

# The Future European Multilingual Information Society

## Vision Paper for a Strategic Research Agenda

“People can’t share knowledge  
if they don’t speak a common language.”

Davenport, Thomas H, and Laurence Prusak, *Working Knowledge: How Organizations Manage What They Know*, Harvard Business School, Boston, 1997, p. 98.



Join the discussion at  
[www.meta-et.eu/forum](http://www.meta-et.eu/forum)

## Contents

<b>1 Introduction.....</b>	<b>3</b>
Language Technology Creates Digital Value and Fosters Social Inclusion.....	3
Visions for a Multilingual Europe.....	4
<b>2 Social and Economic Background: Challenges and Opportunities.....</b>	<b>5</b>
Linguistic, Commercial and Knowledge Barriers .....	5
Ageing Population .....	5
Population with Disabilities .....	5
Immigration and Integration.....	6
Personal Information Services and Customer Care as a Commodity .....	6
Global Cooperation and Embedded Communication .....	6
Social Media and Participation .....	7
Market Awareness and Customer Acceptance .....	7
Single Market, Many Languages .....	7
Secure Europe.....	8
Future Trends .....	8
<b>3 Visions for a Multilingual Europe .....</b>	<b>9</b>
Vision One: A Language-Transparent Web and Media.....	9
Vision Two: Natural and Inclusive Interaction.....	10
Vision Three: Efficient Information Management .....	11
<b>4 Summary and Future Efforts.....</b>	<b>13</b>
<b>5 Appendix: META-NET .....</b>	<b>14</b>
<b>6 Appendix: The Vision Building Process .....</b>	<b>16</b>
<b>7 Appendix: Contributors .....</b>	<b>17</b>
Members of the Vision Groups.....	17
Members of the Technology Council .....	20

## Imprint

Editors:

Dr. Aljoscha Burchardt, DFKI

Dr. Georg Rehm, DFKI

Prof. Dr. Felix Sasaki, DFKI

Approximately one hundred researchers and industry representatives provided input to this document as participants in one of three META-NET Vision Groups or as members of the META Technology Council. Numerous others contributed to online and offline discussions.



The development of this paper has been funded by the Seventh Framework Programme and the ICT Policy Support Programme of the European Commission under contracts T4ME (Grant Agreement 249119), CESAR (Grant Agreement 271022), METANET4U (Grant Agreement 270893) and META-NORD (Grant Agreement 270899).

## 1 Introduction

### Language Technology Creates Digital Value and Fosters Social Inclusion

A single European market that secures wealth and social well-being is possible, but linguistic barriers still severely limit the free flow of goods, information and services. Many Europeans find it difficult to interact with online services and participate in the digital economy. According to a recent report requested by the European Commission, 57% of European Internet users purchase goods and services in other languages. Reading content in a foreign language is accepted by 55% of users while only 35% may use another language when writing emails or posting comments on the web.<sup>1</sup>

Just a few years ago, English was the lingua franca of the web. The vast majority of content on the web was in English. Today, the situation has drastically changed. The amount of online content in other languages, especially Asian and Arabic languages, has exploded. Europe must take action to prepare its 23 official languages and 60 spoken languages for the digital age. European languages are a cultural asset that requires future-proofing! In fact, a recent UNESCO report on multilingualism states that languages are an essential medium for the enjoyment of fundamental rights, such as political expression, education and participation in society.<sup>2</sup> If Europe does not support and promote its diversity of languages, European languages could become irrelevant or underrepresented on the web.

According to some estimates, the European market for translation, interpretation, software localisation and website globalisation was € 8.4 billion in 2008 and was expected to grow by 10% per annum.<sup>3</sup> Yet, this existing capacity is not enough to satisfy current and future needs. Although computers can better handle, process and understand language, machine translation is not a panacea. If we rely on existing technologies, automated translation and the ability to process a variety of content in a variety of languages, a key requirement for the future Internet, will be impossible.<sup>4</sup> The same argument applies to information services, document services, media industries, digital archives and language teaching. There is an urgent need for innovative technologies that help save costs while offering faster and better language services to European citizens.

**Language technology (LT)** is a key, enabling technology for the knowledge society. LT supports humans in everyday tasks, such as writing e-mails, searching for information online or booking a flight. We benefit from language technology when we:

<sup>1</sup> European Commission Directorate-General Information Society and Media, Organization, *User language preferences online*, Flash Eurobarometer #313, 2011 ([http://ec.europa.eu/public\\_opinion/flash/fl\\_313\\_en.pdf](http://ec.europa.eu/public_opinion/flash/fl_313_en.pdf)).

<sup>2</sup> UNESCO Director-General, *Intersectoral mid-term strategy on languages and multilingualism*, Paris, 2007 (<http://unesdoc.unesco.org/images/0015/001503/1503335e.pdf>).

<sup>3</sup> European Commission Directorate-General for Translation, *Size of the language industry in the EU*, Kingston Upon Thames, 2009 (<http://ec.europa.eu/dgs/translation/publications/studies>).

<sup>4</sup> Oxford Internet Institute and SCF Associates Ltd, *Towards a Future Internet: Interrelation between Technological, Social and Economic Trends*, 2010 ([http://cordis.europa.eu/fp7/ict/fire/docs/tafi-final-report\\_en.pdf](http://cordis.europa.eu/fp7/ict/fire/docs/tafi-final-report_en.pdf)).

