When will ITS Speak Your Language?

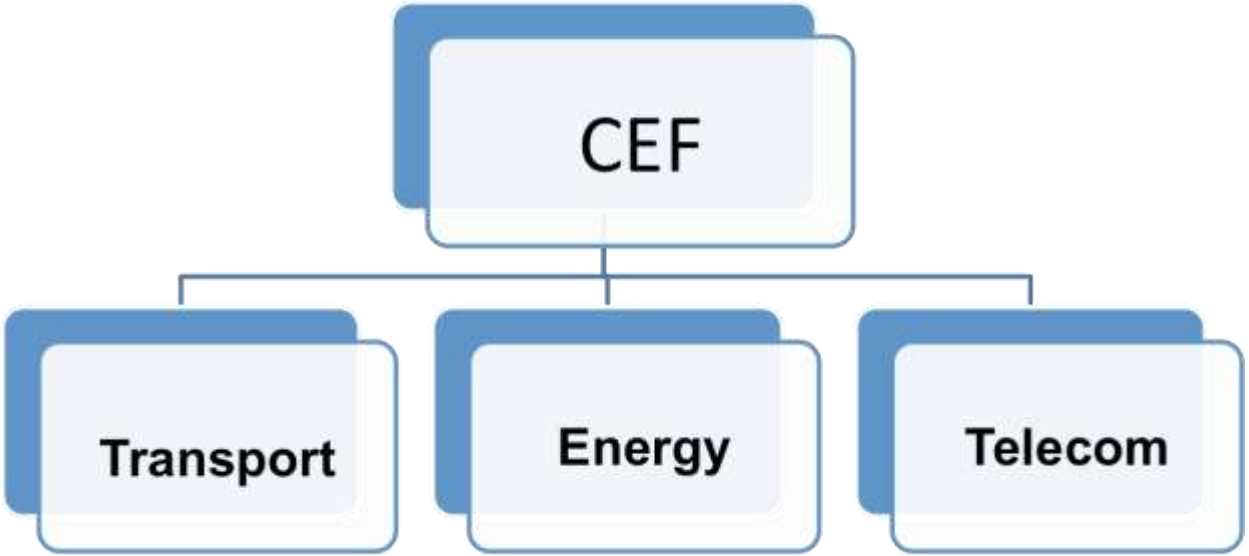
Bringing Multilingual Technologies to CEF Transport to Build Online Digital Services

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

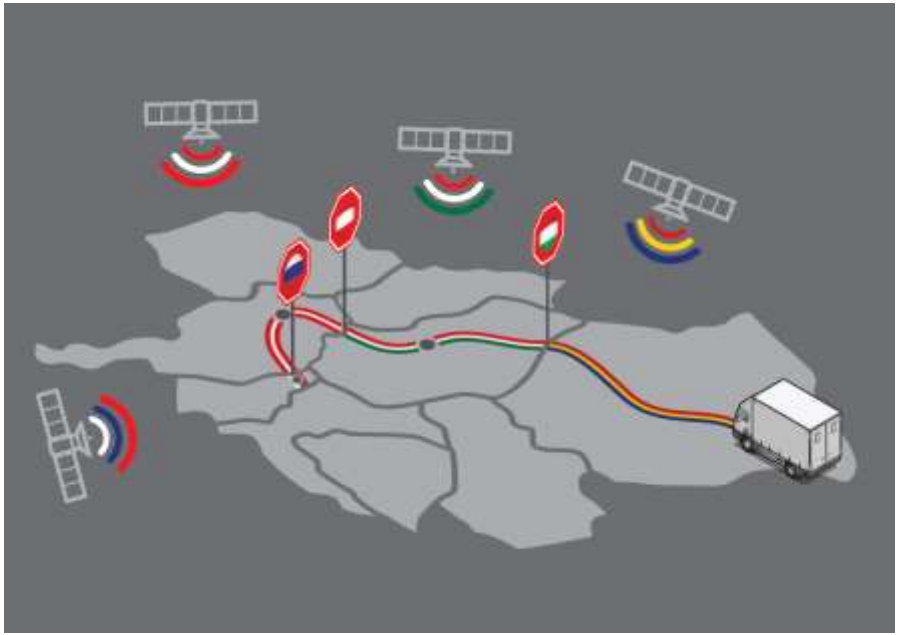
- traffic is inherently international
- national borders no longer a barrier within EU
- however, language barriers still prevail
- huge challenge to real-time travel information services (RTIS)

