The META-NET Strategic Research Agenda for Multilingual Europe

Georg Rehm

Network Manager META-NET
DFKI (German Research Center for Artificial Intelligence), Berlin, Germany

georg.rehm@dfki.de

The Hungarian Language in the Digital Age – Budapest, Hungary
January 18, 2013

Co-funded by the 7th Framework Programme and the ICT Policy Support Programme of the European Commission through the contracts T4ME, CESAR, METANET4U, META-NORD (grant agreements no. 249119, 271022, 270893, 270899).

Outline

- Introduction
- META-SHARE: An Infrastructure for Research & Innovation
- Language White Paper Series: Europe's Languages in the Digital Age
- The META-NET Strategic Research Agenda for Multilingual Europe
- Conclusions – Next Steps

Multilingual Europe

META-NET

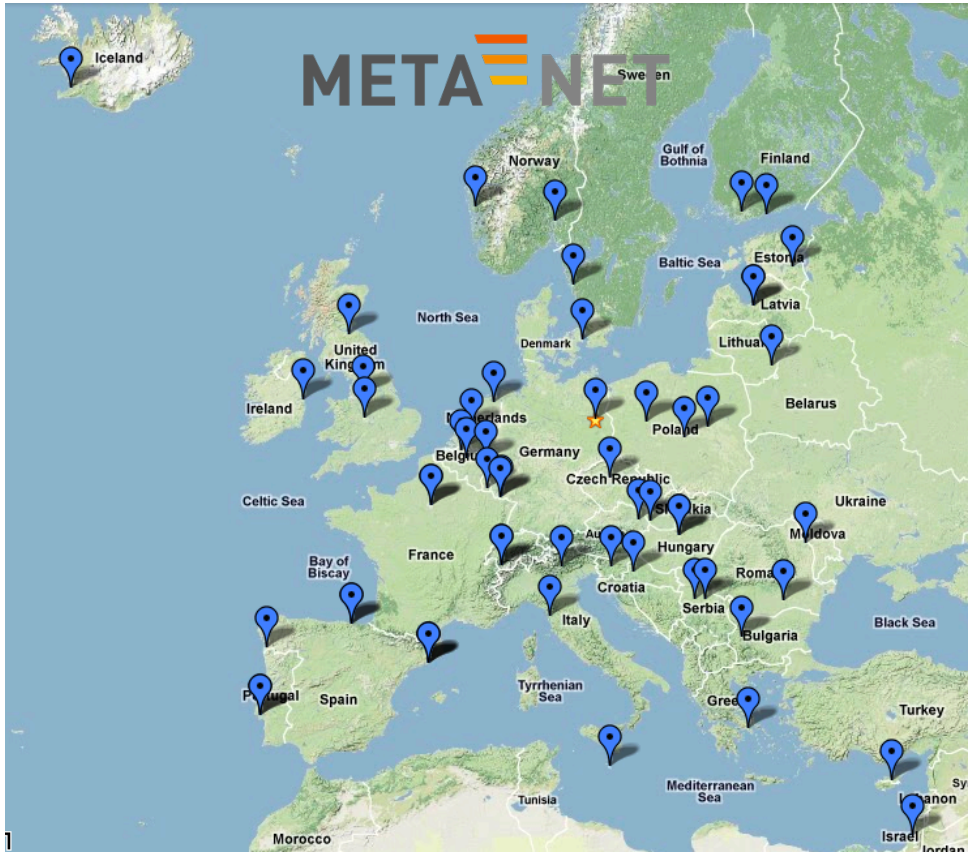
- **Challenge:** Providing each language community with the most advanced technologies for communication and information so that maintaining their mother tongue does not turn into a disadvantage.
- While research has made considerable progress in recent years, the pace of progress is not fast enough to meet the challenge within the next 10-20 years.
- All stakeholders – researchers, LT user and provider industries, language communities, funding programmes, policy makers – should **team up for a major dedicated push**.

Objectives

META-NET is a network of excellence dedicated to fostering the technological foundations of the European multilingual information society.

Four EU-Funded Projects

- Initial project: T4ME (FP7; 13 partners, 10 countries)
- Three ICT-PSP consortia since Feb. 2011: CESAR, METANET4U, META-NORD
- All EU member states and several non-member states covered.
- META-NET in Jan. 2013: **60** members in **34** countries.



http://www.meta-net.eu/members

META-NET

META-SHARE

META-SHARE at a Glance

- Open exchange infrastructure for language resources and tools.
- Language resources and tools are documented, uploaded, stored in repositories, catalogued, can be downloaded, shared, discussed.
- Improve their visibility, documentation, identification, availability, preservation, interoperability.
- Long-term goal: boost research, technology and innovation through wide availability, pooling, openness and sharing of resources.
- Repositories store and maintain inventories of resources and tools.
- Metadata inventories are exported and harvested in the network.
- Currently 19 repositories up and running; ~2.000 LRs available.

META-SHARE portal

Registration - Authentication - Authorisation

Search / Browse | License | Download | User support

Mappings | Reporting/Statistics | Recommendations | Billing / Payment

META-SHARE inventory | META-SHARE inventory | META-SHARE inventory

External repos

Metadata harvesting

Inventory LR repo | Inventory LR repo | Inventory LR repo | Inventory LR repo

META SHARE Editor Profile Logout

Username: t4me-admin

META SHARE

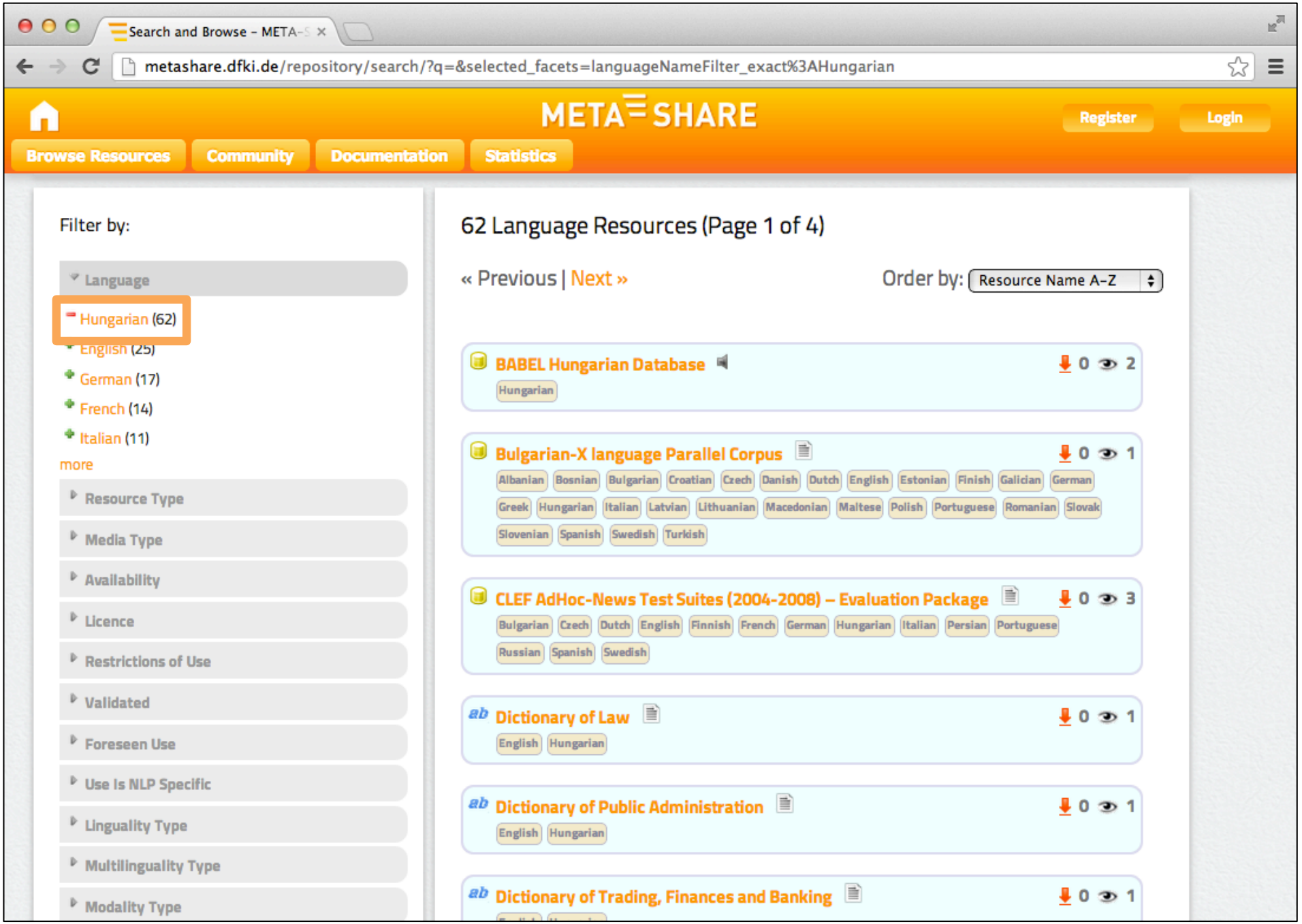
1,248 language resources at your disposal...

