MACHINE TRANSLATION OF EUROPEAN LANGUAGES: PROBLEMS, CHALLENGES, PROSPECTS

Andrejs Vasiļjevs
Tilde, LATVIA
andrejs@tilde.com

The language barriers

The European Union has put a lot of efforts to remove internal barriers for free trade and movement. But we still have invisible but strong barriers making it hard to access information and communicate across countries: the language barriers.

Language is the key element of our social and business communications, an essential part of our identity, social and cultural treasure of every nation. Multilingualism is a source of richness and diversity in Europe. But language diversity also creates barriers for business, communications and cooperation across Europe and globally. Language barriers become an obstacle in advancing European unity and competitiveness.

Each of us has probably faced a situation of being lost in translation. According to the studies by Eurobarometer (2011, 2012) only 39% of European Internet users use other language than their mother tongue to communicate online (e.g. using email, Twitter, Facebook etc.). 46% of Europeans are not able to hold a conversation in other language, and about the same percentage do not read content in a foreign language. 43% of Internet users in Europe do not purchase goods or services in non-native language. For three in five Europeans translation has an important role to play in their everyday life.

Language diversity comes at a price. It is expensive for business. If a small Lithuanian company wants to sell its products all across Europe, it is very costly for it to translate all texts and manuals into almost 30 languages. It is expensive for governments. The EU institutions spend EUR 1.1 billion for translating official communications into 24 official languages of the EU. They are dealing with 552 translation combinations.

Applications of machine translation

How do we reduce language barriers while preserving and strengthening our languages? The exponentially growing volume of multilingual information by far exceeds the capacity of human translators to meet the demand for translation.

Like in many other fields, help can come from technology. Language technologies are the most efficient tool in terms of time and costs, helping to overcome language barriers. Machine translation is the only viable solution for instant and cheap access to information in foreign languages. This is why machine translation (MT) is a critically important technology for multilingual Europe.

It was pure science fiction to imagine just some decades ago there would be computers and robots that understand us and translate into many languages. Now such language technologies are becoming reality. With the help of global services like Google Translate and Microsoft Translator, usage of machine translation has exploded in the recent years.

Automated translation cannot substitute a human translator, but it has proved to be extremely useful in numerous applications. Machine translation almost instantly provides access to information written in a foreign language that otherwise could not be used or would require substantial time and costs to translate. Machine translation can make websites multilingual, it can facilitate cross-lingual information search and analytics, it can be applied not only for written texts but also for spoken language, e.g., dubbing video content. Machine translation boosts productivity of translators by providing a draft translation to be improved and edited. In some specialised areas, machine translation can provide translation of almost publishing quality.

The potential of technologies is vividly demonstrated by the European Commission - one of the world's biggest users of machine-translation. Automated translation is an everyday tool for more than 2000 translators at the European Commission helping to raise productivity and cut expenses.

Innovative language technologies can strongly boost competitiveness of big and small businesses. The same holds true for governments. With machine translation, governments can save money and provide better services to their multicultural citizens.

Language technologies in general and machine translation in particular are the instruments for narrowing the digital divide and overcoming language barriers to facilitate communication and commerce. They are helpful to:

- Enable citizens to access multilingual information across Europe;
- Build European competitiveness by reducing language barriers to enter new markets;
- Facilitate access to the rich multilingual cultural heritage of Europe;
- Increase efficiency and reduce costs of professional translation;
- Reduce the threat of globalization to national identity;
- Facilitate e-democracy: participation of citizens in the pan-European political debate.

The challenge

In the 1950s, when research in this field began, an optimistic approach prevailed: it was thought that automated translation could be made using dictionaries or simply transferring the words from one language into the other. But soon enough it was realised that languages had different structures. In addition, polysemy of the words and the richness of means of expression made this task extremely difficult. A good translator understands the meaning of the text and conveys it into the other language as close as possible. As the computer cannot understand the meaning the way a person does, different syntactic and semantic analysis and statistical methods have to be applied.

