

Paper number ITS-2878

Language Technology in the Service of Intelligent Transport Systems

Tamás Váradi^{1*}, Marko Tadić², András Gulyás³, Mihai Niculescu⁴

1. Hungarian Academy of Sciences Research Institute of Linguistics, Hungary, 1068 Budapest, Benczúr u. 33, e-mail: varadi.tamas@nytud.mta.hu, phone: +36203441403
2. Faculty of Humanities and Social Sciences, University of Zagreb, Croatia
3. University of Pécs, Faculty of Civil Engineering and Informatics, Hungary
4. ITS Romania, Romania

Abstract

The present paper discusses the multilingual challenges in ITS with particular reference to real-time road traffic information service. It is contended that the issue of language is not sufficiently addressed either on the level of policy or in the practice of ITS. We suggest that language technology can offer a full-scale and robust solution to the bottleneck. Building on the ITS data infrastructure represented by DATEX II, the technology proposed in the paper is mature enough to produce high quality machine translation, which, coupled with appropriate speech technology, can deliver the most natural way drivers expect travel information i.e. spoken in their native tongue. We suggest that such technology can be implemented in a variety of settings such as an internet-based service delivered to mobile applications or integrated in GNSS systems.

KEYWORDS:

ITS, multilingual RTTIS, language technology.

The importance of real time traffic information in C-ITS and its deployment

Cooperative intelligent transport systems (C-ITS) can be defined as a subset of the general ITS services based on different types of wireless communications to exchange information between vehicles, roadside infrastructure, and centres (Austroads 2015). C-ITS may promote safer, more efficient and environment friendly solutions for transport compared to former simpler applications. Traveller information systems, dynamic hazard warnings and other elements communicate presently mostly via existing cellular communications and radio broadcast technology. Commercial fleet managers use cellular communications widely as well. However the next generation of information exchange will be the 5.9 GHz Dedicated Short Range Communication (DSRC) especially for time-critical traveller information functions, for example congestion avoidance. The article intends to present a novel interoperable application that can be deployed using either cellular communication or DSRC. One of the aims of connected vehicle research in the US is to provide continuous real-time connectivity among vehicles, infrastructure, and wireless devices (USDOT 2015). An ongoing research work intends to develop interfaces to reduce potential risks and focusing the driver's attention on the roadway to avoid any collision. Real time traffic information provided in an easily understandable way for example by automated speech messages in the drivers own language may lead to risk reduction on high-speed roads where considerable heavy vehicle transit occurs.

A new report by the ERTICO Task Force on communication technologies for future C-ITS services aims at providing recommendations to support successful and efficient deployment

of C-ITS services (ERTICO 2015a). The report is supplemented by a Guide about the technologies for ITS services (ERTICO 2015b). Both materials focus on appropriate communication at necessary level of interoperability that is indispensable for any successful C-ITS services. High quality services for the user can be deployed only along a complex and integrated value chain concerning both infrastructure to vehicle (I2V) and vehicle to vehicle (V2V) communications. Eliminating any language barrier within services deployment provides added value and even higher quality.

Several R&D projects have dealt with the role and importance of real-time traffic information in C-ITS deployment. The FOTsis project (European Field Operational Test on Safe, Intelligent and Sustainable Highway Operation) contains C-ITS deployment pilots from an infrastructure perspective. Among the different recommended FOTsis services there is Service 3 – Intelligent congestion control where real-time traffic information really has an important role. There are two sides of Service 3: the traffic data collection and the notification of the recommended speed to the users, see Figure 1 (FOTsis 2015). This task requires an efficient traffic management system with intelligent traffic control measures and adequate traffic information for drivers.