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

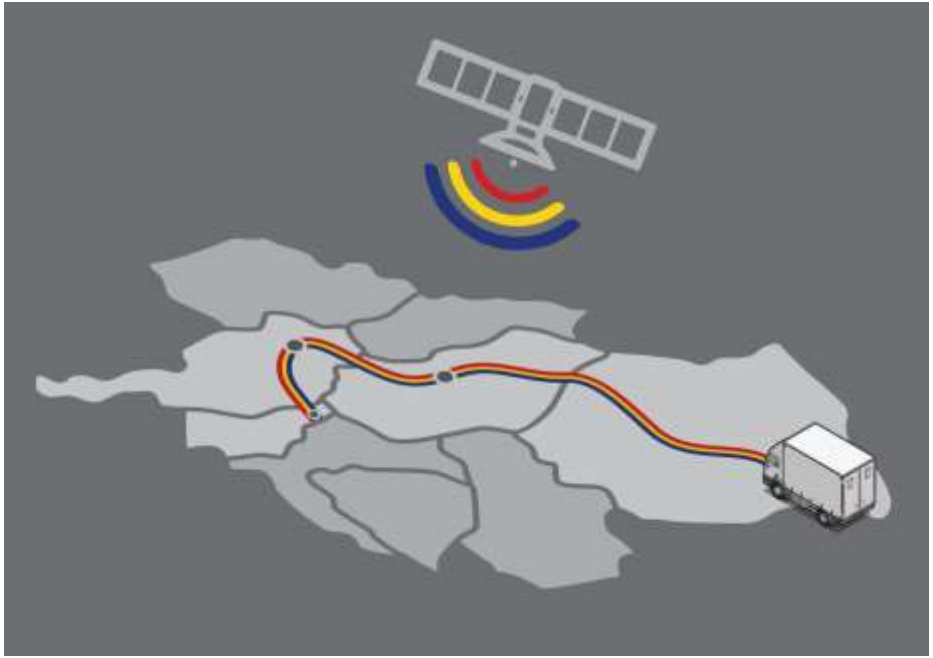


ITS Response to the Multilingual Challenge

- 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

DATEX2 sample

Enumeration name	Enumeration litteral	Designation	Origin	Original code
AccidentTypeEnum	accidentInvolvingRadioactiveMaterial	Accident involving radioactive material	- null -	- null -
AccidentTypeEnum	accidentInvolvingTrain	Accident involving train	- null -	- null -
AccidentTypeEnum	chemicalSpillageAccident	Chemical spillage accident	DATEX	ACE
AccidentTypeEnum	collision	Collision	- null -	- null -
AccidentTypeEnum	collisionWithAnimal	Collision with animal	- null -	- null -
AccidentTypeEnum	collisionWithObstruction	Collision with obstruction	- null -	- null -
AccidentTypeEnum	collisionWithPerson	Collision with person	- null -	- null -
AccidentTypeEnum	earlierAccident	Earlier accident	DATEX	ACA
AccidentTypeEnum	fuelSpillageAccident	Fuel spillage accident	DATEX	ACF
AccidentTypeEnum	headOnCollision	Head on collision	- null -	- null -
AccidentTypeEnum	headOnOrSideCollision	Head on or side collision	- null -	- null -
AccidentTypeEnum	jackknifedArticulatedLorry	Jack-knifed articulated lorry	DATEX	AJA
AccidentTypeEnum	jackknifedCaravan	Jack-knifed caravan	DATEX	AJC
AccidentTypeEnum	jackknifedTrailer	Jack-knifed trailer	DATEX	AJT
AccidentTypeEnum	multipleVehicleCollision	Multiple vehicle collision	- null -	- null -
AccidentTypeEnum	multivehicleAccident	Multivehicle accident	DATEX	ACM
AccidentTypeEnum	oilSpillageAccident	Oil spillage accident	DATEX	AOI
AccidentTypeEnum	other	Other	- null -	- null -
AccidentTypeEnum	overturnedHeavyLorry	Overturned heavy lorry	DATEX	AOL
AccidentTypeEnum	overturnedTrailer	Overturned trailer	- null -	- null -
AccidentTypeEnum	overturnedVehicle	Overturned vehicle	DATEX	AOV
AccidentTypeEnum	rearCollision	Rear collision	- null -	- null -
AccidentTypeEnum	secondaryAccident	Secondary accident	DATEX	ACD
AccidentTypeEnum	seriousAccident	Serious accident	DATEX	ACS
AccidentTypeEnum	sideCollision	Side collision	- null -	- null -
AccidentTypeEnum	vehicleOffRoad	Vehicle off road	- null -	- null -
AccidentTypeEnum	vehicleSpunAround	Vehicle spun around	DATEX	ASP
AlertCDirectionEnum	both	Both	DATEX	B or b
AlertCDirectionEnum	negative	Negative	DATEX	N or n
AlertCDirectionEnum	positive	Positive	DATEX	P or p
AlertCDirectionEnum	unknown	Unknown	DATEX	U or u
AnimalPresenceTypeEnum	animalsOnTheRoad	Animals on the road	DATEX	ANM
AnimalPresenceTypeEnum	herdOfAnimalsOnTheRoad	Herd of animals on the road	DATEX	ANH
AnimalPresenceTypeEnum	largeAnimalsOnTheRoad	Large animals on the road	DATEX	ANL

