MORENA – a project never realized (so far)

Marko Tadić[1]

[1]University of Zagreb, Faculty of Humanities and Social Sciences
marko.tadic@ffzg.hr

1. Introduction

After the successful completion of the ICT-PCP project CESAR within the META-NET initiative in January 2013, some of the partners under the coordination of Tamás Váradi wanted to prolong their fruitful collaboration and started seeking funding opportunities. At that time it seemed that the funding for LT would be scarce and that opportunities similar to the existing ones in earlier times would not be easily available. Therefore, we engaged in finding opportunities beyond our traditional funding sources and programmes that we had known before.

First, we tried our hands at the Scientific support to Danube Strategy. After two meetings including presentations of the CESAR project results with Tamás Váradi, Marko Tadić, and Dan Tufiș1 in attendance in Ispra in March and July 2013, we genuinely believed that we would be able to find funding for the jointly proposed project(s). However, this turned out to be true more or less exclusively for JRC in Ispra, whereas partners from the Danube region were not expected to take considerable part, let alone fully fledged collaborative projects. Nevertheless, we learnt that transport problems and their solutions played one of the major and integrative roles in the Danube Strategy, so we tried to relate that to our field of expertise – language technologies.

In January 2014 at the Networking Day on Horizon 2020 Work Programme 2014-2015 and Connecting Europe Facility in Luxembourg, the partners convened again and we got acquainted with the CEF programme there. Trying the CEF programme as one of the possibilities was initiated by Tamás Váradi, thus he suggested applying to its calls for project proposals beside the H2020 calls that we had been familiar with since we considered them as the prolongation of FP7. The CEF programme predominantly funded the development of systems that were regarded as

[1] Formally not a member of CESAR consortium, but he belongs to the Danube region.

DOI: https://doi.org/10.18135/VT70.15

enhancement to the connections between EEA countries such as transport infrastructure (e.g. intelligent transport systems (ITS), multimodal transport etc.). The CEF Telecom subprogramme, however, focused mainly on the telecommunication infrastructure (e.g. broadband internet connection), therefore a niche for LT opened within the framework of Digital Service Infrastructures (DSIs) where machine translation (MT) became one of the horizontal services that would ensure the multilingual usage of other DSIs.

Throughout our discussions about the possible research topics that would fit into the scope of the first calls and knowing that transport also played an important role not just in the Danube Strategy, but also in CEF, an idea to combine the ITS with LT crystallized in us in the form of a proposal for a service that would offer the multilingual approach to road traffic information. Such information is usually broadcast in the most natural and desired way – as spoken news about road conditions, preferably in the driver's native language.

2. ITS and multilingualism

The issue of language in real time traffic information is a challenge that seemed to be neglected in both the practice and at the policy level in ITS. It seemed obvious that regardless of the source of data and the technology of collecting and presenting them, seamless cross-border traffic information could not be delivered without overcoming language barriers. These impediments have been unable to be solved even within EU member states of Europe where administrative borders have disappeared. These multilingual requirements for a seamless cross-border service were not sufficiently addressed in strategic ITS documents.

Language-independent solutions such as pictograms have their place but their expressive power can't reach that of a natural language, let alone the native language of the driver. It can be stated with justification that ITS Deployment Guidelines seek to eliminate the problem of multilingualism rather than provide a solution (Váradi et al., 2015a).

3. Road Traffic Information Systems

Road traffic information systems still operate in a fragmented manner, isolated by language and even by country borders. Traffic-related information is provided by various local partners, while data collection and processing are carried out by national Traffic Information Centers (TIC).

Users can access information on the website of the national TIC, via mobile apps, through a call centre, or by listening to radio and TV broadcasts. However, the most commonly available channel is public radio broadcasting traffic information relevant to the whole country. Surveys indicated that this is indeed the most popular source and medium of traffic information, which drivers insist on using.

Most countries operate a road traffic information service on a national scale. Cooperation between them is limited. Seamless cross-border traffic information provision still remains only a long-term objective. The potential for data ex-change between the national services has been greatly facilitated by the development of a standardized notation and ontology in the transport domain, namely DATEX II. Nonetheless, no similar progress has been made to remove the linguistic barriers to streamline cross-border traffic information services. Real-time road traffic information services have typically been provided in the official language of a country, accompanied by occasional broadcast of limited scale and/or limited time in a foreign language, i.e. information for tourists during summer in English.

Road users crossing countries normally do not speak the language of the neighbouring country, thus they are isolated in a foreign language medium. If drivers do not understand traffic-related news abroad they are prone to get delayed in congestions, or they may even get involved in accidents in more serious cases (Váradi et al., 2015a).

4. MORENA as a possible solution

The solution to the problem of multilingual traffic information is not to avoid the production of natural language messages, but to deliver the information to the road users in languages they understand and in the most natural way they prefer, i.e. spoken native language. Language technology has reached such a level of maturity that it can offer a comprehensive full-scale solution to this challenge through a combination of real-time machine translation (MT) and natural-sounding speech technology.

We proposed the development of a robust, high quality MT system and its deployment in the ITS domain. We put our confidence in the unique infrastructure of the elaborate terminology and ontology (DATEX II) available for the ITS domain. This language-independent ontology represents a comprehensive and fine-grained conceptual system that allows the description of practically any traffic event and condition.

The task of machine translation should boil down to converting the real-time traffic information into a standard DATEX II representation, which can then be mapped into any particular target language and delivered through a text-to-speech system. The technology was expected to use proven components and promised to yield much higher quality translation than what was available at that time through freely accessible general purpose SMT methods (Váradi et al. 2015a).