Interoperability at information exchange can be enhanced by applying standards and widely used data exchange protocols like DATEX II. There are well-accepted structures for the emerging Cooperative ITS (C-ITS) architectures. The 1st release of standards for C-ITS in response to the European Commission Mandate M/453 tightens these structures. As a good example the C-ITS architecture developed for an urban use case in the SEAMLESS project (Seamless Traffic Data Dissemination across urban and inter-urban Networks financed by ERA-Net Road – Mobility [SEAMLESS 2011]) utilizes DATEX II for communication (Freudenstein–Cornwel 2014). The innovative application presented in the article intends to us DATEX II as well for its deployment.

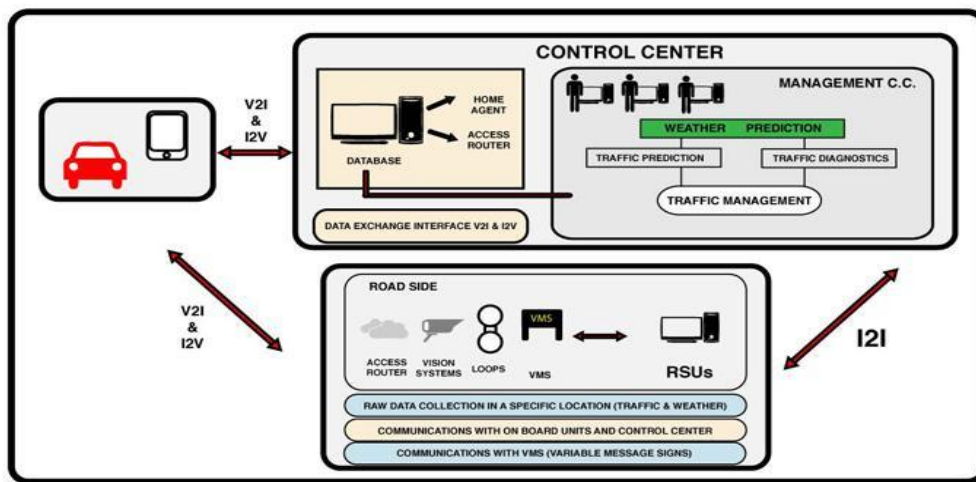


Figure 1 - FOTsis Service 3 Intelligent congestion control communication schema (FOTsis 2015)

ITS policy and multilinguality

The issue of language in real time traffic information is a challenge that seems to be neglected in both the practice and at the policy level in ITS. It seems obvious that whatever the source

of data and the technology of collecting and presenting it, seamless cross-border traffic information cannot be delivered without overcoming language barriers, which are impenetrable even in a continent like Europe where administrative borders have disappeared within the EU.

The multilingual requirements for a seamless cross-border service are not sufficiently addressed in strategic ITS documents. The EU funded EasyWay program in its two phases between 2009 and 2012 worked out a set of ITS deployment guidelines specifying various levels of service for a step-by-step deployment of the service, see Table 1 (EasyWay 2012a). The EasyWay Deployment Guidelines emphasize the fact that "...real time event information influences the route chosen by the road users travelling both short and long distance and inter- and intra-national trips throughout Europe. Therefore, service providers should try to provide the information either in different languages (if possible) or preferably in a language independent format (by using pictograms, symbols, etc.)" (EasyWay 2012b).

Table 1 - Levels of service for a step-by-step deployment of the service for traffic conditions information

CRITERIA	LEVEL 0	LEVEL 1	LEVEL 2
USER INTERFACE	One fixed language	Data available is capable of being provided in a common and shared language (English)	Data available is capable of being provided independent of language
NEIGHBOURING PROVISION	No information exchange	Information exchange to neighbouring only	Neighbouring and beyond Information Provider exchange
LOCAL AND SECONDARY NETWORK INFORMATION	None	Travel information for key routes	Relevant travel information beyond key routes can be provided
LEVEL OF DETAIL (LOCATION REFERENCE)	None	Specific route segments (between 2 junctions)	Specific road segments (approx 1 km)

Language independent solutions like pictograms have their place but their expressive power evidently nowhere near matches that of a natural language, let alone the native language. It can be stated with justification that ITS Deployment Guidelines seek to eliminate the problem of multilingualism rather than provide a solution.

