On 13th of December 2016, experts from major European companies, research centres and universities participated in a consultation workshop on future R&I priorities in Language Technologies. The workshop was organised by CNECT G3 as part of its wider stakeholder outreach activities. The main goal was collecting information, feedback and ideas to take into consideration in the internal discussions on developing DG CNECT’s future R&I activities in language technologies, including the development and extension of the CEF Automated Translation building block. Short and long term challenges of language technologies in the context of creating an inclusive and accessible European Digital Single Market were a recurring subject of discussions. Sessions covered topics such as existing automated translation and language technologies tools, current and future development and research directions and CEF Automated Translation (CEF eTranslation).

**Session 1: Language technologies from the user’s perspective**

The session focused on the multilingual challenges and needs from the perspectives of large scale business users, SMEs and developers in different business sectors. Discussion centred on available and missing language technology support, R&I directions and prerequisites for developing and integrating further innovative solutions. Presentations were given by representatives of the BBC, Daimler, Booking.com, Damon, Tingtun, EurActiv and Facebook.

The most notable challenges mentioned include:

- adapting to constantly changing end-user/consumer behaviour;
- developing and integrating language technology solutions and billing models for different business needs;
- the lack of language resources and sharing data in a re-useable way for research purposes;
- possible new restrictions on data sharing after the European Data Protection Regulation enters into force in May 2018.

Privacy, confidentiality, security and intellectual property concerns limit the use of Cloud solutions. With non-EU based companies dominating the market, finding European partners for collaboration is also difficult.

The needs from the users' perspective were:

- assimilation between content management systems and across borders;
- integration of language technology solutions;
- collection and maintenance of language resources in all languages (with additional focus on lower-resource & minority languages);
- speech recognition and machine translation solutions and text-to-speech and speech-to-speech solutions;
• research in deep learning application and developing neural machine translation.

**Session 2-3: What Research is needed? How to bridge research and innovation?**

The second session focused on research challenges to be resolved in order to improve existing language technology solutions, with presentations from MT/NLP providers and developers and LT researchers (Artificial Solutions, Bertin.it, ESTeam, Tilde, DFKI, JSI, ILSP, FBK, UPM, InterACT). The subsequent discussion explored ways to ensure that research results feed into innovative solutions, and which means should be used to promote the take-up of innovative LT solutions.

The participants suggested the EC consider the creation of an accessible ecosystem of multilingual and semantically interoperable linguistic data, language resources and tools and services that would help provide all European languages with all key technologies, giving EU citizens equal opportunities to use devices, applications and services in their own language. In addition to collecting language resources, their maintenance, and clearing of the legal requirements for licencing and use are also fundamental, as is finding appropriate business models. Research for improving translation models and performance for under-resourced languages in medium and long term and making neural MT more usable and deployable for industry and language service providers is important. Among the research areas to achieve this are conversational interfaces, task-specific language technologies, deep learning and deep language understanding, machine learning technologies, smart linked data technology and AI in parallel with development and research for ease of integration of new solutions.

The general discussion stressed that the EU’s diversity and multilingual nature represent huge potential sources of innovation, and offer a unique position to assemble the best research teams from across Europe. Competition and performance should be placed at the core of EU funded research by involving industry participants as part of the teams. The EC’s role is important in linking proposals with ongoing actions. Targeting higher impact mission- and product-oriented research and adapting funding mechanisms (procurement, start-up-like funding for shorter duration, smaller consortia, funding participants developing open source technology at 100%, funding models with the winning system procured, more agile and more focussed projects) would raise the calls’ attractiveness for industry participation.

Shifting the focus to mission-oriented research aiming for a Language Transparent Society ("human language project") would maximise the impact of the European DSM, enabling citizens to operate free of language barriers across borders. Language technologies are fundamental in facilitating social and economic inclusion of all citizens in the EU. Language technologies are key to enabling access to cross-border services and to promoting equality within the EU, yet they lack political visibility.

**Session 4: How to make European public digital services multilingual**

This session focussed on the EC and private sector's roles in making European public digital services multilingual; and on ways to develop CEF eTranslation so that it is mutually advantageous for the EC, European public administrations and the LT community and industry. The discussions explored its potential, existing features and planned extension.

The long-term goal behind CEF eTranslation is to contribute to the completion of a Digital Single Market where all citizens can gain access to the same public services in their native languages. It will enable citizens to use European public digital services in the language of their choice by making Digital Service Infrastructures (DSIs) accessible and operable in any of the 26 CEF languages. Currently, it provides translators and public administrations of EU institutions, Member States’ national public administrations and some public online services with access to a machine translation service.
Activities funded at present include much needed data gathering that will, through the European Language Resource Coordination initiative, contribute to future sustainable distribution and sharing of language resources for non-commercial and commercial purposes (e.g. training, adaptation and evaluation). In addition to machine translation, multilingual knowledge, interoperability assets and applications development are all crucial to meet the needs of the DSIs, and the development of additional NLP processing tools could address the gaps present in the CEF eTranslation platform’s overall architecture.