Type in your keywords, please...

Search Browse Statistics

What is it? - About the project

META-NET aims at creating META-SHARE, a sustainable network of repositories of language data, tools and related web services documented with high-quality metadata, aggregated in central inventories allowing for uniform search and access to resources. Data and tools can be both open and with restricted access rights, free

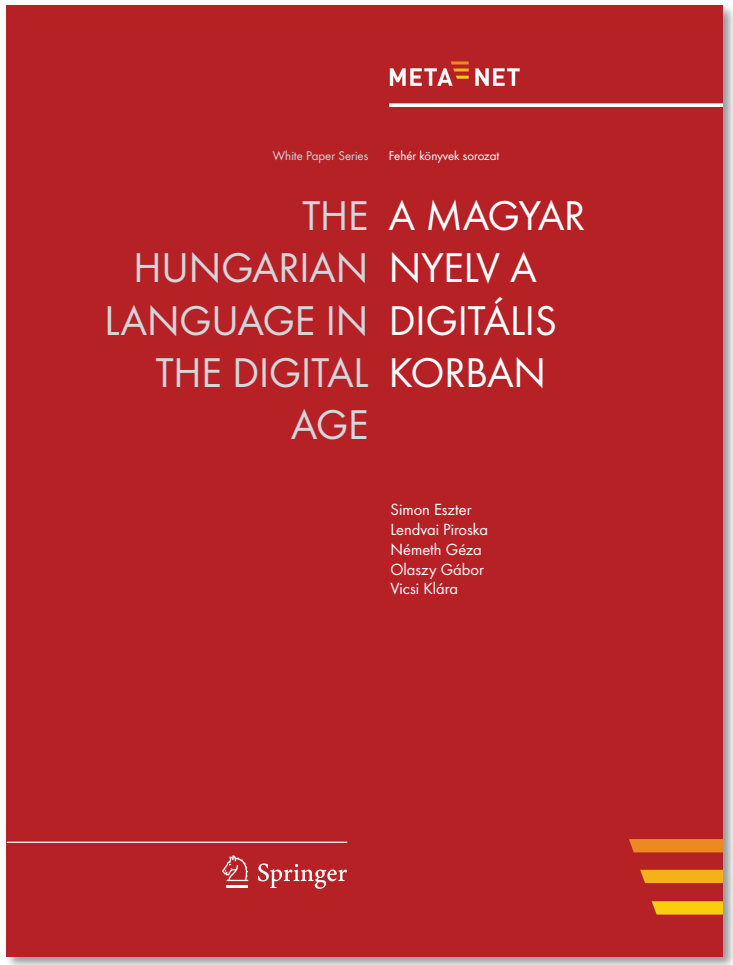

Search and Browse - META-S
metashare.dfki.de/repository/search/?q=&selected_facets=languageNameFilter_exact%3AHungarian
META SHARE
Register
Login
Browse Resources
Community
Documentation
Statistics
Filter by:
Language
Hungarian (62)
English (25)
German (17)
French (14)
Italian (11)
more
Resource Type
Media Type
Availability
Licence
Restrictions of Use
Validated
Foreseen Use
Use Is NLP Specific
Linguality Type
Multilinguality Type
Modality Type
62 Language Resources (Page 1 of 4)
« Previous | Next »
Order by: Resource Name A-Z
BABEL Hungarian Database
0 2
Hungarian
Bulgarian-X language Parallel Corpus
0 1
Albanian Bosnian Bulgarian Croatian Czech Danish Dutch English Estonian Finish Galician German Greek Hungarian Italian Latvian Lithuanian Macedonian Maltese Polish Portuguese Romanian Slovak Slovenian Spanish Swedish Turkish
CLEF AdHoc-News Test Suites (2004-2008) – Evaluation Package
0 3
Bulgarian Czech Dutch English Finnish French German Hungarian Italian Persian Portuguese Russian Spanish Swedish
Dictionary of Law
0 1
English Hungarian
Dictionary of Public Administration
0 1
English Hungarian
Dictionary of Trading, Finances and Banking
0 1

META-VISION

Language White Paper Series

Language White Paper Series

- "Europe's Languages in the Digital Age".
- Reports on the state of our languages in the digital age and the level of support through language technology.
- Series covers 30 languages.
- Key communication instruments to address decision makers and journalists.
- Inform about societal and technological problems and challenges as well as economic opportunities.
- >2 years in the making.
- >200 national experts as contributors.
- >8.000 copies printed and distributed to politicians and journalists.



MT

excellent	good	moderate	fragmentary	weak or no support
	English	French, Spanish	Catalan, Dutch, German, **Hungarian**, Italian, Polish, Romanian	Basque, Bulgarian, Croatian, Czech, Danish, Estonian, Finnish, Galician, Greek, Icelandic, Irish, Latvian, Lithuanian, Maltese, Norwegian, Portuguese, Serbian, Slovak, Slovene, Swedish

Text Analysis

excellent	good	moderate	fragmentary	weak or no support
	English	Dutch, French, German, Italian, Spanish	Basque, Bulgarian, Catalan, Czech, Danish, Finnish, Galician, Greek, **Hungarian**, Norwegian, Polish, Portuguese, Romanian, Slovak, Slovene, Swedish	Croatian, Estonian, Icelandic, Irish, Latvian, Lithuanian, Maltese, Serbian

Speech

excellent	good	moderate	fragmentary	weak or no support
	English	Czech, Dutch, Finnish, French, German, Italian, Portuguese, Spanish	Basque, Bulgarian, Catalan, Danish, Estonian, Galician, Greek, **Hungarian**, Irish, Norwegian, Polish, Serbian, Slovak, Slovene, Swedish	Croatian, Icelandic, Latvian, Lithuanian, Maltese, Romanian

Resources

excellent	good	moderate	fragmentary	weak/no support
	English	Czech, Dutch, French, German, **Hungarian**, Italian, Polish, Spanish, Swedish	Basque, Bulgarian, Catalan, Croatian, Danish, Estonian, Finnish, Galician, Greek, Norwegian, Portuguese, Romanian, Serbian, Slovak, Slovene	Icelandic, Irish, Latvian, Lithuanian, Maltese

Press Campaign

- Headline of press release:

  At Least 21 European Languages in Danger of Digital Extinction.