RTTIS: The State of the Art

Currently, road traffic information systems operate in a fragmented manner, isolated by language and even by country boundaries. Traffic related information is provided by various local partners, while data collection and processing is carried out by national Traffic Information Centers (TIC). The standardised national access points for providing traffic information, especially in order to enhance traffic safety, will be deployed soon following the EU ITS Directive 2010/40/EU (EC 2010) and the regulation 886/2013 (EC 2013) based on that Directive. Until then, users can access information on the national TICs' website, with 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 indicate that this is indeed the most popular source and medium of traffic information, which drivers insist on using (cf. Figure 3). Popular as it is, traditional

public radio broadcasting shows some limitations. It is limited in response time, since traffic related news are on air only a few times an hour. Even radio stations dedicated to traffic news only have the limitation of too broad geographical coverage with consequent lack of attention to local events relevant to individual drivers.

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 exchange between the national services has been greatly facilitated by the development of a standardised notation and ontology in the transport domain. However, no similar progress has been made to remove the linguistic barriers to smooth cross-border traffic information services. At present, real-time road traffic information services are typically provided in the national language of the country, with addition sometimes in a foreign language, typically English, but of a limited scale and/or limited time (e.g. information for tourists only and during summer).

The most severe constraint in the current practice of real-time traffic information services is that they typically operate in the national language mostly. This, however, is inadequate to cater even to the drivers on the national road network because at any time a significant number of drivers are in transit across the country and presumably a large part of them do not speak the local language. According to a survey performed in e.g. Austria, the share of transit vehicles was as high as 64.5% in freight transport alone in 2009 (see Figure 2).

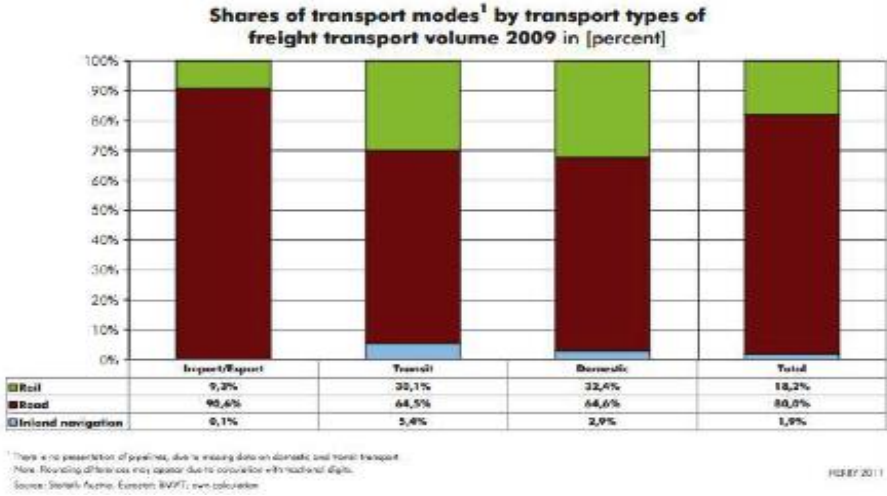


Figure 2 - The share of transit vehicles in Austria in 2009 (Statistik Austria, Eurostat BMVIT, own calculation)

Road users crossing to other countries do not necessarily 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 in more serious cases even to get involved in accidents.

The ever increasing usage of smartphones and other mobile devices is well documented in various studies carried out all over the globe. Constantly, the use case "to gather traffic information" is among the Top10 reasons to use a mobile device on a regular basis

(ASFINAG 2013). The need for natural language traffic information service delivery is not eliminated by the increasing spread of global navigation satellite systems (GNSS) either, which lately provide information on the actual flow of traffic based on sensor data. On the contrary, according to a survey carried out by ASFINAG in the autumn of 2013, 95% of truck drivers on the Austrian motorway network do use still the inflexible and old FM radio as one source of information concerning traffic information while they are on the road (see Figure 3). Based on numerous feedback received so far to the ASFINAG app this is interpreted so that the information is available and considered helpful in principle. A major drawback communicated by the drivers is the way it is presented today (Top3 reasons):

- not always in native language,
- not really available cross-border,
- not as simple as turning on the radio.