To reach maximum impact, it is important to raise awareness of the CEF eTranslation and language technologies at all levels. Policy measures, such as a legal requirement for pan-European online services to be multilingual by 2030, for example, could drive the take-up for CEF eTranslation. Buying and integrating the most developed technologies on the market as elements of the CEF eTranslation system and offering these to public administrations for use (starting with some prioritised services in the form of demonstration projects) would also boost service demand and awareness. Participants mentioned adapting funding, involving industry and the research community, making infrastructure available at affordable expense and/or the EC kick-starting such infrastructure with Member States by contributing to its basic maintenance and creating accessible repository of language resources as possible solutions. The need for computing resources and services in addition to data could be addressed by EU-backed business models which would allow smaller LT players access to European Computer Centres or other infrastructures, considering LT links with high performance computing. Focusing on additional tools and services alongside MT would allow smaller players to benefit and contribute through opportunities that go beyond use by the public sector.

**Session 5: Short and long-term challenges in language technologies for the benefit of end-users in the European Digital Single Market**

The final session was devoted to short and long-term R&I priorities, focusing on identifying innovative European LT solutions which could facilitate the emergence of the multilingual Digital Single Market. A “language transparent” Europe would mean recognizing that challenges and visions for language-technology-enabled Europe beyond H2020 are crucial in the process of creating the DSM.

Participants agreed that in order to achieve the DSM’s mid-term goals, policy needs to shift towards language-technology supported multilingualism backed with political support and determination. European LT research and development capacity is hindered by industrial and market fragmentation and lack of long term vision and coordination, including the lack of coordination between the multiple LT initiatives across Europe and the lack of collaboration between the EU and industry. Due to the cross-linked nature of industrial and research sectors, this impact extends across domains where real-life applications appear.

Technologies such as speech recognition are crucial to interactions with devices performed through speech; adding a semantic layer ensures that information is understood by stakeholders. Projects should be designed with interoperability/cross-border application. This would allow people and systems to work together and would be a prime driver for LT. Launching the “Human Language project” (deep natural language understanding and generation) would generate new research outputs, new approaches and paradigms, would have huge innovation potential and could act as a bridge between research and education.

Focusing on micro-funding to support novel ideas and start-ups in the short-term (incubators, targeting mission-driven innovative solutions, increasing match-making mechanisms between researchers and technology developers, networking, community building and activities on awareness rising) would advance full implementation of the multilingual DSM in the mid-term. Coordination of infrastructures, possibly with Member States involvement, would amplify the potential of a multilingual DSM.

Support is needed not only for development, but for deployment too, using a sectorial approach. Use of CEF eTranslation should be eventually extended to "services of public interest". For its development to be improved, the participants stressed the importance of public procurement on both the EU and MS level, as a
way to generate demand in the LT sector. EC-launched challenges on languages, under CEF eTranslation, with inducement prizes for improving quality for a specific language for example, adjusting funding based on the projects’ results could prove to be profitable for all parties. Future funding programmes taking a mission-oriented approach based on analysis of needs and goals, while being less prescriptive about the choice of methods to reach the required result, would maximise their impact.

The participants confirm that Europe has the potential to lead globally in addressing multilingualism issues, and main international events such as the World Summit on Information Society 2017, could be a good opportunity to hold a session dedicated to multilingualism/language technologies. MT and LT should be seen as part of sustainable growth and offering the possibility to open up new opportunities.

Outcome

The EC is committed to serve the LT community and the reflections and contributions of the workshop will be part of the internal discussions on future funding programmes. The EC recognises the importance of such workshops and meetings with the four groups of stakeholders (private sector users, public sector users, research and providers) in creating a mission-oriented programme, and is keen to institutionalise the discourse in the future. For this purpose, participants were asked to send their suggestions for future debates. The outcome of the current meeting is a list of identified discussion points for further meetings and deliberations to follow.

The EC will consider the need for additional institutionalized events with the four stakeholder groups, focussing on various sectors and with the participation of representatives of the appropriate services offering EU funding opportunities or hold such meetings as side event(s), for example at METAForum 2017. The EC will also consider the feasibility of main events with user-oriented side events (health, tax, etc.) under the umbrella of events in Council presidency countries with European governments and policy makers participating in the dialogue.