- use the spelling and grammar checking features in a word processor;
- view product recommendations at an online shop;
- hear the verbal instructions of a synthetic voice in a navigation system;
- translate web pages with an online service.

Although many language services are provided free of charge, American companies primarily offer these services. The recent success of Watson, an IBM computer system that won an episode of the *Jeopardy* game show against human candidates, illustrates the immense potential of language technology. As Europeans, we have to ask ourselves several urgent questions:

- Should our communications and knowledge infrastructure be dependent upon monopolistic companies?
- How do we respond when the language-related services that we rely upon are switched off?
- Are we actively competing in the global landscape for research and development in language technology?
- Can third parties from other continents, who are not knowledgeable about European languages, solve our translation problems?
- Can the European cultural background help shape the knowledge society by offering better, more secure, more precise, more innovative and more robust high-quality technology?

We believe that secure and innovative language technology made in Europe will significantly contribute to future European economic growth and social stability while establishing for Europe a worldwide, leading position in technology innovation.

### Visions for a Multilingual Europe

This report presents visions for a multilingual Europe that help overcome remaining cultural, economic, linguistic, political, and social challenges with the help of language technology. The visions presented in this report directly result from an exhaustive community outreach programme that involves roughly one hundred distinguished industry participants and researchers in the field of language technology. This work has been conducted under the auspices of META-NET, a European Commission Network of Excellence that consists of 44 research centres from 31 countries.

META-NET is forging the Multilingual Europe Technology Alliance (META) by building a strong community for language technologies. The visions presented in this document are an important contribution to the Digital Agenda for Europe, an initiative of the European Commission. This document will also contribute to a strategic research agenda (SRA) for language technology that will be presented at the META-FORUM 2011 conference, which will be held 27-28 June in Budapest, Hungary (<http://www.meta-forum.eu>).

## 2 Social and Economic Background: Challenges and Opportunities

Many societal changes and economic trends confirm the urgent need to include substantial amounts of language technology in our European information and communication technology (ICT) infrastructure. To go beyond what is possible today, research and development efforts in language technology must increase.

### Linguistic, Commercial and Knowledge Barriers

A report to the Directorate General for Translation of the European Commission on cross-border, online commerce in the EU clearly indicated that language barriers are economic barriers.<sup>5</sup> Only 59% of retailers can handle transactions in more than one language. Translation and localisation costs must be drastically lowered before broad participation in Europe's single digital market can become a reality. In this regard, multilingual language technology is the key, especially for small and medium-sized enterprises (SME). At the same time, user expectations in the information society are increasing: 81% of all Internet users think that websites produced in their country should also be available in other languages. 44% of European users believe they miss interesting information because websites are not available in a language that they understand. These facts can no longer be ignored. Robust and affordable language technology must be developed and integrated into end-user software, such as web browsers and e-mail clients. The availability of reliable language technology can help establish a potentially vast market for information as well as consumer and entertainment goods in any language.

Language barriers are commercial barriers.

### Ageing Population

Demographic changes suggest the need for more assistive technologies, especially those that drastically improve spoken language access. An aging population requires technology that can help master everyday situations and provide proactive guidance. Such technologies could eventually answer the question, "Where did I leave my glasses?" The economic cost of demographic changes will also mean that more health care services and support systems will be required in our homes. Ambient assisted living (AAL) technologies can greatly benefit from a personalised, spoken method of interaction that is possible due to recent developments in the field of robotics. The technology must be affordable and easy to use. The full complexity of language technology must also be hidden from users that have minimal experience using advanced technologies.

### Population with Disabilities

The way we deal with disabilities has changed dramatically in the last 20 years. We have shifted from an approach based on assistance, recovery or maintenance of functional capabilities to a goal of fully integrating individuals. The use of new technologies can help us reach the ambitious goal of achieving equal opportunities, promoting independent living and integrating persons with disabilities. Speech and lan-

<sup>5</sup> European Commission Directorate-General for Translation, *Size of the language industry in the EU*, Kingston Upon Thames, 2009 (<http://ec.europa.eu/dgs/translation/publications/studies>).

guage technologies already help people with disabilities participate in society. Noteworthy examples include screen readers, dictation systems and voice-activated services. Technology can offer much more: automatic sign language recognition; automatic summarization and translation; content simplification; and interactive virtual reality systems. In addition to the social aspect of such developments, there is a huge commercial market for such services. Approximately 10% of Europeans have permanent disabilities, which means, there are about 50 million citizens with disabilities in the EU.

### **Immigration and Integration**

According to the *International Migration Report 2002* of the UN Department of Economic and Social Affairs, 56 million migrants lived in Europe in 2000.<sup>6</sup> The number of migrants has roughly grown to 60 million people today. Facilitating communication, providing access to information in foreign languages and helping people learn European languages can help better integrate migrants into European society. In fact, speech and language technologies can dramatically improve the integration process. Advanced language learning tools; automatic real time subtitling; automatic and simultaneous speech-to-speech, text-to-text or speech-to-text translation; improved search engines; and automatic summarization applications are just some language technologies that can help.

### **Personal Information Services and Customer Care as a Commodity**

Broadband access to information and services is commonplace, and mobile communication is daily routine for millions of Europeans. In this 24/7 economy we expect quick and reliable answers as well as engaging and timely online news broadcasts. But, information overload is common, and it limits exchange in the digital information society. Citizens, governments and industries would greatly benefit from new technologies that help get the situation back under control. Technologies, such as automatic and intelligent question-answering software as well as automatic, personalised and trusted text and speech processing of e-mail messages, news items and other textual content, can make information more relevant, timely and useful.