- Sent out to journalists, politicians and other stakeholder groups on the European Day of Languages (Sept. 26, 2012).
- Overwhelmed by the huge interest in the topic and our key findings!
- 600+ mentions in the press.
- 50+ interviews with META-NET representatives (television, radio).
- News came in from 40+ countries in 35+ different languages.
- City with the most visitors of the META-NET website: *Brussels!*
- Two Parliamentary Questions in the European Parliament on the "digital extinction of languages" topic.

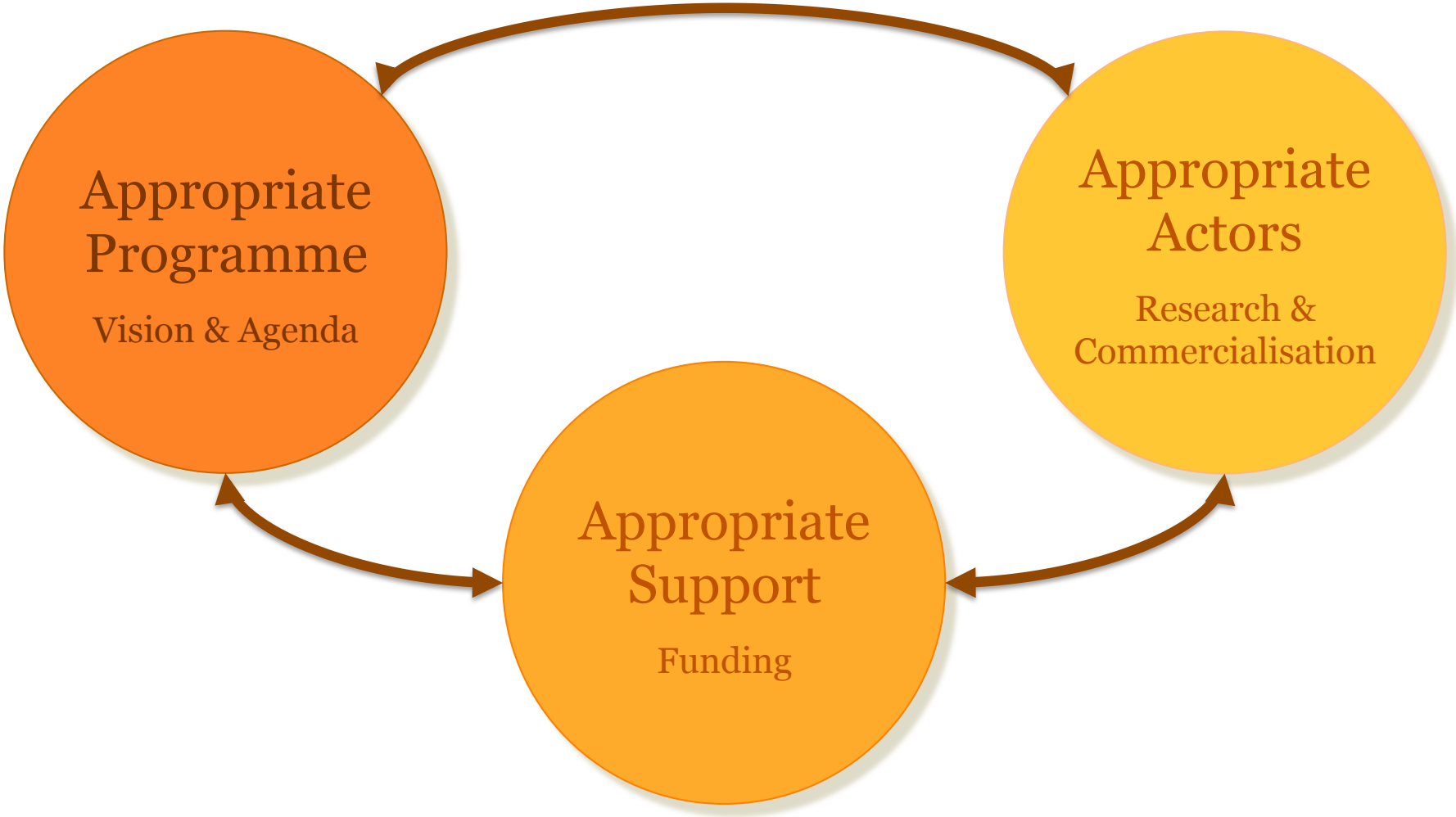
Response: Examples

- **Austria:** Der Standard.
- **Denmark:** Politiken, Berlingske Tidende.
- **Finland:** Tiede.
- **Germany:** Heise Newsticker, Süddeutsche Zeitung.
- **Greece:** in.gr, Πρώτο Θέμα, Prosilipsis.
- **Hungary:** Origo.
- **Iceland:** Fréttablaðið, Morgunblaðið.
- **Italy:** Wired.
- **Norway:** Computerworld.
- **Slovenia:** Delo, Dnevnik, Demokracija.
- **Serbia:** Politika.
- **Spain:** El Mundo.
- **UK:** Huffington Post.
- **USA:** Mashable, NBC News, Reddit.

META-VISION

Strategic Research Agenda

Three Ingredients



Strategic Research Agenda

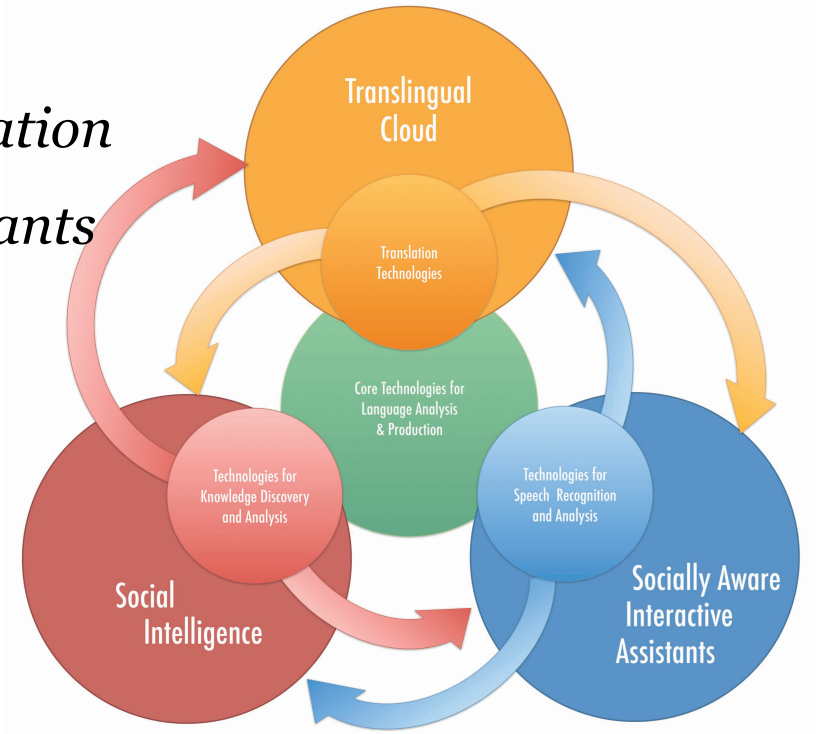
- Addresses the problems we found when preparing the white papers.
- Three priority research themes and application/innovation scenarios.
- Can put Europe ahead of its competitors in this technology area.
- >190 contributors; >2 years.
- Presented and discussed at 83 conferences and major workshops.
- Final version ready on Dec. 1, 2012.
- http://www.meta-net.eu/sra