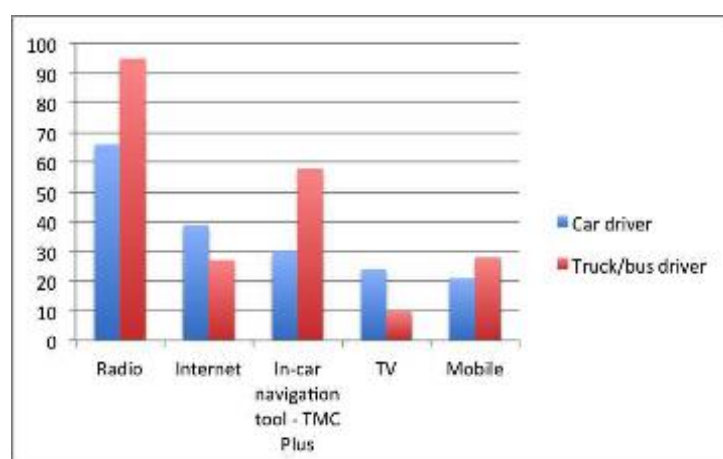


Figure 3 - Results of a survey on sources of traffic information while on the road (ASFINAG 2013)

The solution we propose here will address all of these concerns at the same time. In addition, several studies in the ITS literature (see e.g. [Gilka–Richter 2011]) confirms that the quality of human-machine interfaces plays a key role in user acceptance towards an application. By having a most natural interface, the proposed services will be a much more efficient communicator of traffic information than current practices.

The Solution

In our view, the solution to the problem of multilingual traffic information is not to avoid the production of natural language (spoken) 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 now reached a 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.

While machine translation rightly receives criticism when it is applied to the general domain, it has proved itself in translating texts in specific narrow areas (an early success story is the Canadian system METEO which was functioning for decades since the seventies (Thouin

1982). We propose here the development of a robust, high quality MT system and its deployment in the ITS domain. We base our confidence on the unique infrastructure of the elaborated 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. Furthermore, this underlying conceptual scheme is already available translated in different natural languages. The task of machine translation boils 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 uses proven components and promises to yield much higher quality translation than is possible through freely accessible general purpose statistical machine translation methods.

The technology is amenable to be integrated in ITS solutions in a variety of settings. It provides a generic solution to the multilingual challenge of RTTISs and as such it deserves to be integrated in long-term deployment strategic thinking in ITS. The technology is neutral to modalities of deployment and can be embedded in all sorts of applications providing traffic information. So far we have been focusing on how it can be deployed to automate traditional traffic information services and make them multilingual. In the next section we outline a proposal for a pilot project that aims to implement this technology for five languages and delivers a regional service through an independent online application. This, however, is only one of the possible ways this technology can be employed.

The envisaged technology¹ presents a ground-breaking solution to pre- and on-trip traffic information provision by delivering traffic related information in the most natural and user friendly manner, i.e. in the form of natural, spoken messages in one's native tongue, something like an automated traffic news internet radio. It will offer high quality machine translation through a hybrid approach integrating best practices and 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 language technology draws on the standardized data dictionaries and protocols developed in the ITS domain and thus exploits the synergies between the two disciplines involved in ITS.