### **Global Cooperation and Embedded Communication**

Companies need to address new markets where multiple languages are spoken and support multinational teams at disperse locations. Many jobs cannot be filled today because linguistic barriers exclude otherwise qualified personnel. A flexible and mobile population requires multilingual language skills. Improvements in language technology can enable richer interactions and provide more advanced video conferencing services. Advances such as simultaneous translation, automatic minute taking, video indexing and video searching will increase productivity. Future technologies like a three dimensional Internet can enable new modes of situation-based collaboration in the workplace as well as support more realistic training and education scenarios. We will soon be able to participate in virtual events as new forms of entertainment, cultural exchange and tourism. Combining multilingual lan-

<sup>6</sup> UN Department of Economic and Social Affairs Population Division, *International Migration Report 2002*, New York, 2002 (<http://www.un.org/esa/population/publications/ittmig2002/2002ITTMIGTEXT22-11.pdf>).

guage technology with 3D virtual worlds and simulations will let us experience being European in a brand new way.

### Social Media and Participation

Participation in online social media networks is a key characteristic of the early twenty-first century. Social media has a tremendous impact on practically all areas of society and life. Social media can also help us solve technical problems, research products, learn about interesting places or discover new recipes. At the same time, recent developments in North Africa demonstrate the ability of social media to bring citizens together so they can express political power. Social media will play a role in the discussion of important, future topics for Europe like a common energy strategy and a common foreign policy.

A key problem is that certain groups are becoming more and more detached from these developments. One can even speak of a broken link regarding communication cultures. This is an issue since both types of bottom-up movements sketched above are highly relevant for politicians, marketing experts, or journalists who would like to know what their customers or citizens think about their initiatives, products, or publications and to be able to react accordingly. However, it is not possible to process the sheer amount of information generated in multiple languages on social networks with manual approaches. Although crowdsourcing could help gather information to a certain extent, the use of sophisticated language technology is the only way to analyse developments in real time.

### Market Awareness and Customer Acceptance

Language technology is a key part of business and consumer software. The exact size of this market is difficult to pinpoint because language technologies are often hidden inside other, more visible products. Customer acceptance of language technology has been recently shown to be high. For example, market research by the Ford Motor Company indicates that the voice control system, Ford SYNC, is widely accepted.<sup>7</sup> 60% of Ford vehicle owners use voice commands in their cars. Non-Ford owners report a three-fold increase in their willingness to consider Ford models while 32% of existing customers admit that the technology played an important or critical role in their purchase decision. Language technology has a tremendous market potential, especially if European companies can quickly enter this emerging market that covers many different languages.

### Single Market, Many Languages

Support for the 23 official languages of the EU has major economic and social implications, but the political dimension is equally important. Europe currently lags behind countries such as India and South Africa. Government programmes in these two countries actively foster the development of language technology for a significant number of official languages, especially for mobile devices. Mobile devices will become an even more important connection point between humans and information technology. Google already provides free translation services in

### The Political Dimension

<sup>7</sup> Ford Motor Company, *Fact Sheet: Ford SYNC Voice-Controlled Communications & Connectivity System* ([http://media.ford.com/article\\_display.cfm?article\\_id=33358](http://media.ford.com/article_display.cfm?article_id=33358)).

3,306 different language pairs (including Basque and Catalan) as well as voice input for 16 languages and speech output for 24 languages. The Apple iTunes and App Store has demonstrated how premium content and products can be marketed for free and for payment. Europe must address this global competition.

### Secure Europe

The evolving information and knowledge society has improved human communication and information access, but the same communication networks also help some commit crimes like identity theft and Internet fraud. The effective persecution of illegal activities requires automatic tools that can help detect crimes and monitor offenders. Systems that can monitor, analyse and summarise large amounts of text, audio and video data in different languages (European and non-European) and from different sources (websites and social media) demand sophisticated language technologies.

### Future Trends

The visions presented in this document are strongly influenced by larger trends, such as cloud computing, social media, mobile apps and web services. Many of these products and services are only available online, for example, limiting access to Facebook and Twitter strongly influenced the course of recent political developments of North Africa. Multilingualism has now become a global norm rather than an exception. Less than 30% of web-based content is in English while other estimates suggest that only 50% of Twitter messages are in English. Future 3D applications that embed information and communication technology require sophisticated language technology. Autonomous robots with language technology capabilities could potentially help in catastrophes by rescuing travellers from public transportation or by giving first aid. Language technology can significantly contribute towards improving social inclusion. Language technology can also help us provide answers to urgent social challenges while creating genuine business opportunities.

Less than 30% of all web content is in English.



### 3 Visions for a Multilingual Europe

#### Vision One: A Language-Transparent Web and Media



**The web is multilingual and multimedia.** Although the World Wide Web makes it easy to consume or publish content for business or private use, most of the content remains hidden from view. Information is often confined by the social silos of closed linguistic communities, or it is kept inside proprietary media files (video formats or presentations) that users cannot access or search. This valuable knowledge and information should become part of the European social and economic communications space.

