When will ITS Speak Your Language?

Bringing Multilingual Technologies to CEF Transport to Build Online Digital Services

Tamás Váradi
Research Institute for Linguistics
Hungarian Academy of Sciences
varadi.tamas@nytud.mta.hu

CEF Conference Riga Summit
29th April 2015
Synergies between CEF sectors
Why Intelligent Transport Systems (ITS)?

- position of ICT within H-2020 programme
- ITS combines Transport and ICT
- multilingual issues grossly neglected
  - in practice
  - at policy level
The Problem

- traffic is inherently international
- national borders no longer a barrier within EU
- however, language barriers still prevail
- huge challenge to real-time travel information services (RTIS)
The State of the Art

- RTIS typically provided in the national language of the country
- which is *not* the native language of up to 40% of the drivers within the country
- progress on data exchange between national services
- seamless cross-border service still a long-term objective
A Use Case Scenario

- Bucharest truck driver
- speaks only Romanian
- drives through three countries with no RTIS support in Romanian
ITS Response to the Multilingual Challenge

- the issue of language is grossly underrated
- ITS Deployment Guidelines propose mainly language independent solutions
- *pictograms*, useful but severely limited
- drivers prefer travel information broadcast (i.e. spoken) to them in their native language
MT Response to the Challenge

- Cloud based service
- Native language spoken
- Travel information
- Delivered to mobile devices
- Quality MT combined with speech technologies
- MORENA pilot project
Synergies between ITS and MT

- ITS developed sophisticated data infrastructure
- DATEX 2: data dictionary and protocol
- standardized and widely deployed
<table>
<thead>
<tr>
<th>Enumeration name</th>
<th>Enumeration literal</th>
<th>Designation</th>
<th>Origin</th>
<th>Original code</th>
</tr>
</thead>
<tbody>
<tr>
<td>AccidentTypeEnumeration</td>
<td>accidentInvolvingRadioactiveMaterial</td>
<td>Accident involving radioactive material</td>
<td>- null-</td>
<td>- null-</td>
</tr>
<tr>
<td>AccidentTypeEnumeration</td>
<td>accidentInvolvingTrain</td>
<td>Accident involving train</td>
<td>- null-</td>
<td>- null-</td>
</tr>
<tr>
<td>AccidentTypeEnumeration</td>
<td>chemicalSpillageAccident</td>
<td>Chemical spillage accident</td>
<td>DATEX</td>
<td>ACE</td>
</tr>
<tr>
<td>AccidentTypeEnumeration</td>
<td>collision</td>
<td>Collision</td>
<td>- null-</td>
<td>- null-</td>
</tr>
<tr>
<td>AccidentTypeEnumeration</td>
<td>collisionWithAnimal</td>
<td>Collision with animal</td>
<td>- null-</td>
<td>- null-</td>
</tr>
<tr>
<td>AccidentTypeEnumeration</td>
<td>collisionWithObstruction</td>
<td>Collision with obstruction</td>
<td>- null-</td>
<td>- null-</td>
</tr>
<tr>
<td>AccidentTypeEnumeration</td>
<td>collisionWithPerson</td>
<td>Collision with person</td>
<td>- null-</td>
<td>- null-</td>
</tr>
<tr>
<td>AccidentTypeEnumeration</td>
<td>earlierAccident</td>
<td>Earlier accident</td>
<td>DATEX</td>
<td>ACA</td>
</tr>
<tr>
<td>AccidentTypeEnumeration</td>
<td>fuelSpillageAccident</td>
<td>Fuel spillage accident</td>
<td>DATEX</td>
<td>ACF</td>
</tr>
<tr>
<td>AccidentTypeEnumeration</td>
<td>headOnCollision</td>
<td>Head on collision</td>
<td>- null-</td>
<td>- null-</td>
</tr>
<tr>
<td>AccidentTypeEnumeration</td>
<td>headOnOrSideCollision</td>
<td>Head on or side collision</td>
<td>- null-</td>
<td>- null-</td>
</tr>
<tr>
<td>AccidentTypeEnumeration</td>
<td>jackknifedArticulatedLorry</td>
<td>Jack-knifed articulated lorry</td>
<td>DATEX</td>
<td>AJA</td>
</tr>
<tr>
<td>AccidentTypeEnumeration</td>
<td>jackknifedCaravan</td>
<td>Jack-knifed caravan</td>
<td>DATEX</td>
<td>AJC</td>
</tr>
<tr>
<td>AccidentTypeEnumeration</td>
<td>jackknifedTrailer</td>
<td>Jack-knifed trailer</td>
<td>DATEX</td>
<td>AJT</td>
</tr>
<tr>
<td>AccidentTypeEnumeration</td>
<td>multipleVehicleCollision</td>
<td>Multiple vehicle collision</td>
<td>- null-</td>
<td>- null-</td>
</tr>
<tr>
<td>AccidentTypeEnumeration</td>
<td>multipleVehicleAccident</td>
<td>Multivehicle accident</td>
<td>DATEX</td>
<td>ACM</td>
</tr>
<tr>
<td>AccidentTypeEnumeration</td>
<td>oilSpillageAccident</td>
<td>Oil spillage accident</td>
<td>DATEX</td>
<td>AOI</td>
</tr>
<tr>
<td>AccidentTypeEnumeration</td>
<td>other</td>
<td>Other</td>
<td>- null-</td>
<td>- null-</td>
</tr>
<tr>
<td>AccidentTypeEnumeration</td>
<td>overturnedHeavyLorry</td>
<td>Overturned heavy lorry</td>
<td>DATEX</td>
<td>AOL</td>
</tr>
<tr>
<td>AccidentTypeEnumeration</td>
<td>overturnedTrailer</td>
<td>Overturned trailer</td>
<td>- null-</td>
<td>- null-</td>
</tr>
<tr>
<td>AccidentTypeEnumeration</td>
<td>overturnedVehicle</td>
<td>Overturned vehicle</td>
<td>DATEX</td>
<td>AOV</td>
</tr>
<tr>
<td>AccidentTypeEnumeration</td>
<td>rearCollision</td>
<td>Rear collision</td>
<td>- null-</td>
<td>- null-</td>
</tr>
<tr>
<td>AccidentTypeEnumeration</td>
<td>secondaryAccident</td>
<td>Secondary accident</td>
<td>DATEX</td>
<td>ACD</td>
</tr>
<tr>
<td>AccidentTypeEnumeration</td>
<td>seriousAccident</td>
<td>Serious accident</td>
<td>DATEX</td>
<td>ACS</td>
</tr>
<tr>
<td>AccidentTypeEnumeration</td>
<td>sideCollision</td>
<td>Side collision</td>
<td>- null-</td>
<td>- null-</td>
</tr>
<tr>
<td>AccidentTypeEnumeration</td>
<td>vehicleOffRoad</td>
<td>Vehicle off road</td>
<td>DATEX</td>
<td>ASP</td>
</tr>
<tr>
<td>AccidentTypeEnumeration</td>
<td>vehicleSpunAround</td>
<td>Vehicle spun around</td>
<td>- null-</td>
<td>- null-</td>
</tr>
<tr>
<td>AlertDirectionEnumeration</td>
<td>both</td>
<td>Both</td>
<td>DATEX</td>
<td>B or b</td>
</tr>
<tr>
<td>AlertDirectionEnumeration</td>
<td>negative</td>
<td>Negative</td>
<td>DATEX</td>
<td>N or n</td>
</tr>
<tr>
<td>AlertDirectionEnumeration</td>
<td>positive</td>
<td>Positive</td>
<td>DATEX</td>
<td>P or p</td>
</tr>
<tr>
<td>AlertDirectionEnumeration</td>
<td>unknown</td>
<td>Unknown</td>
<td>DATEX</td>
<td>U or u</td>
</tr>
<tr>
<td>AnimalPresenceTypeEnumeration</td>
<td>animalsOnTheRoad</td>
<td>Animals on the road</td>
<td>DATEX</td>
<td>ANM</td>
</tr>
<tr>
<td>AnimalPresenceTypeEnumeration</td>
<td>herdOfAnimalsOnTheRoad</td>
<td>Herd of animals on the road</td>
<td>DATEX</td>
<td>ANH</td>
</tr>
<tr>
<td>AnimalPresenceTypeEnumeration</td>
<td>largeAnimalsOnTheRoad</td>
<td>Large animals on the road</td>
<td>DATEX</td>
<td>ANL</td>
</tr>
</tbody>
</table>
Synergies between ITS and MT