It should be pointed out that this technology can be usefully employed in a monolingual context as well to automate the work of national language traffic information service providers. When this technology is applied to the same language, we get a fully automated traffic information service that is capable of 'understanding' the stream of incoming traffic

¹ A detailed project proposal has been developed to implement the technology described in the present paper on a regional scale, called the MORENA project (cf. <http://www.morena-project.eu>). The backbone of the whole solution is a cloud-based system that provides the machine translation of traffic information. This is called the MORENA service, which allows deployment in a variety of settings. As one possible implementation, the project offers to develop a multi-platform mobile application for mobile devices (smartphones, tablets, etc.), which will be referred to as the MORENA application or MORENA app. The MORENA application is not intended to be the sole deployment of the MORENA service and the MORENA service should be judged in terms of the potential it offers as embedded technology in the ITS domain and should not be evaluated on the merits of the MORENA application alone.

reports and producing standard traffic information messages more flexible and at wider scale than the traditional RDS-TMC system with its very limited content. Removing the heavy manual work-load currently involved at both ends of the information chain offers the prospect that traditional traffic information services can be more efficient and closer to real-time.

Language technology approach

This innovative approach integrates existing best practices in two domains, the ITS domain and the language technology (LT) domain. Within the LT domain this approach makes use of proven concepts and technology in the fields of terminology management, machine translation, computer-assisted translation (CAT) using translation memories (TM), and controlled natural language (CNL) systems. All methods and technologies suggested for this approach have been tested and are robust enough for the real-world applications. What is innovative within the language technology domain in its configuration for this solution is the specific interactive model of three levels/methods/approaches that so far have co-existed with each other, but that have not yet been integrated into a single, coherent and production-oriented framework:

1. terminology management,
2. hybrid process models for MT plus translation memories,
3. CNL authoring approaches.

The specific reasons why this integrative and interactive approach is chosen lie in the inherent limitations of each of these three levels and methods when it comes to overcome cross-lingual barriers in information systems as well as communication processes.

With the new EU regulations on ITS and the already mature DATEX II information interchange format, the context of the LT application is very well specified. The constrained structure of the natural messages, the limited vocabularies (see <http://www.datex2.eu/content/datex-ii-v20-data-dictionary>), the existing formal multilingual terminologies, and the clear classification of the relevant traffic events allow LT tools to achieve a higher accuracy and much faster processing of the ITS relevant messages than in the case of unrestricted texts. The accuracy and response time are crucial elements in this application domain. We see the DATEX II XML encoding schema as a language independent representation level (interlingua) for the information which could have been initially provided in any of the languages of the project (and whatever other languages in the future). Vice-versa, the messages in DATEX II interlingua representation are source for production of any target language messages (in whatever languages in the future).

As mentioned before, this interlingua-based approach to this translation tasks is facilitated by the limited domain, very well structured and using restricted linguistic constructs and closed vocabularies. It has all the advantages advocated by classics of the interlingua-based MT and resists all early criticism regarding the expressivity and coverage. Moreover, due to the restricted language and precise terminology, one may discard expensive NLP phases (lexical and syntactic disambiguation, discourse phenomena) and use much faster IE techniques (e.g. named entity recognition, regular grammar patterns, event frames). In translating DATEX II messages into natural language, the same specificities of the application domain turn the process into a very fast procedure supported by standardized multilingual patterns and multilingual lexicons and terminologies. (For a more detailed discussion of the language technology aspects, see Váradi et al. 2015)

The described technology addresses the limitations noted above in RTTIS provision on several counts:

1. It will eliminate language barriers through high quality MT of traffic related information from and into the national languages of the participating countries (plus English offered as a lingua franca to speakers of other languages until the service is up-scaled to include their language).
2. It will deliver the traffic information in spoken language, the most natural medium. This is an obvious requirement from the perspectives of safe driving.
3. It will cover traffic information that is of direct relevance to drivers both in terms of their geolocation point and their destination. This personalized and user oriented traffic information service has a clear advantage over current practices and on top of that it is delivered in the medium, spoken news, that drivers predominantly prefer while driving.
4. This user customized service is delivered through the Internet to their personal mobile devices, typically smartphones, which are becoming widespread on a massive scale. The viability of this service is ensured with the planned abolishment of roaming charges within the European Union.

