

Human language technology enables breaking communication barriers within the EU

HLT in Slovenia: Overcoming Language Barriers

Jerneja Žganec Gros

HLT: the European Dimension

Human Language Technology (HLT) plays a truly unique role in the European Union from the cultural, social and economic perspectives. For all citizens to become e-included in the information society, the products and services of that society must be accessible in their own languages. Language transparency, in which products and services are offered cross-lingually and in a localised manner, is one of the major prerequisites for the successful establishment of a common European market.

Research and development in HLT is nowadays rapidly transferred into commercial systems, and it is becoming increasingly pervasive in everyday life. Apart from the well-known areas of **machine translation (MT)**, **automatic speech recognition (ASR)** and **text-to-speech synthesis (TTS)** – all of which are now packaged into inexpensive applications of varying quality for the commercial and home markets, at least for languages with a broad speaker base – other previously unexpected areas of use are emerging, such as speech-to-speech translation systems.



Dr Jerneja Žganec Gros, COO

122

alpineon



Is HLT equally well supported for all EU languages?

The extent to which the languages spoken in Europe have been researched systematically varies widely from language to language, with a minority being well investigated – such as English, French and German – while some are hardly being addressed at all. Most often, the new Member States have not been in a position to develop strong technologies for their own spoken and written languages.

HLT in Slovenia

Slovenia is a new EU Member State in which about 2 million speakers communicate in Slovene. Although the market for language technologies is rather small, this can also be viewed

as a competitive advantage, because most of the resources and expertise of the language are based within Slovenia and large market players are not yet ready to invest in developing HLT products for Slovene.

According to the e-Europe action plans, and most recently the i2010 Communication, the Slovenian government has decided to foster the development of future-oriented public services and to create a dynamic environment for e-commerce. In the past 5 years, the Slovenian Ministry of Information Society, later followed by the Slovenian Ministry of Defence, launched a series of calls for targeted research projects aimed at developing HLT resources and pilot speech-enabled systems for Slovene.

Alpineon, a Slovenian research SME and member of the Slovenian ICT Centre of Excellence, is coordinating two of these projects; the other partners involve three faculties of the University of Ljubljana, the Jožef Stefan Institute and the Fran Ramovš Institute. The first targeted project was commissioned by the former Slovenian Ministry of Information Society in response to the needs of the public sector for speech technologies to be

applied in systems for information access for all citizens, and especially for impaired persons, via voice in Web and voice portals. It is based on the European action plan "Accessibility of Public Web Sites and Their Content". The project builds on the experience gained in creating the HOMER system – an award-winning voice-driven Slovenian Web reader especially designed for visually impaired users – in close cooperation with the Slovenian Association of Blind and Partially Sighted Persons.

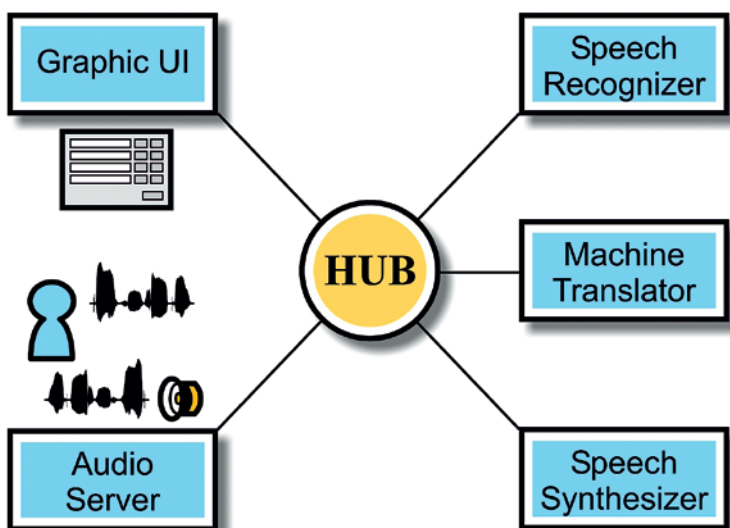
Many of the basic language resources for Slovene were created within the project. More tangible results are pilot speech technology engines with standard interfaces, which can be used as plug-in modules in standard voice portal and voice browser applications that allow access to public service information via voice. They can further be applied in assistive technology products, and for improving access to online facilities for visually impaired people, such as to e-libraries and other e-learning platforms. The public sector, in turn, must seek to ensure the accessibility of all public web pages by complying with W3C standards, with an emphasis on the Web Accessibility Initiative recommendations for e-content accessibility to the impaired.

The second targeted research project, VoiceTRAN, focuses on speech-to-speech translation. The aim of the project is to build a robust multimodal speech communicator able to translate simple domain-specific sentences between Slovene and English; more complex phrases can be entered via keyboard using a graphical user interface. The VoiceTRAN communicator is composed of a number of HLT servers – audio server, graphical user interface,



The VoiceTRAN project received the eChallenges exhibitor award. Dr Jerneja Žganec Gros and European Commissioner Janez Potočnik

speech recogniser, machine translator and speech synthesiser. The initial version of the VoiceTRAN communicator was presented at the eChallenges e-2005 event. It received the runner-up exhibitor award for the presentation in the Slovenian ICT Pavillion.



Alpineon is a high-tech SME providing turnkey software and hardware solutions in the ICT, automotive and industrial automation control fields. The team has extensive experience in hardware and software development, including CTI applications, VoIP devices and services, and speech technologies.

Alpineon specialises in developing state-of-the-art speech technology products for Slovene and other Slavonic languages. Its members have built the first commercial speech technology-enabled services in Slovenia, including a carrier-grade SMS-to-voice application along with an e-mail reader service. Its reference list ranges from voice portal applications to individual speech engines, such as text-to-speech synthesis, speech recognition, speech-to-speech translation, to be used in carrier-grade applications as well as in embedded devices.

Alpineon is a member of the Slovenian ICT Technological Network and several technological platforms. The company is currently involved in numerous industrial, national and international R&D projects, where it cooperates with leading academic institutions, institutes and major industrial players in Slovenia and abroad. Alpineon is coordinating the e-inclusion Eureka project E!3864 iTEMA and the largest HLT targeted research project in Slovenia, VoiceTRAN, funded by the Slovenian Ministry of Defence, which aims at building a speech-to-speech communicator to be used by the armed forces in peacekeeping operations.