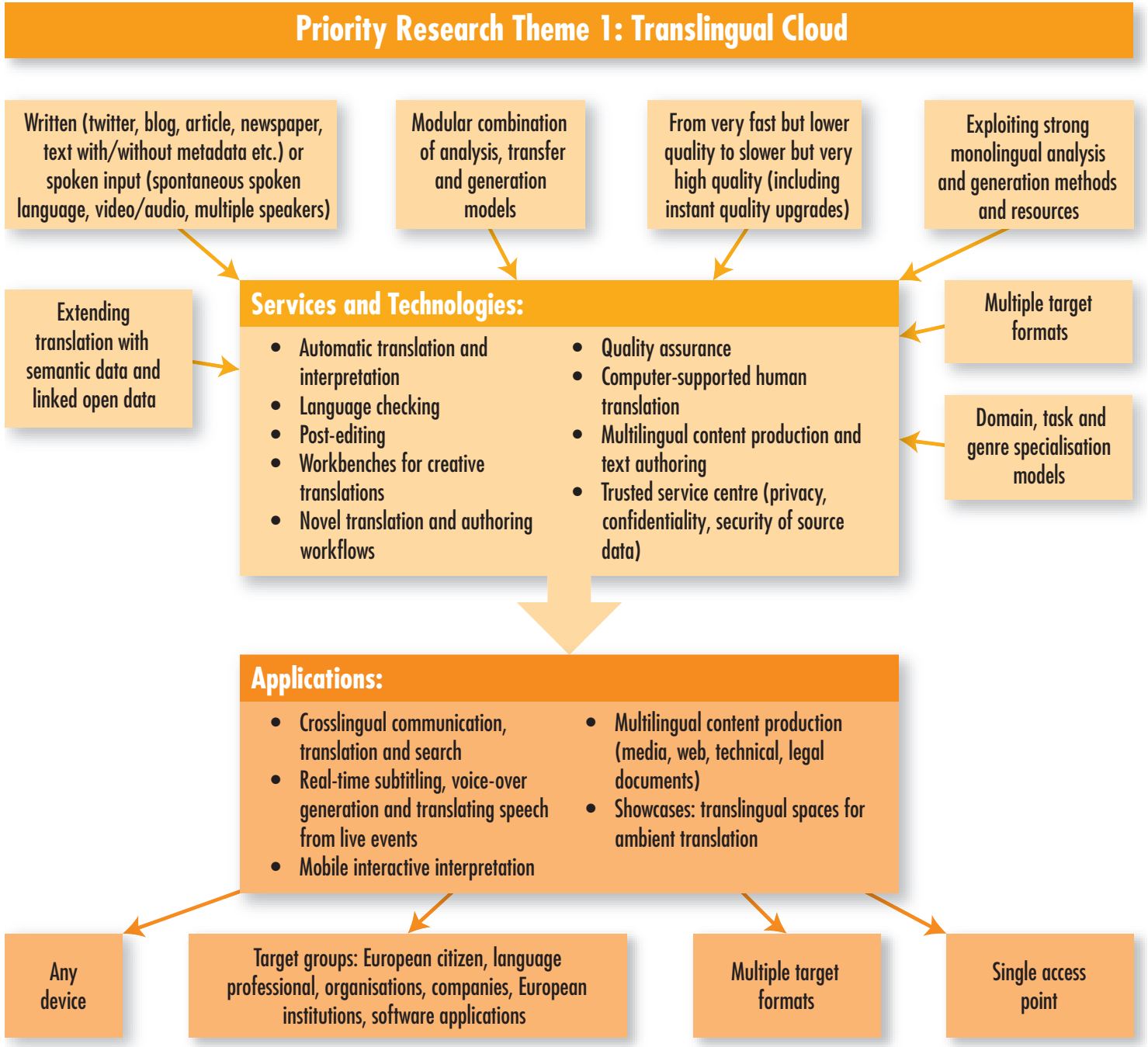
SRA: Contents – Brief Glimpse

- Set the stage and describe the European situation, the needs and the LT research and industry.
- Discuss the state of IT, predictions and mega-trends.
- Our technology vision for 2020.
- Select and specify priority themes.
- Suggest a model for speeding up innovation.
- Outline proposals for the organisation of research and innovation.

Priority Themes: 3 + 2

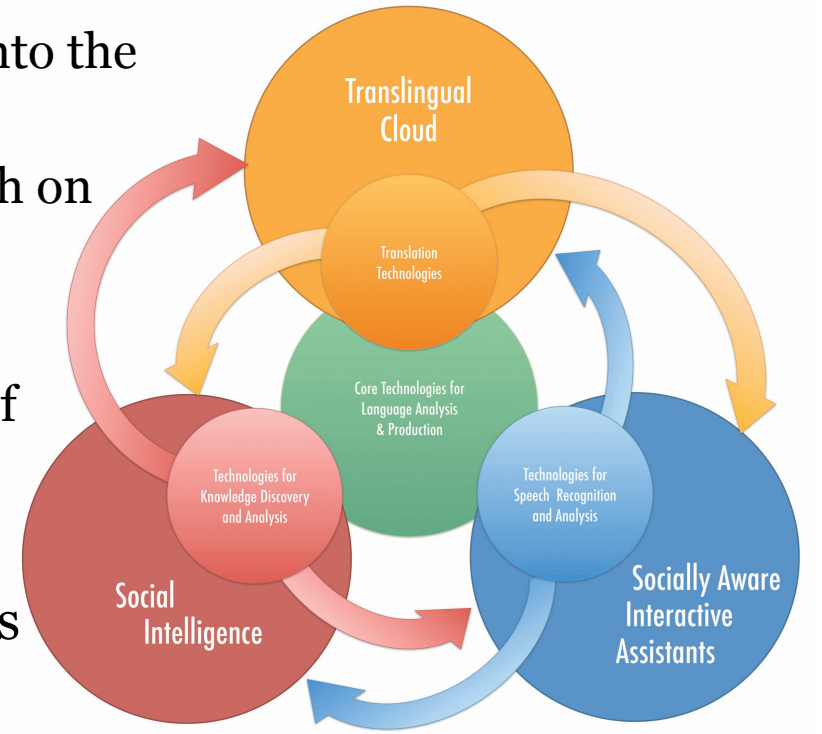
- We decided on priority themes that (a) support technology progress, (b) lead to solutions that European society needs and (c) solutions from which European industry will benefit as users or as providers.
  - *Translingual Cloud*
  - *Social Intelligence and e-Participation*
  - *Socially-Aware Interactive Assistants*
- Two additional themes:
  - *European Service Platform for Language Technologies*
  - *Core Technologies for Language Analysis and Production*