Our goal is to facilitate and enable the use of advanced, networked ICT for European citizens. Speech-based interaction and other more novel interaction approaches will enable the use of ICT in situations where touch and visual interaction methods are not possible or feasible. We will also observe a stronger integration of social applications, media use and information access. Watching a quiz show on television will be more like edutainment when access to interactive background information becomes available and when friends and family can join in a videoconference. Social applications will further help those who are temporarily or permanently immobile, so they can participate in social life and activities.

We envision a language-transparent web and language-transparent media. This vision will help realise a truly multilingual mode of online and media interaction for every citizen regardless of age, education, profession, cultural background, language proficiency or technical skills. Transparency means everyone can contribute in their own language and be understood by members of other linguistic communities. Language technology will provide the necessary technical functionality for this layer of transparency, and it will create the foundation for truly multilingual online communication, commerce and education.

Examples of this vision include:

- ▣ *Cross-lingual information access to the web and to media in all languages.* Language technology will help provide access to the entire world of the multilingual web by supporting 200 to 1000 languages with innovative applications, such as cross-lingual queries, automated question-answering, natural language search, conversational agents, and social media that includes the automatic translation of chats, tweets and e-mails. Books, laws, movies, newspapers, radio broadcasts, television broadcasts and other sources of information produced in the 27 European member states will be provided to citizens in their own language(s). Online advertising will also benefit from the application of multilingual, semantic technologies for better and less expensive customisation.
- ▣ *Multimedia multi-language subtitling.* Language technology will provide speech recognition for automatic subtitling. Such efforts will respond to forthcoming legal requirements in many EU countries to provide subtitles for television programmes in real time. Current, manual approaches are not feasible or scalable. Personalisation is another important aspect, so people with different reading speeds and language capacities can have subtitles tailored to meet their specific needs. At the same time, audio and video translation will be easily available for all Euro-

pean languages while ensuring flexible approaches that provide full translation or partial support for difficult passages. The required applications will foster access to information for ethnic minorities and for new immigrants as well as facilitate the exchange of information between EU countries.

- *Making documents understandable.* Language technology will provide the means to rephrase complicated documents like legal texts or prescription package inserts into everyday, simplified language. Such technology can potentially overcome digital illiteracy and exclusion by adapting human-computer-interaction according to language proficiency. Existing LT algorithms for tasks, such as language simplification or automatic summarisation, can be further developed along with language generation technology. For some texts like technical manuals, such developments can help companies better address requirements for unrestricted access.

### Vision Two: Natural and Inclusive Interaction

**Digital communication does not have any borders.** The technical infrastructure for digital communications has grown to a global scale, but the means of supporting natural communication and interaction has not kept up with the pace of development. In many situations involving human communication or interaction between humans and machines, participation is limited to a small number of supported languages, and participation also depends on other factors, such as cultural norms or social circumstances.

We envision language technology helping to overcome the invisible borders in human-to-human and human-to-machine communication. Language technology will enable natural interaction and inclusion while addressing the growing cultural and social diversity in Europe. Language technology will assure that diversity will no longer be a challenge, but a welcome enrichment for Europe both socially and economically.

Examples of this vision include:

- *Natural interaction with agents and robots.* Language technology will be the basis for developing self-learning, context-aware, personalised agents that have speech, language and multi-modal input and output abilities. Language technology will help in augmented reality contexts, human-machine interactions and human communications that are mediated by machines. The agents will also help to perform low-level tasks, such as processing e-mails, voice messages or telephone calls. Agents will also exist as autonomous robots that will naturally interact with humans while learning from such interaction.
- *Assistive applications—technologies that help.* Language technology will provide interactive information systems like personalised speech technology systems for persons with reduced motor control as well as home care interfaces for older or disabled persons. Language technology will also help companies, governments and other organisations respond to forthcoming legal requirements to provide information for people with disabilities (blind, hearing impaired, speech impaired, motion disabled and others) and to support education, rehabilitation and training for disabled persons. For example, sign language



recognition, synthesis and translation will allow deaf people to communicate via remote communication networks and to communicate with people who cannot sign. Such technologies will give the deaf access to the same services as European citizens who do not require sign language interpreters.

- *Cross-lingual E-learning.* We expect E-learning to significantly benefit from improved language technology. Lifelong learning, continuing education, on-the-job training and traditional education programs at universities and schools will become more practical, task-driven, interactive, and international once ICT, augmented with innovative LT, becomes more widely available. Such applications demonstrate the social impact of LT and the high level of interoperability that can be reached. Social, cross-lingual, E-learning environments can make it possible for participants from different parts of the world to cooperate in tandem or in larger groups. New types of interactive systems are also necessary for schools and universities, so they can support remote students, who may want to use their own preferred languages. Such systems must additionally support self-assessment techniques that check the quality of spoken utterances (pronunciation, timbre, intonation, rhythm, accent and so on) in Computer Assisted Language Learning (CALL) systems.
- *Cross-lingual meeting assistants that support speech-to-speech translation.* Language technology will lead to the development of social applications that are based on human interactions, such as meeting support systems, videoconferencing applications, remote meeting systems and personal devices that can monitor the current communication context. A cross-lingual assistant for virtual and face-to-face meetings will definitely benefit global teams, who videoconference with other worldwide office locations, and assistants can help include participants, who speak different languages. Each person can use their own preferred language without wearing headphones or using microphones. Translation engines will provide instant translations that imitate distinctive features of the original voice—its pitch, speed, tone or emotional emphasis. At the same time, language technology will help immediately transform slides, presentations and handwritten notes on a whiteboard into a preferred language. During meetings, minutes will be automatically produced and video recordings will be automatically indexed to support voice searching, transcription and translation.