Deployment

In addition to the innovative core multilingual language technology employed, the anticipated deployment of the suggested service will be innovative as well. It will be a cloud based service delivered as a mobile application for smart phones or other mobile devices. The recently announced phasing out of roaming charges by mid-2017 represents a major breakthrough, leading to mobile devices as the main broadcasting channel for traffic information services (TIS). An online service has the additional benefit over current practice that the information flow through this channel can be tailored for individual users since all relevant information such as preferred language, current GPS position, planned destination, etc. are at disposal. So delivery of information is personalised not only for language, but also for GPS position and destination.

RTTIS vis-à-vis GNSS

In the light of the various satellite navigation systems and widely available GPS equipped mobile devices, the question arises whether spoken traffic information messages are not likely to become obsolete. It is not the task of this paper to justify current traffic information services or discuss how they relate to GNSS based services. We merely note that the information that traditional traffic information services provide partly overlaps, partly complements what one finds in navigation systems and there is no reason why navigation systems should not integrate the information services provided by the traditional service providers. Table 2 presents a comparison of the information content typically presented in RTTIS and GNS systems.

Table 2 - Comparison of GNSS systems with RTTI services

Category of traffic information	Navigation devices	RTTI (web)services
Traffic speed flow	Available, if equipped with RDS/TMC receiver or DAB/TPEG receiver or Internet connection. Only current situation can be displayed.	Available. Historical and predicted information can also be provided.
Traffic events	Available, if equipped with RDS/TMC receiver or DAB/TPEG receiver or Internet connection. Typically displayed as a pictogram or as a list, with minimal details.	Available. Detailed information is provided, for example: source of the data, duration of the event, description, timestamp of last update.
Road weather events	Available, if equipped with DAB/TPEG receiver or Internet connection.	Available.
Weather forecasts	Available, if equipped with DAB/TPEG receiver or Internet connection.	Available.
Animal warnings	Available, if equipped with DAB/TPEG receiver or Internet connection.	Available.
Special events	Available, if equipped with DAB/TPEG receiver or Internet connection.	Available.
Police check points	Available, no data connection required.	Availability depending on the baseline map used.
Road tolls	Available, no data connection required.	Availability depending on the baseline map used.
Parking places	Available, no data connection required.	Availability depending on the baseline map used.
Webcams	Not available. Just position of speed cameras might be provided.	Available with live video feed.
Bike stations	Available, no data connection required.	Availability depending on the baseline map used.

It is easy to see how the two systems usefully complement each other. Navigation systems are great in indicating up-to-the-minute traffic conditions. However, drivers may well be interested in the reasons why the particular congestion ahead has developed, how long it is estimated to last, etc. An important part of traffic information consists of warnings about road conditions, weather, animals, etc. that navigation systems currently usually do not report. As was discussed earlier, current GNSS based applications operate in various languages but the localisation is limited to the user interface and the limited set of predefined navigation messages. If information about real-time traffic is conveyed at all, it is presented mostly visually with all the evident limitations. In conclusion, we can state that to the extent the whole spectrum of traffic information is integrated in GNSS based applications, possibly as an optional component, the spoken automatic Machine Translation of messages can be successfully deployed in the navigation systems. In this respect the proposed technology can be seen not as a rival to navigation systems, but as complementary technology that can be included in existing navigation systems and make them more useful and adaptable to their user’s needs.

Compatibility with built-in digital radio and TPEG-based systems

Another area to consider is built-in car traffic information systems. Here again we foresee no major problems in deploying this technology. Although it is inherently based on DATEX II as

the conceptual backbone, in so far as its output is digital speech it can be deployed in digital radio and TPEG-based information systems as well.

Conclusions

Customized native spoken language in RTTIS has been neglected so far, although drivers still prefer audio channel and their native language for receiving traffic information. This fact should be given the attention it deserves. The suggested service can provide a solution to this problem in the form of spoken machine translated traffic information in the user preferred language that is pertinent to their GPS position and driving direction. Spoken machine translation applied to the traffic domain is expected to deliver a high quality not otherwise attainable by general purpose statistical machine translation methods. It is suggested that this technology can be deployed further in many different systems, traffic modalities and applications.