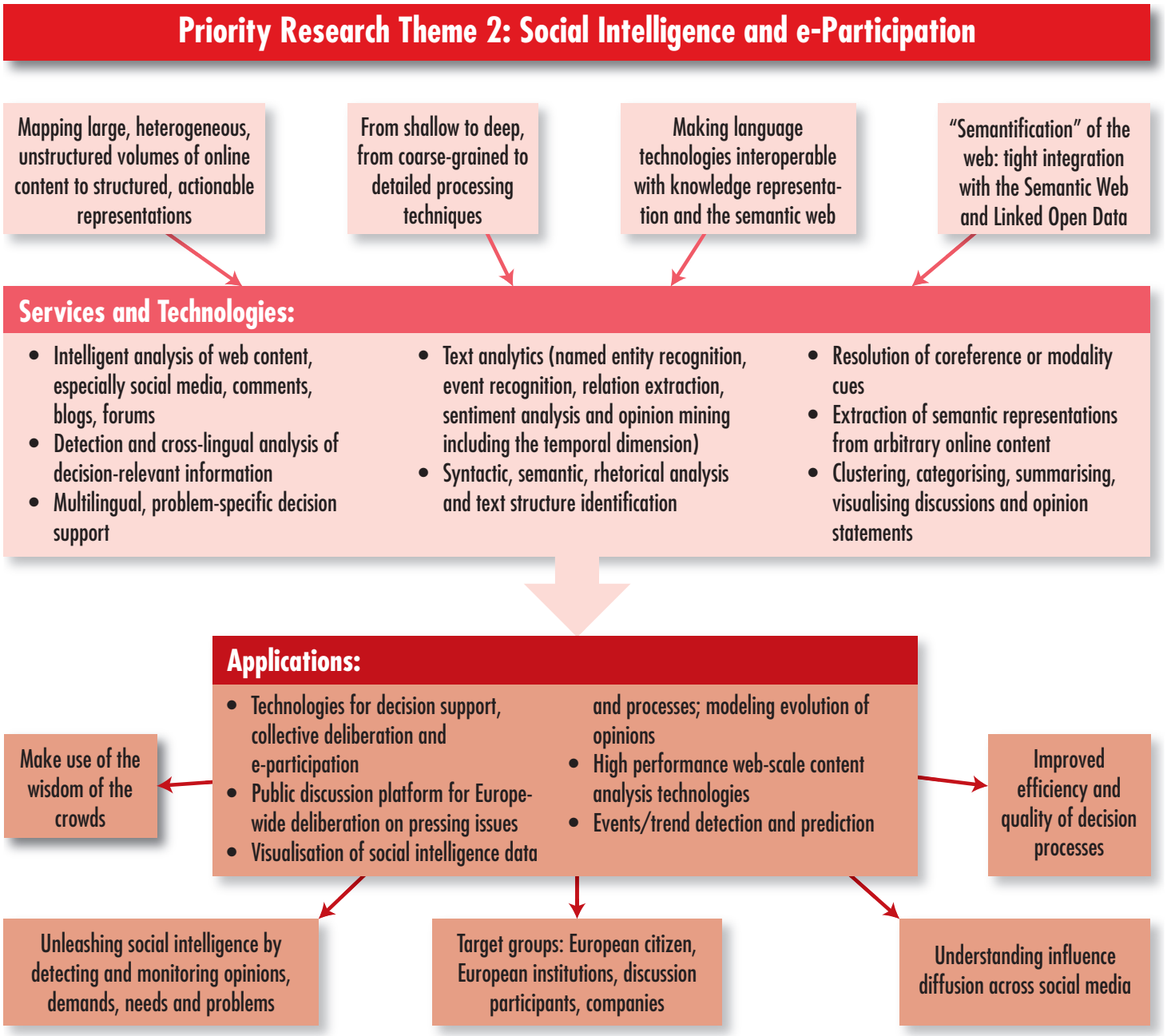



Priority Research Theme 1: Translingual Cloud
Written (twitter, blog, article, newspaper, text with/without metadata etc.) or spoken input (spontaneous spoken language, video/audio, multiple speakers)
Modular combination of analysis, transfer and generation models
From very fast but lower quality to slower but very high quality (including instant quality upgrades)
Exploiting strong monolingual analysis and generation methods and resources
Extending translation with semantic data and linked open data
Services and Technologies:
Automatic translation and interpretation
Language checking
Post-editing
Workbenches for creative translations
Novel translation and authoring workflows
Quality assurance
Computer-supported human translation
Multilingual content production and text authoring
Trusted service centre (privacy, confidentiality, security of source data)
Multiple target formats
Domain, task and genre specialisation models
Applications:
Crosslingual communication, translation and search
Real-time subtitling, voice-over generation and translating speech from live events
Mobile interactive interpretation
Multilingual content production (media, web, technical, legal documents)
Showcases: translingual spaces for ambient translation
Any device
Target groups: European citizen, language professional, organisations, companies, European institutions, software applications
Multiple target formats
Single access point

PT1: Translingual Cloud

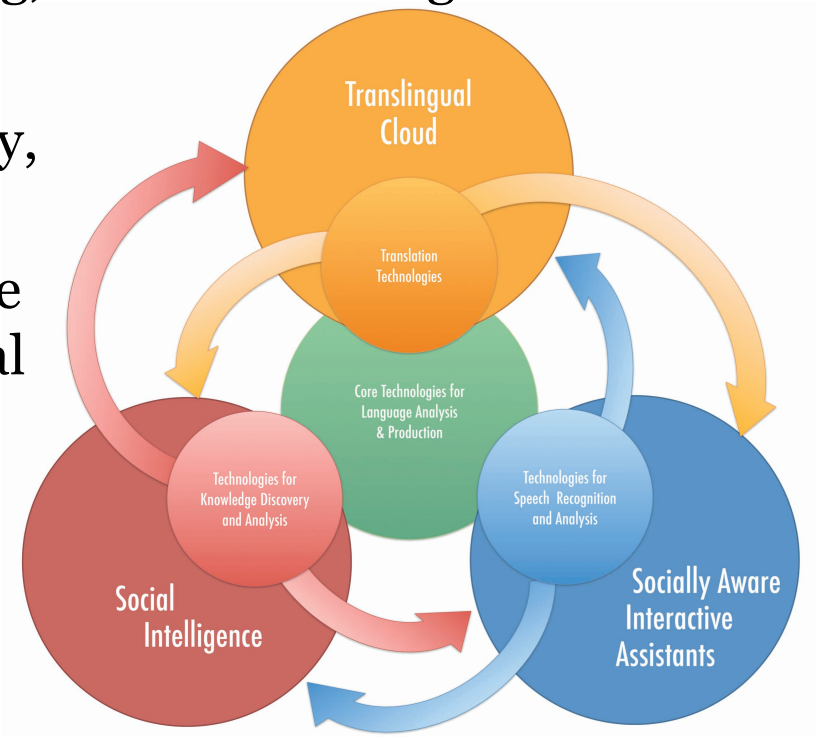
- Europe has a big need for translations of publishable quality.
- Focus on high-quality translation.
- New research paradigms
  - Inclusion of professional translators into the research process
  - Inclusion of technologists into research on human translation processes
- Different technological approaches
  - Stronger emphasis on the properties of individual languages
  - A central role for semantics
- Methods for specific genres & domains