Today, this effort would most probably include NMT methods, but it should be investigated in advance whether the DATEX II could function as controlled interlingua (glass box) or it should be left to neural networks themselves to form the connection network (black box).

The envisaged technology at that time presented a ground-breaking solution to pre- and on-trip traffic information provision by delivering traffic-related information in the most user-friendly manner, i.e. in the form of spoken messages in one's native tongue, something like an automated traffic news internet radio. It was planned to offer high quality MT through a hybrid approach integrating proven concepts in the fields of machine translation (MT), terminology management, computer-assisted translation using translation memories (TM), and controlled-natural language (CNL) systems. The LT was planned to draw on the standardized data dictionaries and protocols developed in the ITS domain and thus it would have utilized the synergies between the two disciplines involved (Váradi et al. 2015a).

In the first half of 2014, a detailed project proposal was developed to implement the technology described above. We called it MORENA. The backbone of the whole solution was a cloud-based system that would provide the machine translation of traffic information. This was called the MORENA service, which would allow the deployment in a variety of settings. As one possible implementation, the project offered to develop a multi-platform mobile application for mobile devices (smartphones, tablets, etc.), which we referred to as the MORENA application or MORENA app. The MORENA application was not intended to be the sole deployment of the MORENA service and the MORENA service should be assessed in terms of the potential it could offer as embedded technology in the ITS domain and should not be evaluated on the merits of the MORENA application alone (Váradi et al. 2015a). At that time we even had the idea to propose the MORENA service to become one of the CEF DSIs since it could have played an integrative role within the CEF Transport and/or CEF Telecom.

Then, the anticipated deployment of the suggested service was innovative as well. It was supposed to be a cloud-based service, delivered as an application for mobile devices. The drop in roaming charges regarding EEA countries by mid-2017 represented a major breakthrough, enabling mobile devices to become the main channel for broadcasting traffic information services (TIS). Such an online service would have presented the additional benefit over current practice because the information flow through this channel can be tailored to individual users since all relevant information such as preferred language, current GPS position, planned destination, etc. were at disposal from the device itself. Accordingly, delivery of information would have been personalized not only for language, but also for GPS position and destination (Váradi et al., 2015a).

5. Project presentations and proposals

Keeping an eye on the CEF programme, we still decided to submit the MORENA project proposal to the H2020-ICT-2014-1 call in April 2014. Unfortunately, this proposal was not successful, however, that didn't stop the consortium from further developing the idea and trying again.

We tried to look for additional opportunities, so Tamás Váradi and Marko Tadić also attended the Danube Region Transportation Days in October 2014 in Ljubljana, where a presentation was given on the MORENA project as an already established one, with its proposal submitted, as well as the first leaflet produced.

Figure 1. MORENA leaflet at the Danube Region Transportation Days, Ljubljana, 2014-10-21.

In early 2015 we submitted a more refined proposal again, but this time to the CEF Transport Multiannual Call: Specific Call for Cohesion Funds under the Transport Sector. The Funding Objective 2 in this call was defined as "Deployment of new technologies and innovation in all transport modes, with a focus on decarbonisation, safety and innovative technologies for the promotion of sustainability, operation, management, accessibility, multimodality and efficiency of the network". We expected the MORENA project to qualify for "new technologies and innovation" as well as for "safety and innovative technologies". The consortium was enlarged with respective national players in TIC who were included as providers of real-time data on national road traffic conditions. We also expected that their involvement as key stakeholders would help the project to be accepted. However, this proposal was evaluated by experts in transport and they didn't find that the deployment of LT in ITS would be worth-supporting at this stage.

We began to prepare for the CEF Telecom call in 2016, but the consortium was successful with another proposal, the MARCELL project, so the idea behind MORENA was gradually put aside.

Nevertheless, the work on this idea resulted in two papers presented in October 2015 at the 15th ITS Congress held in Bordeaux (see Varádi et al., 2015a and 2015b). Today, these two papers may be considered as pioneering papers towards introducing the usage of language technologies in ITS. In spite of all efforts, they have remained to represent the untapped potential behind the meticulous pioneering work spearheaded by Tamás Váradi.

6. Conclusions

On this solemn occasion we present the activities based on the idea and several proposals for the MORENA project in which our celebrated person was involved in the manner he has always pursued – entirely, thoroughly and with the best interest of the consortia.

Natalis honorem Tamasi Varadii septuagesimum.

References

EasyWay. Data Exchanges DATEX II Supporting Guideline (2012). Available at http://www.rits-net.eu/uploads/media/EW-DG-2012_DTX-DG01_DatexII_02-00-00.pdf

Freudenstein, J., Cornwel, I.: Tailoring a reference model for C-ITS architectures and using a DATEX II profile to communicate traffic signal information, Transport Research Arena 2014, Paris (2014)

Gilka, P., Richter, T.: Result assessment for user acceptance and safety evaluation on motorways with I2V-communication. In Proceedings of the 18th World Congress on ITS, Orlando, FL, USA (2011)

ITS Deployment Guidelines Library (2012). Available at https://dg.easyway-its.eu/DGs2012.

Váradi, T. Tadić, M. Gulyás, A. Niculescu, M.: Language Technology in the Service of Intelligent Transport Systems, Paper number ITS-2878, 22nd ITS World Congress, Bordeaux, France, 5–9 October 2015 (2015a)

Váradi, T., Tadić, M., Gulyás, A., Niculescu, M.: When Will ITS Speak Your Language? Paper number ITS-2955, 22nd ITS World Congress, Bordeaux, France, 5–9 October 2015 (2015b)