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

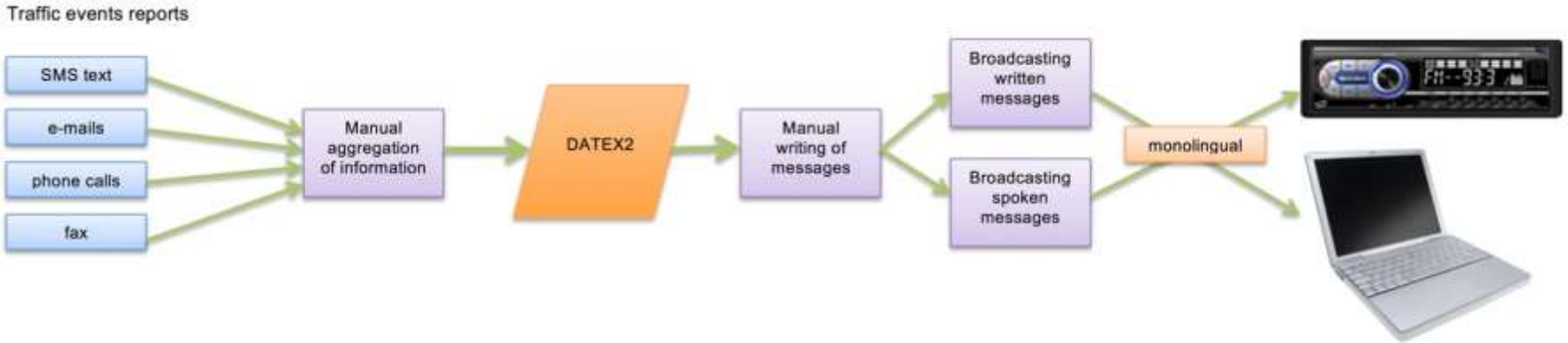
Innovative ITS aspects

- 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

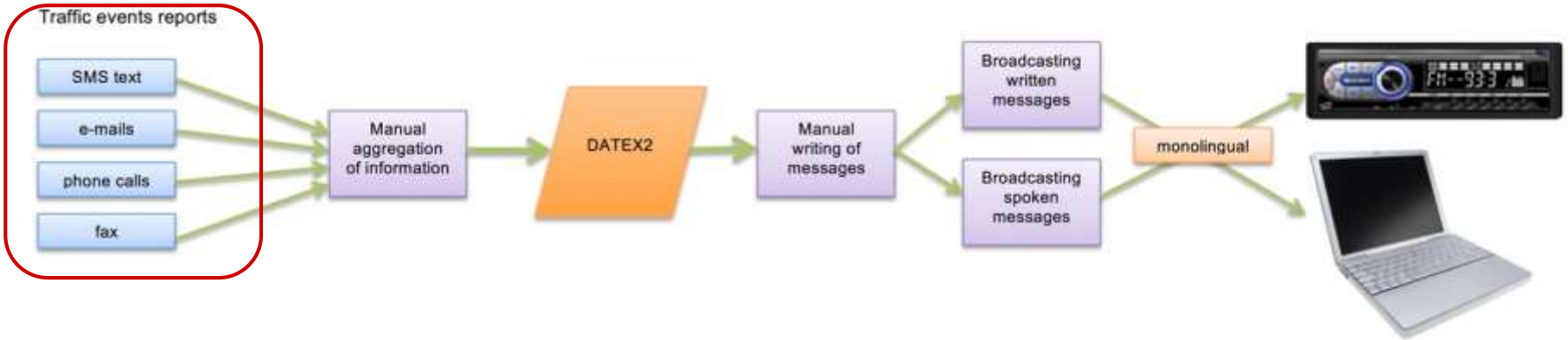
Language Technology Approach

- integrates best practices and proven concepts in
  - 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