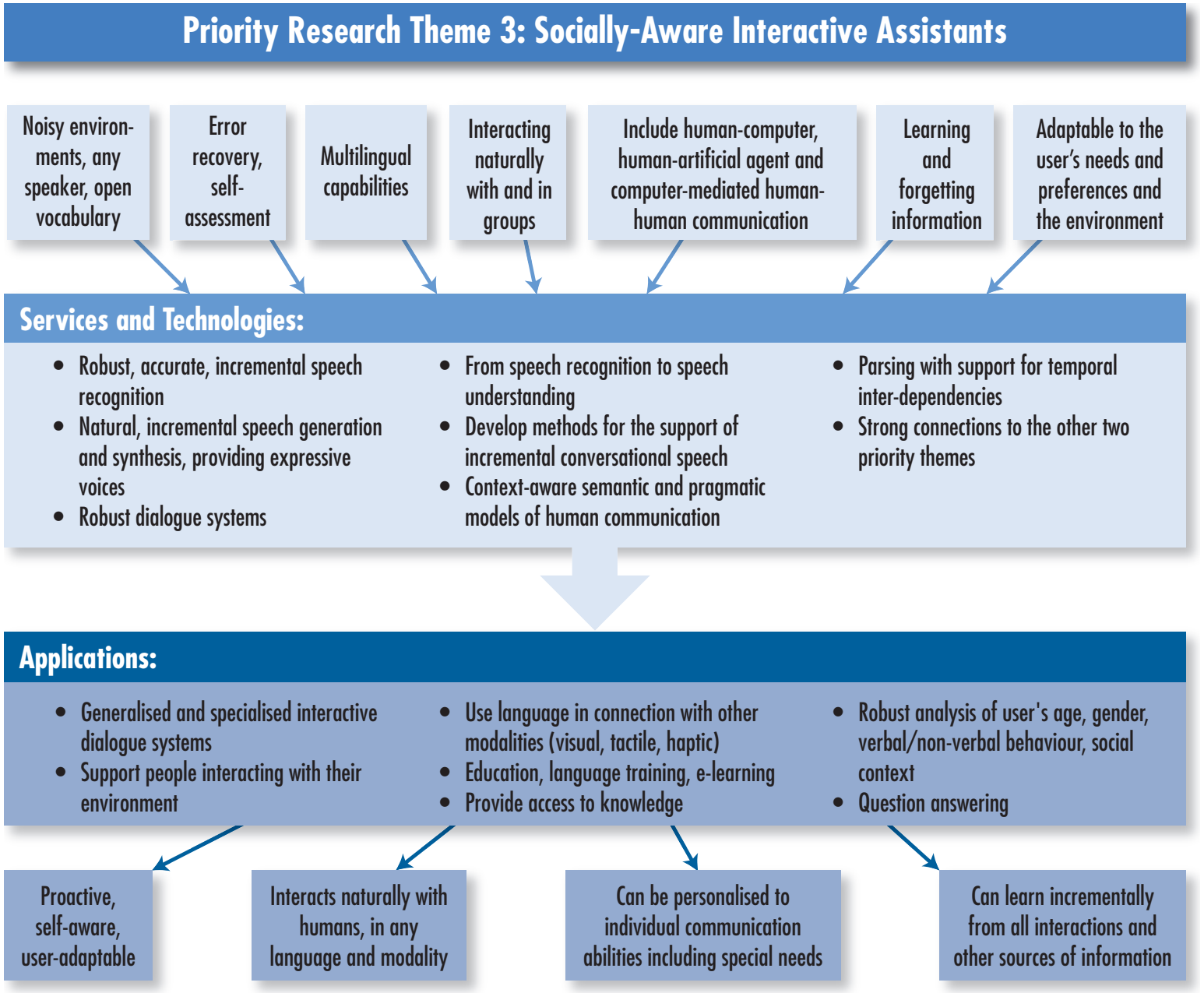


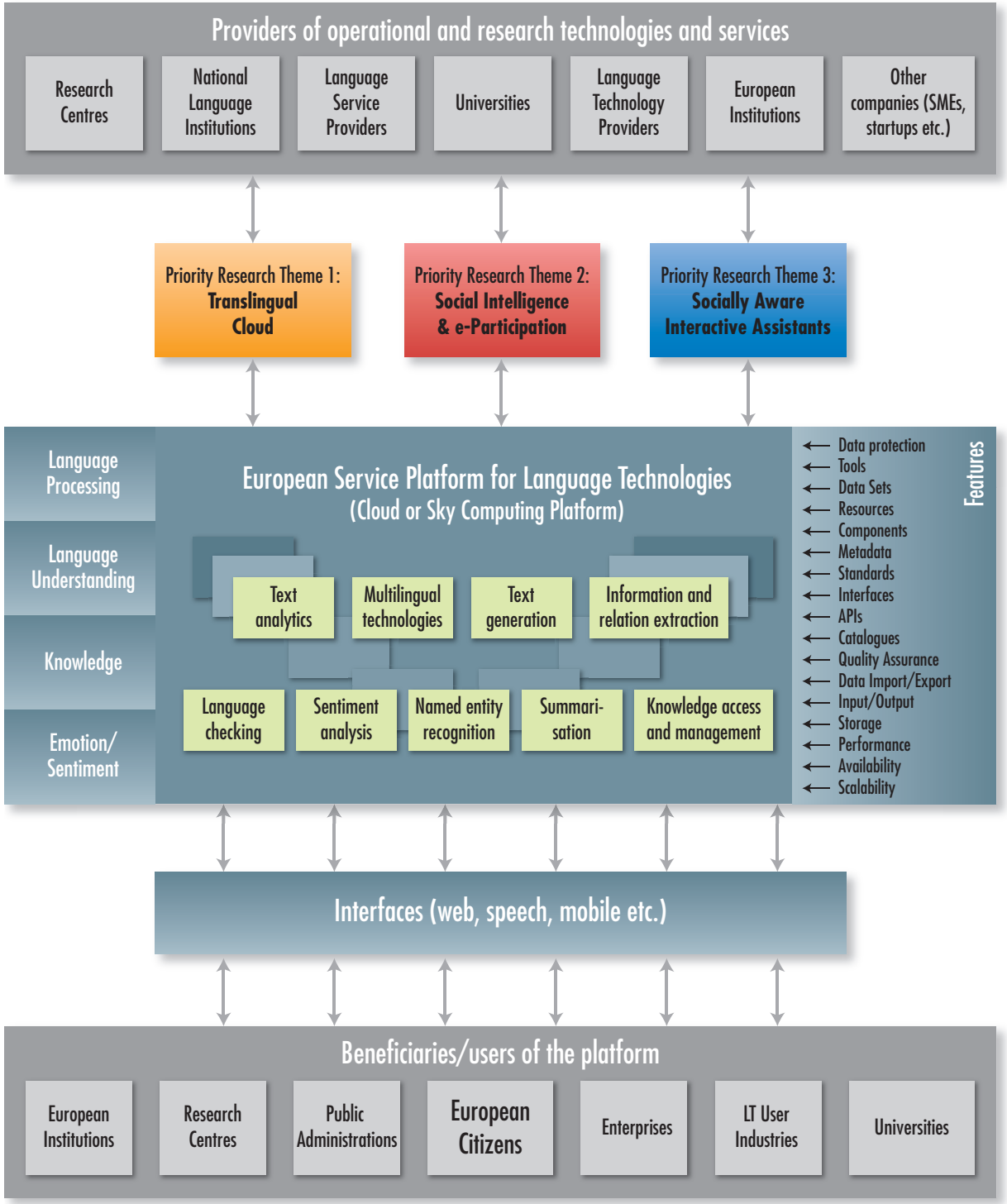

Priority Research Theme 2: Social Intelligence and e-Participation
Mapping large, heterogeneous, unstructured volumes of online content to structured, actionable representations
From shallow to deep, from coarse-grained to detailed processing techniques
Making language technologies interoperable with knowledge representation and the semantic web
"Semantification" of the web: tight integration with the Semantic Web and Linked Open Data
Services and Technologies:
Intelligent analysis of web content, especially social media, comments, blogs, forums
Detection and cross-lingual analysis of decision-relevant information
Multilingual, problem-specific decision support
Text analytics (named entity recognition, event recognition, relation extraction, sentiment analysis and opinion mining including the temporal dimension)
Syntactic, semantic, rhetorical analysis and text structure identification
Resolution of coreference or modality cues
Extraction of semantic representations from arbitrary online content
Clustering, categorising, summarising, visualising discussions and opinion statements
Applications:
Technologies for decision support, collective deliberation and e-participation
Public discussion platform for Europe-wide deliberation on pressing issues
Visualisation of social intelligence data and processes; modeling evolution of opinions
High performance web-scale content analysis technologies
Events/trend detection and prediction
Make use of the wisdom of the crowds
Improved efficiency and quality of decision processes
Unleashing social intelligence by detecting and monitoring opinions, demands, needs and problems
Target groups: European citizen, European institutions, discussion participants, companies
Understanding influence diffusion across social media

PT2: Social Intelligence

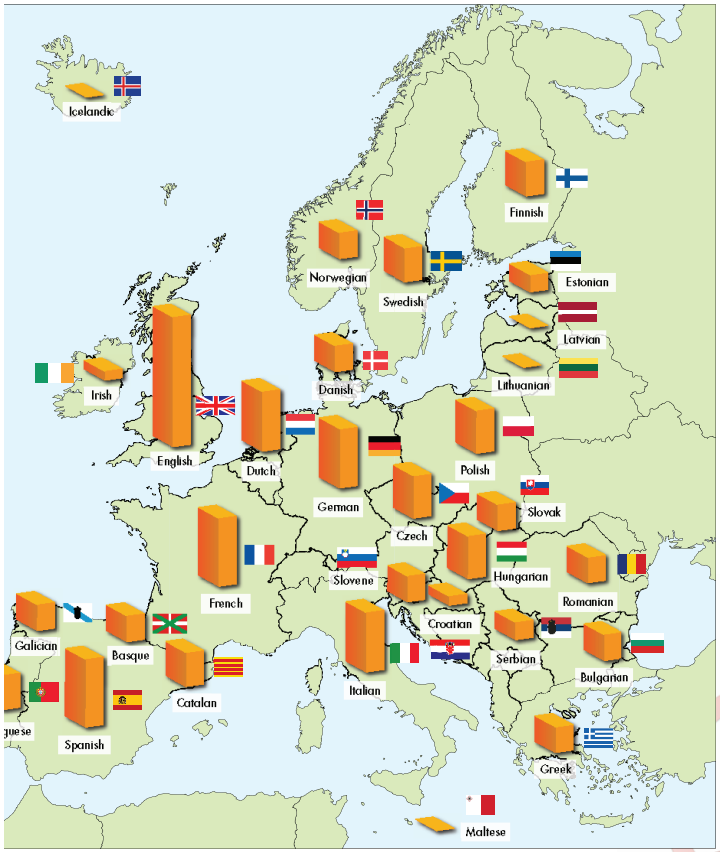
- Better decisions by monitoring social media
