

Speech of Kimmo Rossi (European Commission, CNECT.G3) at the meeting of the chairpersons of the Committees on Education, Science and Culture and the Committees on the Development of Information society, session II, at 14.00 on 27 September 2013

In the recent speech on the State of the Union, Commission's president Jose Manuel Barroso said: "We know that in the future, trade will be more and more digital. Isn't it a paradox that we have an internal market for goods but when it comes to digital market we have 28 national markets? How can we grab all the opportunities of the future that are opened by the digital economy if we don't conclude this internal market?"

While languages are a cultural and societal richness to Europe, the multitude of languages also makes it make it difficult to conclude an internal market, especially online. For example: I want to buy something from a Polish online shop, but if the web page is only in Polish and I don't understand it, shopping stops there. In fact, 80% of European eCommerce sites are in one language only. This is even more alarming as all areas of commerce and business are moving to the online space and rely increasingly on consumers serving themselves in online shops.

In the Internet era, the amounts of text that we need to translate are far too vast for any number of translators to translate. The web content is constantly growing and becomes more and more multilingual. For example, the European Union's website, Europa, has some 6 million pages of content. If we want to translate all of that into the 24 official languages of the Union, it means some 140 million pages– this work would require some 100.000 translators working full time for one year. In general, buyers of translations are more cost-aware than ever. Also, translations are needed almost instantly – otherwise they are no longer relevant. As a result of all this, too much interesting text never gets translated. Or it is translated by poor machine translation, distorting the message. This problem is more prominent for smaller languages, like the Baltic languages, Finnish, Hungarian, Maltese etc.

Actually technology can help a lot to bridge the gap between supply and demand of translation services. At the same time, it helps the small languages. Machine translation is revolutionizing the way people find and use information in the internet. Unfortunately, most of the online translation engines tend to work best when translating into English or into another major world language. If I want to use Google or Microsoft translator to translate into Finnish or Lithuanian, the result may be unreadable, and at best, there are errors in almost every sentence. But even poor machine translation makes these languages accessible.

What does the European Commission do to address the gaps and problems?

In the 7th framework programme for ICT research and development, language technologies have been supported by over 150 MEUR. As a result, we have arrived a huge step closer to a Europe without language barriers. Promising pilots have created automated translation solutions for less widely spoken languages, and systems that let citizens and small companies customize their own machine translation systems tailored to their own text types. Part of this 150 MEUR investment has been used to set up an integrated repository of language resources, which still needs to be completed. We have also created an alliance uniting research in linguistics, language

technology and created links to the guardians of the 24 European official languages in the Member States.

The work done so far forms a solid basis for the missing mile that we have to pass before language barriers disappear from the digital market. To do this, we have three main tools, the Horizon 2020, the Connecting Europe Facility (CEF), and the structural funds.

Horizon 2020 unites the research and innovation programs that were previously separate. It allows better integration of the different phases in the life cycle of technology, allowing a smoother path of research results into innovative pilot projects and better solutions for citizens.

To complete the technology life cycle, research and innovation must be followed by effective deployment. We must produce cost-effective services that citizens and companies need, to provide the ground for growth, jobs and prosperity. This is what the Connecting Europe Facility is about. The Commission proposal for CEF foresees substantial funding to build pan-European digital service infrastructures. They will ensure, for example, that any European company can easily participate in a call for tender launched in any other EU country, or that an entrepreneur can set up a company in another country, where the customers are, or that patients can receive the right treatment or medication anywhere in the EU. These pan-European digital services must of course speak the languages of their users: all Europeans. This is why the CEF legislative proposal contains a specific “building block” called “Automated Translation”. It is called a building block in the CEF proposal, because it will become an integral part of all digital services and will make them multilingual: for example, translating the tender specifications or medical prescriptions into all EU languages.

EU funding programs are a good catalyst for ground-breaking research, innovation and pan-European infrastructures. But they are not enough. We have to do more to move from roadmaps to useful systems and services, from policies to added value for taxpayers. And this is where we need more collaboration and coordination between EU programs and national activities in the EU Member States. We should learn to work in synchrony, for a common goal. This is especially true for language technologies. For example, if we want to develop powerful automated translation technology for Lithuanian, we need, as “raw material”, Lithuanian language resources: for example, collections of texts and terminologies. These, in turn, are more likely to be found and managed in Lithuania, and not in Brussels. What can be done at EU level is to create a common repository and infrastructure that allows all these linguistic assets to be shared, and allows powerful translation technology to be developed equally for all EU languages, not only for the big languages.

EU member states need to be empowered to do more and to decide more about their own language resources. This comes with responsibility on results. We need local, regional and national initiatives to complement the more centralized EU programs on language technology. This is because your own language is the prime interest of your country and your region. And by working together we can support each other and make sure that all languages are preserved in the online space. By doing so, we make sure that valuable and useful online services are available to all Europeans, no matter what language they speak.

To conclude, I wish to express my congratulations and my gratitude to the Lithuanian presidency and to the Lithuanian parliament for their extraordinary efforts for the

cause of European multilingualism. And I thank for this opportunity to present and explain the European initiatives and my view our common challenges in the future. Thank you for your attention.