The wide-spread deployment of this technology serves to eliminate the current TIS fragmentation, ensuring seamless traffic information service across regions, between EU member states and across language borders, thus overcoming the only remaining borders in the European Union.

References

1. Austroads 2015: Faber, F., Green, D. (2015). *Concept of Operations for C-ITS Core Functions* Austroads. Available at <https://www.onlinepublications.austroads.com.au/items/AP-R479-15> (last visited 12-07-2015)
2. EasyWay (2012a). *Data Exchanges DATEX II Supporting Guideline*. Available at http://www.rits-net.eu/uploads/media/EW-DG-2012_DTX-DG01_DatexII_02-00-00.pdf (last visited 01-08-2015)
3. EasyWay (2012b). *Traveller Information Services Forecast and Real Time Event Information Deployment Guideline*. Available at http://www.rits-net.eu/uploads/media/EW-DG-2012_TIS-DG02_ForecastAndRealTimeEventInformation_02-00-00.pdf (last visited 01-08-2015)
4. EC 2010: European Commission (2010). *Directive 2010/40/EU of the European Parliament and of the Council*. Available at <http://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1438350993825&uri=CELEX:32010L0040> (last visited 01-08-2015)
5. EC 2013: *Commission Delegated Regulation (EU) No 886/2013 of 15 May 2013*. Available at <http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32013R0886> (last visited 01-08-2015)
6. ERTICO 2015a: Task Force on Communication technologies for C-ITS (2015a). *Communication Technologies for future C-ITS service scenarios ERTICO*. Available at http://erticonetwork.wpengine.com/wp-content/uploads/2015/05/images_easyblog_images_1213_Report---Communication-technologies-for-future-C-ITS-v1-1.pdf (last visited 12-07-2015)

7. ERTICO 2015b: Task Force on Communication technologies for C-ITS (2015b). *Guide about technologies for future C-ITS services ERTICO*. Available at http://erticonetwork.wpengine.com/wp-content/uploads/2015/05/images_easyblog_images_1213_Guide-about-technologies-for-future-C-ITS-services-v1-0-2.pdf (last visited 12-07-2015)
8. FOTsis 2015: European Field Operational Test on Safe, Intelligent and Sustainable Highway Operation (2015). *Service 3 – Intelligent congestion control*. Available at <http://www.fotsis.com/index.php/services/intelligent-congestion-control> (last visited 12-07-2015)
9. Freudenstein, J., Cornwel, I. (2014). *Tailoring a reference model for C-ITS architectures and using a DATEX II profile to communicate traffic signal information*, Transport Research Arena 2014, Paris.
10. Gilka, P., Richter, T. (2011). 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.
11. *ITS Deployment Guidelines Library* (2012). Available at <https://dg.easyway-its.eu/DGs2012> (last visited 03-08-2015)
12. SEAMLESS 2011: Seamless traffic data dissemination across urban and inter-urban networks, initiated by ERA-NET ROAD II, *Deliverables 1-4*. Available at http://www.eranetroad.org/index.php?option=com_content&view=article&id=109:2011-mobility&catid=31:standard&Itemid=46#SEAMLESS (last visited 03-08-2015)
13. Thouin, B. (1982). The METEO System. In Lawson, V. (ed.), *Practical Experience of Machine Translation*, North-Holland Publishing Company, pp. 39–44.
14. USDOT (2015). *Connected Vehicle Research in the United States*. Available at http://www.its.dot.gov/connected_vehicle/connected_vehicle_research.htm (last visited 12-07-2015)
15. Váradi, T., Tadić, M., Gulyás, A., Niculescu, M. (2015). *When Will ITS Speak Your Language?* Paper number ITS-2955, 22nd ITS World Congress, Bordeaux, France, 5–9 October 2015.