



**A Network of Excellence forging the
Multilingual Europe Technology Alliance**

Final Report

Period covered: February 1, 2010 to January 31, 2013

Editor: Georg Rehm

Dissemination Level: Public

Date: July 1, 2013



Grant agreement no.	249119
Project acronym	T4ME Net (META-NET)
Date of latest version of Annex I against which the assessment will be made	Version 3.7 of 8 November 2011 as agreed with the EC services
Project full title	Technologies for the Multilingual European Information Society
Funding scheme	Network of Excellence
Coordinator	Prof. Hans Uszkoreit (DFKI)
Start date, duration	1 February 2010, 36 months
Distribution	Public
Contractual date of delivery	31 March 2013
Actual date of delivery	01 July 2013
Deliverable title	Final Report
Type	Report
Status and version	Final
Number of pages	62
Contributing partners	all
WP/Task Leader	DFKI
Editor	Georg Rehm
EC project officer	Kimmo Rossi (2010 to early 2012: Hanna Klimek)
The partners in META-NET are:	Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI), Germany
	Barcelona Media (BM), Spain
	Consiglio Nazionale Ricerche – Istituto di Linguistica Computazionale “Antonio Zampolli” (CNR), Italy
	Institute for Language and Speech Processing, R.C. “Athena” (ILSP), Greece
	Charles University in Prague (CUP), Czech Republic
	Centre National de la Recherche Scientifique – Laboratoire d’Informatique pour la Mécanique et les Sciences de l’Ingénieur (CNRS), France
	Universiteit Utrecht (UU), Netherlands
	Aalto University (AALTO), Finland
	Fondazione Bruno Kessler (FBK), Italy
	Dublin City University (DCU), Ireland
	Rheinisch-Westfälische Technische Hochschule Aachen (RWTH), Germany
Jozef Stefan Institute (JSI), Slovenia	
Evaluations and Language Resources Distribution Agency (ELDA), France	

For copies of reports, updates on project activities and other META-NET-related information, contact:

DFKI GmbH

META-NET

Dr. Georg Rehm

Alt-Moabit 91c

10559 Berlin, Germany

office@meta-net.eu

Phone: +49 (30) 23985-1833

Fax: +49 (30) 23985-1810

Copies of reports and other material can also be accessed via <http://www.meta-net.eu>

© 2013, The Individual Authors

No part of this document may be reproduced or transmitted in any form, or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission from the copyright owner.

Table of Contents

1	Executive Summary: The Highlights	4
2	Summary: Objectives – Work – Results	4
2.1	Three Lines of Action.....	5
2.1.1	META-VISION: Highlights.....	6
2.1.2	META-SHARE: Highlights.....	7
2.1.3	META-RESEARCH: Highlights	9
2.2	META-RESEARCH: Building Bridges to other Fields	10
2.2.1	WP 1: Bringing more Semantics into Translation	11
2.2.2	WP 2: Optimising the Division of Labour in Hybrid MT	12
2.2.3	WP 3: Exploiting the Context for Translation	14
2.2.4	WP 4: Empirical Base for Machine Translation	16
2.2.5	WP 5: Novel MT Methodology: the Challenges.....	17
2.3	META-SHARE: Building an Open Resource Infrastructure.....	18
2.3.1	WP 6: Infrastructure Preparation and Design.....	19
2.3.2	WP 7: Infrastructure Functional and Technical Specification.....	21
2.3.3	WP 8: Infrastructure Implementation	22
2.3.4	WP 9: Pilot services – Assessment and evaluation	24
2.3.5	WP 10: META-SHARE Support Actions.....	25
2.4	META-VISION: Building an Active Community.....	26
2.4.1	WP 11: Charting the Field.....	27
2.4.2	WP 12: Vision and Community Building	29
2.4.3	WP 13: External Liaisons, Promotion and Mobilisation.....	33
2.4.4	WP 14: Training	36
2.4.5	WP 15: Enlisting Outside Competence for Community Mobilisation.....	37
2.5	Global Work Packages	38
2.5.1	WP 16: Virtual Information Centre	38
2.5.2	WP 17: Coordination	39
3	Socioeconomic Impact.....	39
3.1	Dissemination Strategies.....	39
3.2	Dissemination Activities	40
3.2.1	Conferences, Presentations and Publications.....	40
3.2.2	Community Building.....	40
3.3	Planned Use of Results After T4ME/META-NET	41
4	Deliverables, Publications, Presentations.....	42
4.1	Deliverables.....	42
4.2	Publications	46
4.3	Presentations.....	55
4.4	Conferences, Workshops, Meetings.....	59
4.4.1	Conferences Organised by META-NET	59
4.4.2	National Conferences and Roadshow Events.....	59
4.4.3	Meetings and Workshops Organised by META-NET	60
4.4.4	Conferences in which META-NET Played a Major Role.....	61
4.4.5	Invited Keynote Speeches by META-NET Members.....	62