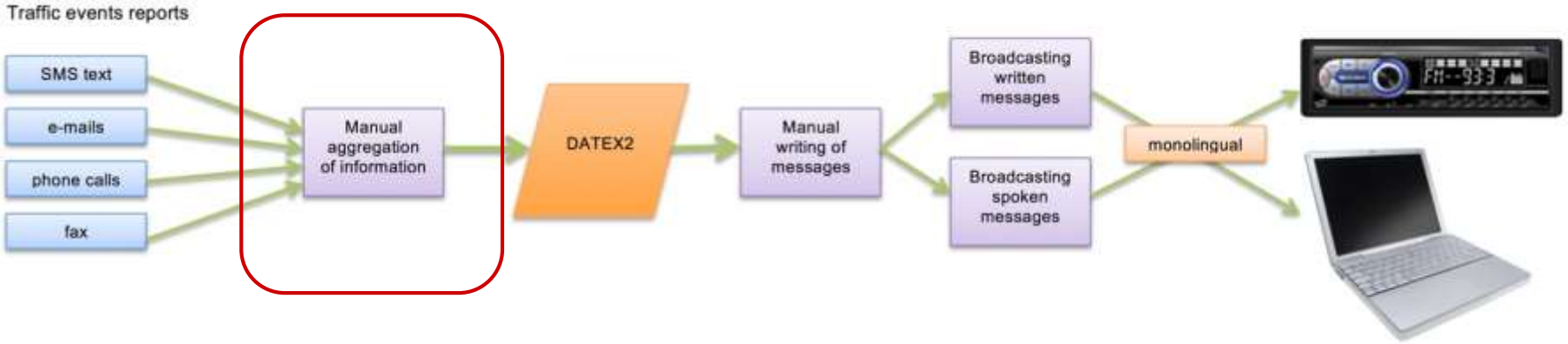


Language Technology Approach



prepocessing including ASR

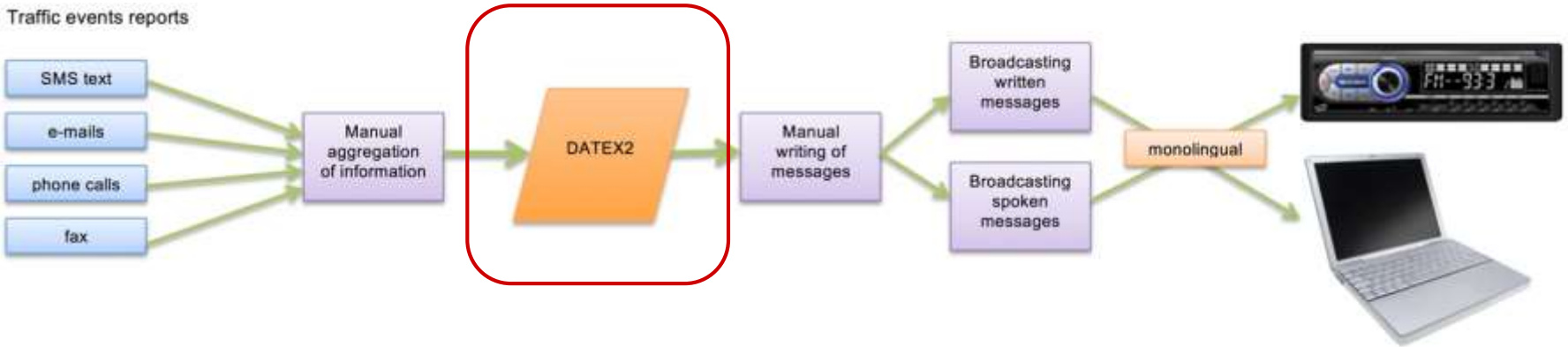
Language Technology Approach



prepocessing including ASR

conversion to DATEX2 multilingual IE

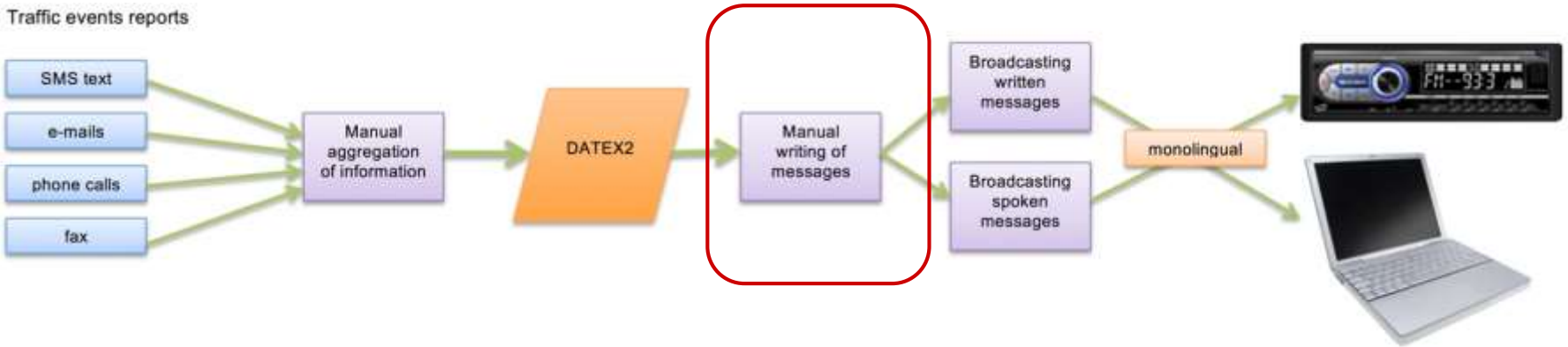
Language Technology Approach



prepocessing including ASR