### Vision Three: Efficient Information Management

**Information is growing without limits.** Digital communication is responsible for the tremendous, worldwide growth of content and information. The availability of high-quality, free information is just one benefit of this development. On the other hand, the sheer amount of information makes it increasingly difficult, if not impossible, to sensibly manage information. Without the means of accessing relevant information for a given situation, the value of the digital information space itself is in danger.



We envision that language technology will ensure and extend the value of the digital information space. LT will help deliver personalised information access and management. Users will be able to aggregate the information they need, evaluate the quality of the information, put it in context and share it. Language technology applications will result the transformation of information into knowledge, which will create new business opportunities for information brokers and knowledge providers.

Examples of this vision include:

- ▣ *Federated multilingual audio-visual search.* Language technology will improve the search for audio and video materials across languages by combining many LT techniques, such as the recognition and identification of objects, persons and actions as well as optical character recognition (OCR) and speech recognition of ordinary (untrained) voices. The efficient utilisation of millions of available knowledge bases such as linked data/annotated web and social/real time streams will provide a larger number of texts in European languages that will help improve the performance and scalability of LT systems. Semantic analysis of audio and video content and other intelligent search techniques will help create search engines that can provide reliable answers to everyday questions.
- ▣ *Personalised information assistants.* Language technology will help create personal assistants that can offer task-centred, interactive support in gathering and structuring information. This new generation of information assistants will provide support just like human assistants by filing documents, reformatting materials, copying information from one document to another, preparing standard letters and answering information requests. When writing a document or preparing a presentation, the assistant can look for similar documents on a local hard drive, a network, an intranet, or the Web. The assistant can provide information that already exists and examples that match the user's needs. In social media environments, the assistant can support text mining, handle information overload or perform semantic searches.
- ▣ *Life logging.* Language technology will help capture every utterance and conversation during the day while semantically structuring the information into meaningful bits and pieces. Such efforts will create a so-called gazetteer of concepts for gathering information on a massive scale. Such materials will greatly aid language technologies since they require large volumes of sample data to improve their algorithms.

## 4 Summary and Future Efforts

Language technology is a key enabling technology for a multilingual Europe. Secure and innovative language technology made in Europe will significantly contribute to future European economic growth and social stability while establishing for Europe a worldwide, leading position in technology innovation.

This paper presented future visions for innovative language technology applications in three areas:

- ▣ language-transparent web and media experience,
- ▣ natural and inclusive interaction,
- ▣ efficient information management.

The visions directly result from an exhaustive community outreach programme that involves roughly one hundred distinguished industry participants and researchers in the field of language technology. This work has been conducted under the auspices of META-NET, a European Commission Network of Excellence that consists of 44 research centres from 31 countries.

Based among other things in this vision report, META-NET will collaboratively prepare, establish and promote a strategic research agenda (SRA) for language technology in Europe. It will be presented on 27-28 June during the META-FORUM 2011 conference in Budapest, Hungary.

This paper addresses journalists, political referents, language communities, scientists, and other people that might be interested in supporting the common goal of establishing a truly multilingual Europe. You can help META by:

- ▣ spreading the word about our initiative;
- ▣ participating in our online discussion forum (<http://www.meta-net.eu/forum>);
- ▣ joining META as a registered organisation (<http://www.meta-net.eu/join>).

META-NET and META offer many more opportunities for participation. Please check out [www.meta-net.eu](http://www.meta-net.eu) for information on upcoming events and activities.

## 5 Appendix: META-NET

META-NET is a Network of Excellence funded by the European Commission. The network currently consists of 44 research centres that represent 31 European countries. META-NET is fostering the Multilingual Europe Technology Alliance (META), a growing community of language technology professionals and organisations in Europe.

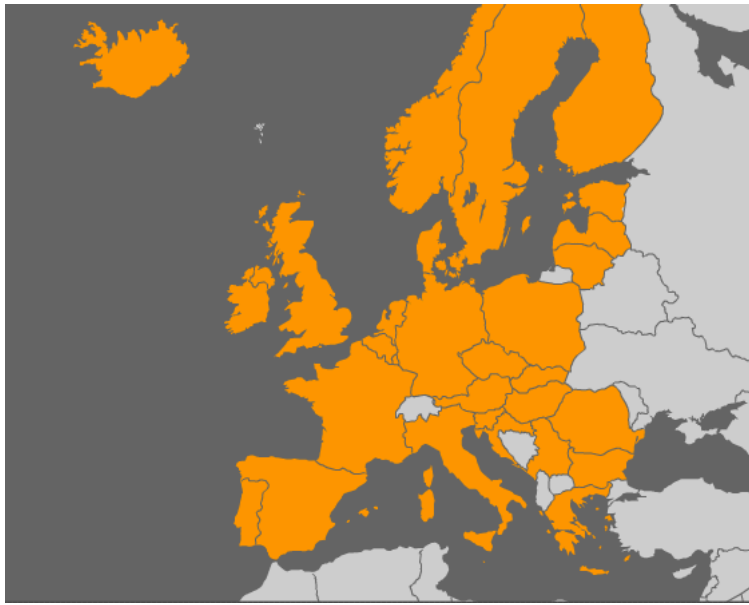


Figure 1: Countries Represented in META-NET

META-NET stimulates and promotes multilingual technologies for all European languages. The technologies enable automatic translation, content production, information processing and knowledge management for a wide variety of applications and subject domains. The network wants to improve current approaches, so better communication and cooperation across languages can take place. Europeans have an equal right to information and knowledge regardless of language.