- Inclusion of citizens into collective decision processes
- Opinion formation, consensus building, decision making
- Evolution of new solutions
- New forms of democracy: e-democracy, massive participation, transparency
- Dialogues and debates across language boundaries and across parties, political alliances, social classes
- Better than binary voting
- Documented transparent decision processes



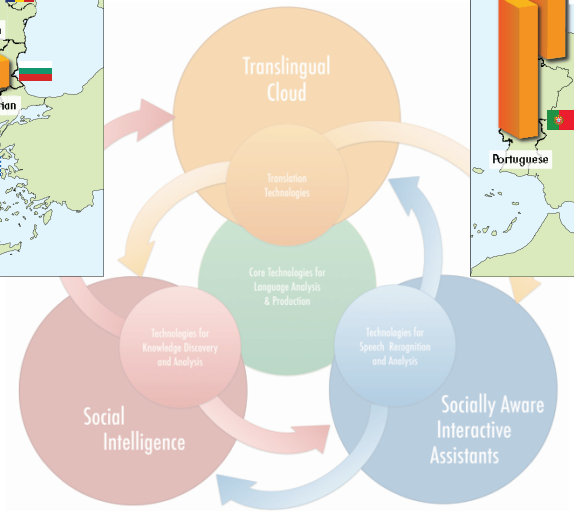

Priority Research Theme 3: Socially-Aware Interactive Assistants
Noisy environments, any speaker, open vocabulary
Error recovery, self-assessment
Multilingual capabilities
Interacting naturally with and in groups
Include human-computer, human-artificial agent and computer-mediated human-human communication
Learning and forgetting information
Adaptable to the user's needs and preferences and the environment
Services and Technologies:
Robust, accurate, incremental speech recognition
Natural, incremental speech generation and synthesis, providing expressive voices
Robust dialogue systems
From speech recognition to speech understanding
Develop methods for the support of incremental conversational speech
Context-aware semantic and pragmatic models of human communication
Parsing with support for temporal inter-dependencies
Strong connections to the other two priority themes
Applications:
Generalised and specialised interactive dialogue systems
Support people interacting with their environment
Use language in connection with other modalities (visual, tactile, haptic)
Education, language training, e-learning
Provide access to knowledge
Robust analysis of user's age, gender, verbal/non-verbal behaviour, social context
Question answering
Proactive, self-aware, user-adaptable
Interacts naturally with humans, in any language and modality
Can be personalised to individual communication abilities including special needs
Can learn incrementally from all interactions and other sources of information