conversion to DATEX2 multilingual IE

Language Technology Approach

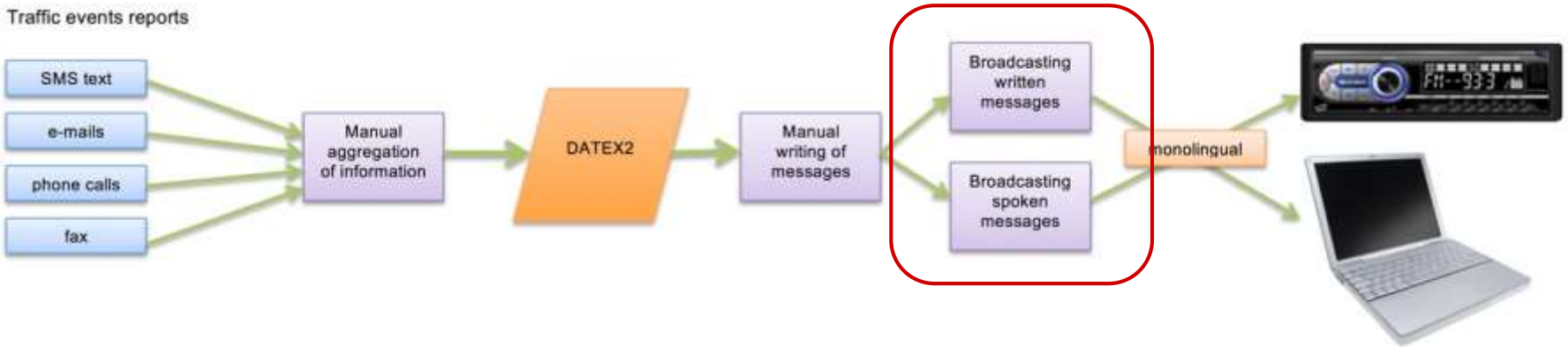


prepocessing including ASR

conversion to DATEX2 multilingual IE

automatic language generation

Language Technology Approach



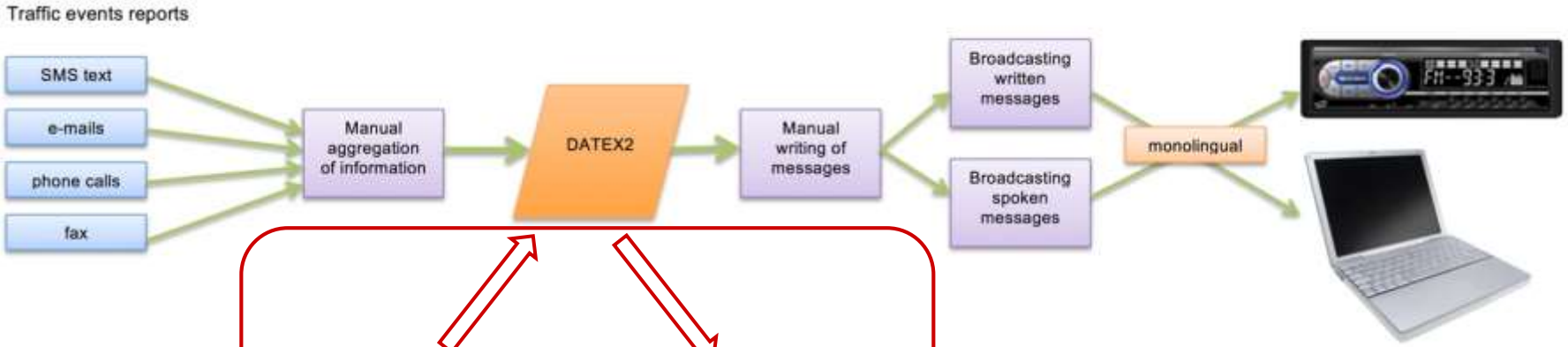
prepocessing including ASR

conversion to DATEX2 multilingual IE