META-NET launched on 1 February 2010 with the goal of advancing research in language technology. The initiative supports a Europe that unites as a single, digital market and information space. META-NET has conducted several activities that further its goals. META-VISION, META-SHARE and META-RESEARCH are the network's three lines of action.

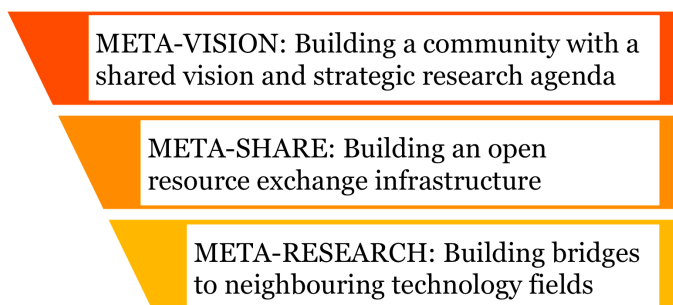


Figure 2: Three Lines of Action in META-NET

**META-VISION** fosters a dynamic and influential stakeholder community that unites around a shared vision and a common strategic research agenda (SRA). The main focus of this activity is to build a coherent and cohesive LT community in Europe by bringing together representatives from highly fragmented and diverse groups of stakeholders. In META-NET's first year, presentations at the FLAReNet Forum (Spain), Language Technology Days (Luxembourg), JIAMCATT 2010 (Luxembourg), LREC 2010 (Malta), EAMT 2010 (France) and ICT 2010 (Belgium) centred on public outreach. According to initial estimates, META-NET has already contacted more than 2,500 LT professionals to share its goals and visions with them. At the META-FORUM 2010 event in Brussels, META-NET shared the initial results of its vision building process to more than 250 participants. In a series of interactive sessions, the participants provided feedback on the visions presented by the network.

**META-SHARE** creates an open, distributed facility for exchanging and sharing resources. The peer-to-peer network of repositories will contain language data, tools and web services that are documented with high-quality metadata and organised in standardised categories. The resources can be readily accessed and uniformly searched. The available resources include free, open source materials as well as restricted, commercially available, fee-based items. META-SHARE targets existing language data, tools and systems as well as new and emerging products that are required for building and evaluating new technologies, products and services. The reuse, combination, repurposing and re-engineering of language data and tools plays a crucial role. META-SHARE will eventually become a critical part of the LT marketplace for developers, localisation experts, researchers, translators and language professionals from small, mid-sized and large enterprises. META-SHARE addresses the full development cycle of LT—from research to innovative products and services. A key aspect of this activity is establishing META-SHARE as an important and valuable part of a European and global infrastructure for the LT community.

**META-RESEARCH** builds bridges to related technology fields. This activity seeks to leverage advances in other fields and to capitalise on innovative research that can benefit language technology. In particular, this activity wants to bring more semantics into machine translation (MT), optimise the division of labour in hybrid MT, exploit context when computing automatic translations and prepare an empirical base for MT. META-RESEARCH is working with other fields and disciplines, such as machine learning and the Semantic Web community. META-RESEARCH focuses on collecting data, preparing data sets and organising language resources for evaluation purposes; compiling inventories of tools and methods; and organising workshops and training events for members of the community. This activity has already clearly identified aspects of MT where semantics can impact current best practices. In addition, the activity has created recommendations on how to approach the problem of integrating semantic information in MT. META-RESEARCH is also finalising a new language resource for MT, the Annotated Hybrid Sample MT Corpus, which provides data for English-German, English-Spanish and English-Czech language pairs. META-RESEARCH has also developed software that collects multilingual corpora that are hidden on the web.

## 6 Appendix: The Vision Building Process

META-NET will collaboratively prepare, establish and promote a strategic research agenda (SRA) for language technology in Europe. The SRA is a long-term plan that will engage academic and commercial research and development up to the year 2020. The SRA will contain high-level recommendations and suggestions for joint action. META-NET will eventually present the SRA to the European Commission, national funding agencies and regional bodies.

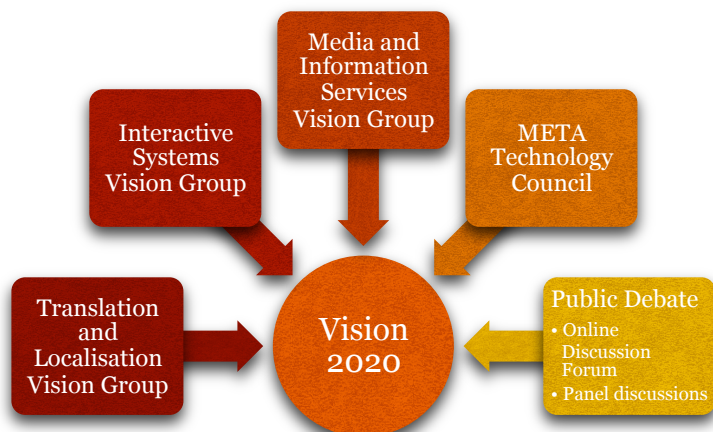


Figure 3: The Vision Building Process

Work on the SRA has just begun. The document, which will be initially published in Summer 2011, is understandably complex. It must reflect the participation of various stakeholders in the language technology community. To this aim, META-NET organised three vision groups and an online forum. The vision groups and the forum encourage stakeholders to express their opinion and be heard.