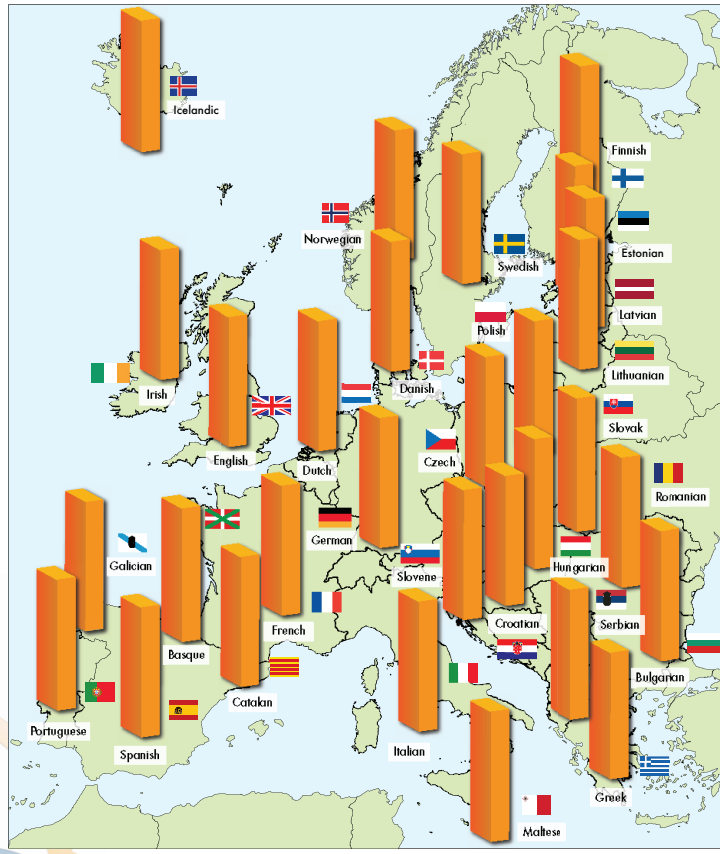

Providers of operational and research technologies and services
Research Centres
National Language Institutions
Language Service Providers
Universities
Language Technology Providers
European Institutions
Other companies (SMEs, startups etc.)
Priority Research Theme 1: Translingual Cloud
Priority Research Theme 2: Social Intelligence & e-Participation
Priority Research Theme 3: Socially Aware Interactive Assistants
Language Processing
Language Understanding
Knowledge
Emotion/ Sentiment
European Service Platform for Language Technologies
(Cloud or Sky Computing Platform)
Text analytics
Multilingual technologies
Text generation
Information and relation extraction
Language checking
Sentiment analysis
Named entity recognition
Summari-sation
Knowledge access and management
Features
Data protection
Tools
Data Sets
Resources
Components
Metadata
Standards
Interfaces
APIs
Catalogues
Quality Assurance
Data Import/Export
Input/Output
Storage
Performance
Availability
Scalability
Interfaces (web, speech, mobile etc.)
Beneficiaries/users of the platform
European Institutions
Research Centres
Public Administrations
European Citizens
Enterprises
LT User Industries
Universities

Core Resources & Technologies



2013





2020

META-NET

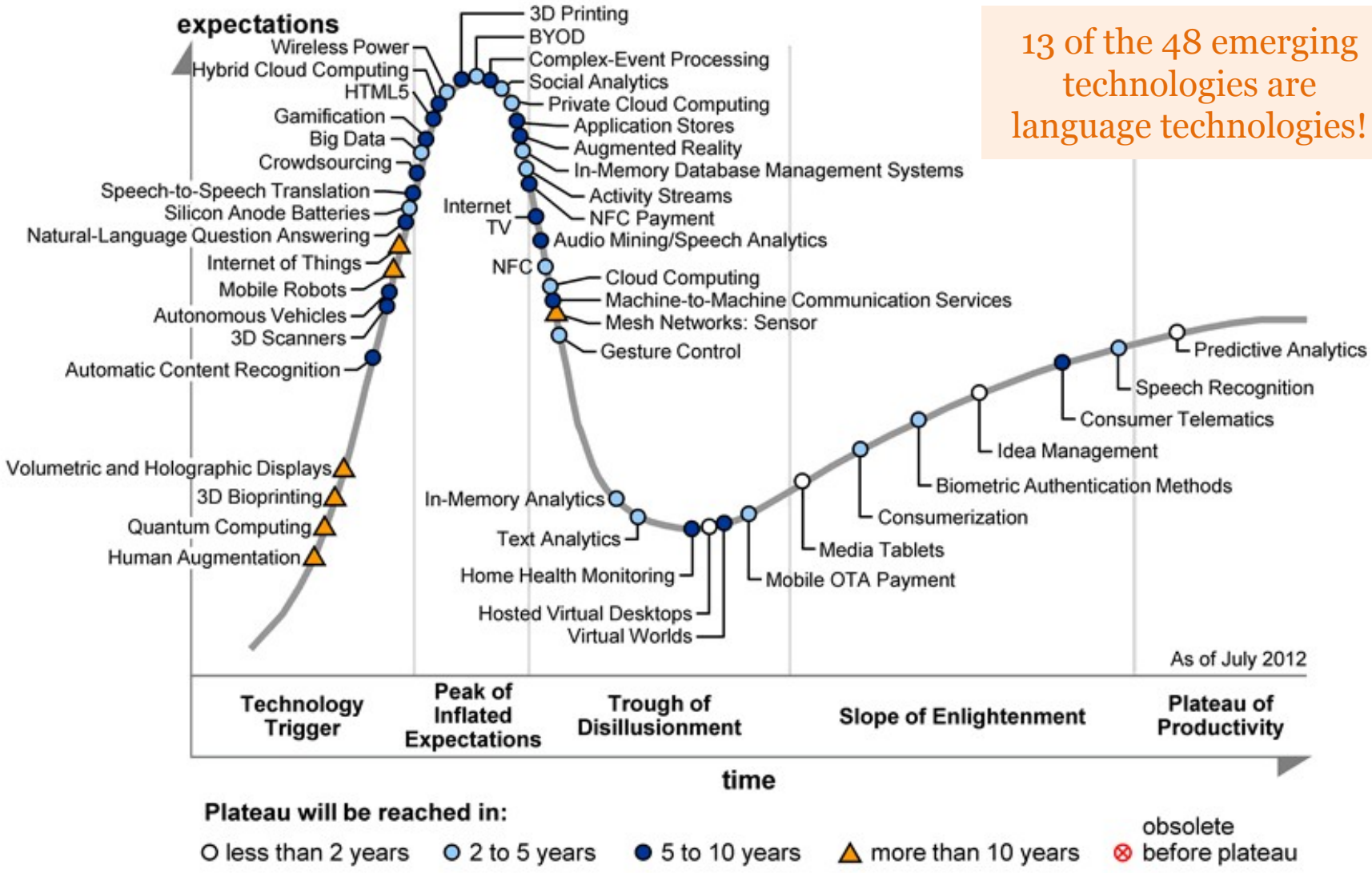
Conclusions and Next Steps

Conclusions and Next Steps

- Europe is extremely interested in and passionate about its languages.
- Our Strategic Research Agenda for LT research and innovation can put Europe ahead of its competitors in this technology area.
- Provides useful and attractive solutions to European society, at the same time creating huge business opportunities for European industry.
- Now is the time to move forward with a continent-wide, systematic push and to invest in strategic research. A modest investment is required.
- This push will generate a *countless* number of opportunities.
- H2020 and CEF can provide sufficient resources to make our visions for Europe's citizens and economy a reality.
- SRA press campaign, launch on Jan. 21, 2013.
- META-SHARE and SRA launch event on Jan. 24/25, 2013 in Berlin.

Gartner Hype Cycle 2012

META NET



13 of the 48 emerging technologies are language technologies!

Q/A

Köszönöm szépen!

office@meta-net.eu

http://www.meta-net.eu
http://www.facebook.com/META.Alliance