Particularly, statistical methods achieved a major break-through in machine translation and are at the core of the Google Translate and many other systems. At the same time, there is a lot to improve in the quality of translation. New generation technologies should come from the European research and help to overcome the quality deficiencies, letting to get closer to the structure and meaning behind human language.

Particularly challenging are the smaller European languages with complex morphology and linguistic structure. Due to economic reasons, current business developments are mostly focused to the largest languages only. For example, although Google Translate covers over 70 languages, translation quality for Baltic languages is significantly worse than for English, French, Spanish and other larger languages.

While assessing the level of development of language technology, experts concluded[1] that digital support (machine translation and other language technologies) for 21 European languages is weak and the future of these languages is endangered in the digital age.

Dedicated actions are needed to address this "technological-gap". As stated by Valdis Dombrovskis, the Prime Minister of Latvia, "The only way to ensure future existence of our language is to provide its users with equal opportunities with the users of larger languages."

The vision

European network of excellence META-NET has created a compelling vision and roadmap for the advancement of translation technologies. According to the META-NET Strategic Research Agenda[2], if the European community makes a dedicated push, we can get rid of many language barriers by 2020 and thus fully implement the common

[1] META-NET White Paper Series, Springer, 2012, www.meta-net.eu
[2] META-NET Strategic Research Agenda for Multilingual Europe 2020, Springer, 2012, www.meta-net.eu/sra

digital market. Europeans will be able to communicate with one another, with their governments and with web services in their mother tongue.

The goal is a multilingual European society, in which all citizens can use any service, access all knowledge, enjoy all media and control any technology in their mother tongues. This will be a world where written and spoken communication is not hindered by language barriers anymore and where even specialised high-quality translation is affordable.

Achievements

The vision for the future is based on the recent accomplishments. The EU Framework Programme for Research and development FP7 and Competitiveness and Innovation CIP have resulted in several world-leading technologies in machine translation. Probably the most prominent example is the Moses toolkit that has become the most widely used machine translation technology in the world.

Tilde is proud to lead the development of the online platform LetsMT! for generation of specialized multilingual machine translation systems. LetsMT! is based on recent advances in statistical machine translation, resulting in a tool that "learns" how to translate by analysing millions of translated sentences. The LetsMT! platform enables the development of various specialised machine translation systems that are particularly suited for smaller languages and specialised translation. It helped *Tilde* to create numerous domain- and task-specific MT systems with significantly better translation quality than Google Translate. Good progress has also been achieved in the creation of the TaaS (*Terminology as a Service*) platform to address the need for correct terminology in both human and machine translation.

Latvia and Lithuania target the EU Structural funds for support to the promotion of their languages. Lithuania recently began implementing a large-scale programme to develop the Lithuanian language for the needs of Information Society. Machine translation is one of the key projects in this programme. *Tilde* is in charge of designing a public machine translation service for Lithuanian citizens. In Latvia, machine translation development is advanced under the *Language Shore* initiative in the framework of the Competence Centre Programme. A major step has been taken by the Latvian Government with the creation of a specialized e-Government machine translation solution to make online services accessible to multilingual communities in Latvia and abroad. Estonia and several other countries also have dedicated programmes for the development of language technologies, including resources for machine translation.

A need for targeted actions

In order to achieve the vision, concerted actions involving the major stakeholders from governments, research and industry are necessary. The following areas should be covered:

- Advanced research for the next major break-through in machine translation quality;
- Facilitating the creation of an open European translation and language service infrastructure;
- Fostering usage of machine translation in innovative applications and integration in various information services and systems;
- Application of automated translation to address major societal challenges such as education, job market, democratic participation and public services;
- Addressing the "technological-gap" in machine translation and other technologies for smaller languages;
- Promotion of the creation and distribution of linguistic resources for machine translation and other language technologies;
- Elaboration of the tools and methods of evaluation and assessment of machine translation;
- Inclusion of translation professionals and enterprises into the entire research and innovation process.

For this, it is important to mobilize and coordinate national and international stakeholders, initiate and synchronize national R&D activities, target EU R&D programmes to achieve a break-through in innovation and global European leadership.