Figure 4: Timeline of the Vision Building Process

This document presents the outcome of several vision group meetings, which were conducted in 2010 and 2011, and many suggestions from the online forum. The complete reports for each vision group are available online (<http://www.meta-net.eu/vision>). An additional group of external experts, the META Technology Council, is currently consolidating the various findings and feedback gathered from an additional online forum. The technology council will prepare a draft of the SRA for comment and review by the community.



## 7 Appendix: Contributors

The members of the META-NET Vision Groups and the META Technology Council come from diverse backgrounds including commercial businesses, government agencies, industry, research organisations, software companies, technology providers, and European universities. The general public was invited to the discussion via an online discussion forum.

### Members of the Vision Groups

META-NET assembled three vision groups. Each group targets stakeholders from a different technology sector:

- ▣ translation and localisation,
- ▣ media and information services,
- ▣ interactive systems.

The groups, who each met three times in 2010 and 2011, bring together developers, integrators, researchers and users of language technology applications, products and services. The groups address the needs of service providers and users. The following table lists the current members of the vision groups.

Name	Organisation	Country	Sector
Sophia Ananiadou	University of Manchester	UK	LT research
Toni Badia	BM	Spain	Speech and natural language processing (NLP) research; META-NET
Christoph Bauer	ORF	Austria	Broadcasting technologies
Nozha Boujemaa	INRIA	France	Multimedia content search; Digital libraries
Antonio Branco	University of Lisbon	Portugal	LT research; META-NET
Andrew Bredenkamp	Acrolinx	Germany	Authoring support
Gerhard Budin	Universität Wien	Austria	LT research; META-NET
Axel Buendia	SpirOps	France	Robotics and games
Aljoscha Burchardt	DFKI	Germany	LT research; META-NET
Nicoletta Calzolari	CNR	Italy	LT research; META-NET
Nick Campbell	Trinity College Dublin	Ireland	Speech technology
Jean Carrive	INA	France	Audiovisual archives
Khalid Choukri	ELDA	France	Language resources; LT; META-NET
Ann Copestake	University of Cambridge	UK	LT research
Morena Danieli	Loquendo	Italy	Spoken dialogue; Text analysis for text-to-speech and emotional text-to-speech
Claude de Loupy	Syllabs	France	Information services; Semantic search
Maarten de Rijke	University of Amsterdam	Netherlands	LT research
Marin Dimitrov	Ontotext	Bulgaria	Knowledge management; Semantic Web; Web services
Petar Djekic	Sound Cloud	Germany	Audio technologies
Christoph Dosch	IRT	Germany	Broadcasting technologies; Multimedia content search
David Filip	Moravia Worldwide	Czech Republic	Translation and localisation
Dan Flickinger	Stanford University	USA	LT research

Name	Organisation	Country	Sector
Gil Francopoulo	Tagmatica; IMMI	France	NLP; Standards
Piotr W. Fuglewicz	TiP	Poland	NLP; MT
Robert Gaizauskas	University of Sheffield	UK	IE; Dialogue modelling
Martine Garnier-Rizet	Vecsys and IMMI	France	Mobile applications; Call centres
Simon Garrett	British Telecom	UK	eCommerce
Stefan Geissler	Temis	Germany	Semantic text technologies
Edouard Geoffrois	DGA	France	Defence applications
Daniel Grasmick	Lucy Software	Germany	MT
Gregory Grefenstette	Exalead	France	NLP; IE; IR
Marko Grobelnik	JSI	Slovenia	LT research; META-NET
Joakim Gustafson	KTH	Sweden	Speech technology
Jan Hajic	Charles University	Czech Republic	LT research; META-NET
Paul Heisterkamp	Daimler	Germany	Automotive industry
Mattias Heldner	KTH	Sweden	Speech technology
Manuel Herranz	PangeaMT	Spain	Translation and localisation
Timo Honkela	Aalto University	Finland	Speech technology
Krzysztof Jassem	Poleng	Poland	Translation
Kristiina Jokinen	University of Helsinki	Finland	Interactive Systems
John Judge	DCU	Ireland	LT research; META-NET
Martin Kay	University of Saarland; Stanford University	Germany; USA	MT research
Christopher Kermorvant	A2iA	France	Document classification; Advanced data extraction
Simon King	University of Edinburg	UK	Speech technology
Maria Koutsombogera	ILSP	Greece	LT research; META-NET
Steven Krauwer	University of Utrecht	Netherlands	NLP and digital humanities
Verena Krawarik	APA	Austria	Press agency; Innovations expert
Stefan Kreckwitz	Across Systems	Germany	Translation workflow support
Simon Krek	JSI	Slovenia	LT research; META-NET
Brigitte Krenn	OFAI	Austria	LT research
Michal Küfhaber	Skrivanek	Czech Republic	Translation
Jimmy (Siegfried) Kunzmann	European Media Laboratory	Germany	Human-machine interfaces; Location-based services; Mobile users
Bernardo Magnini	FBK	Italy	LT research; META-NET
Gudrun Magnusdottir	ESTeam AB	Sweden	Machine translation
Elisabeth Maier	CLS Communication AG	Switzerland	Language services
Joseph Mariani	LIMSI-CNRS; IMMI	France	Speech technology
Margaretha Mazura	EMF	Belgium	Media and multimedia
Wolfgang Menzel	Universität Hamburg	Germany	LT research
Sukumar Munshi	Across Systems	Germany	Translation workflow support
Bart Noe	Jabbla	Netherlands	Educational software; Assistive technologies
Jan Odijk	Utrecht University	Netherlands	LT research; META-NET
Mehmed Özkan	Medical Institute Bogazici University	Turkey	Biomedical applications