automatic language generation

Text-to-Speech generation

Language Technology Approach



prepocessing including ASR

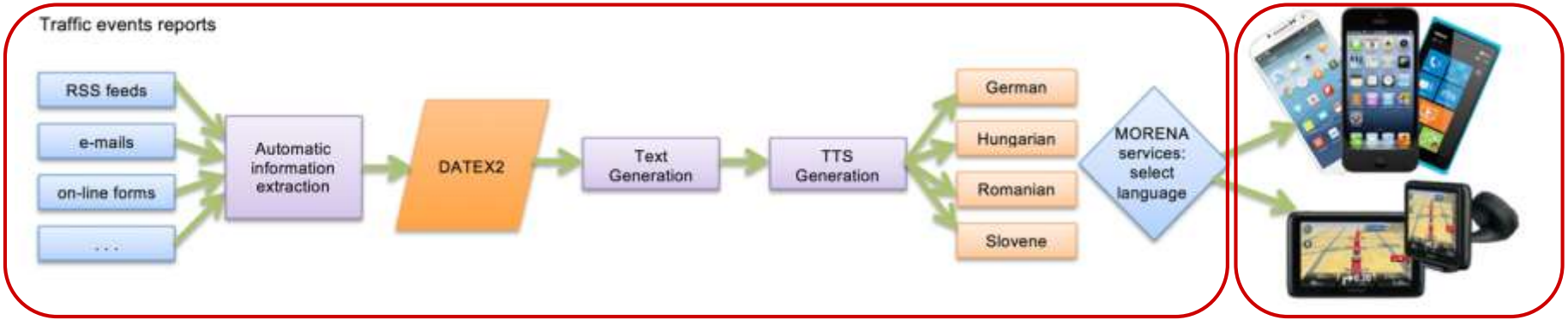
conversion to DATEX2 multilingual IE

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!

MORENA
Multilingual Mobile Regional Road Traffic Information Service

www.morena-project.eu