- ITS developed sophisticated data infrastructure
- DATEX 2: data dictionary and protocol
- standardized and widely deployed
- MORENA uses DATEX 2 as core component
- full-scale solution: covers the full cycle from event reporting to traffic messages
Innovative ITS aspects

- bringing travel information to drivers in their most preferred manner
- maximising outreach to drivers within their country
- providing timely and customised service to drivers
- potential to fully automate their RTIS even in the national language
Language Technology Approach

● integrates best practices and proven concepts in
  o terminology management, machine translation, translation memories, natural controlled language systems and speech technology

● synergies with ITS allow high-quality MT owing to standard, well-defined ontology infrastructure (Datex 2)

● MT optimally conforms to the data and workflow of the application domain
Language Technology Approach
Language Technology Approach

preprocessing including ASR
Language Technology Approach

 preprocessing including ASR

 conversion to DATEX2 multilingual IE
Language Technology Approach

-preprocessing including ASR
- conversion to DATEX2
- multilingual IE
Language Technology Approach

preprocessing including ASR

conversion to DATEX2
multilingual IE

automatic language generation
Language Technology Approach

- preprocessing including ASR
- conversion to DATEX2 multilingual IE
- automatic language generation
- Text-to-Speech generation
Language Technology Approach

preprocessing including ASR

conversion to DATEX2 multilingual IE

manual aggregation of information

manual writing of messages

automatic language generation

Text-to-Speech generation

interlingua-based machine translation
Language Technology Approach

Morena service

Mobile apps
Conclusions

- transdisciplinary effort, requires cooperation of ITS stakeholders
- paradigm shift needed in ITS policy vis-à-vis language
- LT approach built of proven components and maximizes the synergy with the application domain
- usable in other areas of ITS and in other sectors and domains
Conclusions

- current pilot requires scaling up to European level
- a pan-European online digital service is born!
- a candidate for a future DSI
Thank you for your attention!