Name	Organisation	Country	Sector
Karel Oliva	Council for Research, Development and Innovation	Czech Republic	
Alexandre Passant	DERI	Ireland	Semantic Web; Social software
Pavel Pecina	DCU/CNGL	Ireland	MT
Manfred Pinkal	Saarland University	Germany	LT research
Stelios Piperidis	ILSP	Greece	LT research; META-NET
Jörg Porsiel	VW	Germany	Translation
Gabor Proszeky	Morphologic	Hungary	NLP; Speech; MT
Artur Raczynski	European Patent Office	Germany	MT
Georg Rehm	DFKI	Germany	LT research; META-NET Manager
Steve Renals	Speech I/O	UK	Speech technology
Peter Revsbech	Ordbogen	Denmark	Translation
Giuseppe Riccardi	Univ. Trento (ex AT&T)	Italy	Speech technology
Johann Roturier	Symantec	Ireland	Localisation
Dimitris Sabatakakis	Systran	France	MT
David Sadek	Institut Télécom (formerly Orange)	France	Telecommunications
Sergi Sagàs	MediaPro	Spain	Broadcasting technologies
Felix Sasaki	DFKI; FH Potsdam	Germany	Standardisation
Jana Šatková	ACP Traductera	Czech Republic	Localisation
Ruud Smeulders	RABO Bank	Netherlands	Financial industry; Mobile services; Translation services
Svetlana Sokolova	ProMT	Russia	MT
Juan Manuel Soto	Fonetic	Spain	Speech analytics; Speech services
Volker Steinbiss	RWTH; Accipio	Germany	LT research; Speech research; META-NET
Daniel Tapias	Sigma Technologies	Spain	Telecommunications voice applications
Alessandro Tescari	Pervoice	Italy	Multilingual speech
Lori Thicke	Lexcelera; Translators Without Borders	France	Translation and language services
Gregor Thurmair	LinguaTec	Germany	MT; Localisation
Hans Uszkoreit	DFKI	Germany	LT research; META-NET Coordinator
Erik Van der Goot	JRC	Italy	Media monitoring
Peggy Van der Kreeft	Deutsche Welle	Belgium	Broadcast media; Innovations expert
René van Erk	Wolters Kluwer	Netherlands	Global information services
Josef van Genabith	DCU and CNGL	Ireland	MT; META-NET
Arjan van Hessen	Telecats; Twente University	Netherlands	Research
David van Leeuwen	TNO; Radboud University	Netherlands	Speech technology
Claire Waast	EDF	France	Automated call centres
Philippe Wacker	EMF	Belgium	Media and multimedia
Wolfgang Wahlster	DFKI	Germany	Artificial Intelligence (AI) research
Alex Waibel	CMU; KIT; Jibbig	Germany; USA	Speech translation for mobile devices
Jakub Zavrel	Textkernel	Netherlands	Document understanding; Web mining; Text matching solutions

Name	Organisation	Country	Sector
Chenqing Zong	Chinese Academy of Sciences	China	LT research

### Members of the Technology Council

The META Technology Council has currently 21 members from leading organisations in the field of language technology. The council takes the recommendations of the vision groups and consolidates them. The result will be a strategic research agenda that expresses the consensus of the stakeholders and communicates the long-term vision of the European language technology community. The following table provides a list of the members of the META Technology Council.

Name	Organisation	Country
Nicoletta Calzolari	CNR	Italy
Claude de Loupy	Syllabs	France
Bill Dolan	Microsoft Research	USA
Yota Georgakopolou	European Captioning Institute	UK
Jan Hajic	Charles University Prague	Czech Republic
Theo Hoffenberg	Softissimo	France
Thomas Hofmann	Google	Switzerland
Elisabeth Maier	CLS Communication	Switzerland
Joseph Mariani	CNRS/LIMSI	France
Penny Marinou	European Union of Associations of Translation Companies	Greece
Stelios Piperidis	ILSP	Greece
Georg Rehm	DFKI	Germany
C.M. Sperberg-McQueen	World Wide Web Consortium (W3C)	USA
Daniel Tapias	Sigma Technologies	Spain
Alessandro Tesconi	Pervoice	Italy
Rudy Tirry	Lionbridge	Belgium
Hans Uszkoreit	DFKI	Germany
Jaap van der Meer	TAUS	Netherlands
Josef van Genabith	DCU/CNGL	Ireland
Michel Vérel	Vecsys	France
Alex Waibel	CMU; University of Karlsruhe; Jibbig	Germany; USA

The strategic research agenda will detail the community’s research focus. The agenda will list main themes and provide a timeline for short-term, mid-term and long-term priorities as well as projects and actions.