1 Executive Summary: The Highlights

During its runtime of 36 months – the first funded period of the META-NET initiative – the project T4ME Net (GA 249149) has developed very successfully and had an impressive and significant impact on the field of language technologies in Europe. In its META-RESEARCH line of action ca. 100 scientific publications, language resource as well as software development significantly pushed forward the state of the art in the area of machine translation. Bridges were built to neighbouring scientific fields. In the second line of action, the open resource exchange infrastructure META-SHARE was designed, implemented, set up and populated with more than 2,000 language resources and technologies. META-SHARE is a peer-to-peer network that is driven by more than 20 nodes up and running at the same number of participating organisations. An infrastructure of this kind has been demanded by the industrial and scientific research community for decades – now this gap has finally been closed. In the third line of action, META-VISION, the META-NET White Paper Series “Europe’s Languages in the Digital Age” was written. This huge, unprecedented endeavor which comprises 31 volumes and more than 3,000 pages in total, took more than 200 co-authors about two years to prepare state-of-the-art descriptions of the state of language technology support for 30 European languages including all 24 official languages of the European Union. One of the main findings of this study warns that as many as 21 European languages are in danger of digital extinction. The corresponding press release generated more than 600 mentions in the international press as well as dozens of radio interviews with META-NET representatives as well as television reports. The initiative also prepared the *META-NET Strategic Research Agenda for Multilingual Europe 2020*. This document which also involved ca. 200 co-authors and contributors, specifies three priority research themes for the next decade of Language Technology research in Europe. META-VISION also succeeded in bringing together the highly heterogeneous European LT community by forging META, the Multilingual Europe Technology Alliance. During the course of the project the META-NET Network of Excellence has grown to 60 research centres in 34 European countries which were supported through a total of four funded European projects.

2 Summary: Objectives – Work – Results

During its runtime of 36 months – the first funded period of the META-NET initiative – the project T4ME Net has developed very successfully and had an impressive and significant impact on the field of language technologies in Europe. In the following, we present a summary of the objectives, work and results obtained, structured by the three pillars and lines of action of the project, META-RESEARCH, META-SHARE and META-VISION. T4ME Net started on February 1, 2010 and was finished on January 31, 2013, with a budget of 5.99 MEUR (EC contribution). The project was coordinated by DFKI GmbH (German Research Centre for Artificial Intelligence), Language Technology Lab. Hans Uszkoreit was the coordinator, Georg Rehm the network manager and project manager.

While the initiative had a founding consortium that consisted of 13 partners in 10 countries (the members of T4ME Net), META-NET was supposed to operate on a European level. This is why, in November 2010, the network was extended for the first time. The enlarged network consisted of 44 partners in 31 countries and was presented to the public at META-FORUM 2010. Most of the new members participated in three ICT-PSP projects (CESAR, METANET4U and META-NORD), their mission being to support the META-NET objectives by systematically collecting Language Resources as well as Language Technologies, curating and describing them with metadata records, mobilising the communities in their partners’ countries and general awareness raising activities. After several additional rounds of extending the network, the META-NET initiative grew to 60 research centres in 34 countries.

More information and further reading:

- META-NET website: <http://www.meta-net.eu>
- Members of META-NET: <http://www.meta-net.eu/members>
- META-FORUM 2010: <http://www.meta-net.eu/events/meta-forum-2010>
- T4ME: <http://www.meta-net.eu/projects/t4me/>
- CESAR: <http://www.meta-net.eu/projects/cesar/>, <http://cesar.nytud.hu>
- METANET4U: <http://www.meta-net.eu/projects/METANET4U/>, <http://metanet4u.eu>
- META-NORD: <http://www.meta-net.eu/projects/meta-nord/>, <http://www.meta-nord.eu>

2.1 Three Lines of Action

Speaking one's mother tongue, be it Latvian, Hungarian, or Portuguese must not become a social or economic disadvantage in the networked European society of the 21st century. Language Technology (LT) has the potential to become the key solution to this challenge if it is robust, cost-effective and available for all European languages and to all European citizens. In order to achieve these goals the pace of research and development has to be accelerated by a major, dedicated push. A push on this level and with the magnitude needed can only be put into motion through a joint action of all stakeholder groups involved. This highly diverse group consists of researchers, user and provider industries, technology integrators, language communities, administrators, politicians, journalists and society in general.

META-NET is a Network of Excellence forging the Multilingual Europe Technology Alliance (META) through a concerted effort to build a strong European community around LT. Its goal is to prepare the grounds for applications that enable automatic translation, multilingual information and knowledge management as well as content production

across all European languages. This effort will enhance the development of intuitive language-based interfaces to technology ranging from household electronics, machinery and vehicles to computers and robots. The objective is to advance LT so that communication and cooperation across languages becomes possible and to secure users of any language equal access to information and knowledge. We aim to further LT as a means towards realising the vision of a Europe united as one single digital market and information space. In its first funded phase, META-NET supported these goals by pursuing three lines of action:

1. META-VISION. This action line was concerned with bringing together the fragmented European LT community and with collaboratively preparing, establishing and promoting a Strategic Research Agenda (SRA) for European LT. The SRA is a long-term instrument. It covers the period until 2020 and is meant to be an umbrella for industrial and academic research. The SRA contains recommendations and suggestions for joint actions to be presented to the European Commission and national as well as regional bodies. The complex process of preparing the SRA included Vision Groups that consisted of external experts from industry, who provided ideas for innovative LT applications and scenarios for the future knowledge society. The META Technology Council consolidated their findings into the SRA.

2. META-SHARE. META-NET built META-SHARE, a sustainable network of repositories of language data, tools and web services documented with high-quality metadata, aggregated in central inventories allowing for uniform search and access to resources. Data and tools can be open and with restricted access rights, free and for-a-fee. META-SHARE targets existing but also new and emerging language data, tools and systems required for building and evaluating new technologies, products and services. Reuse, combination, repurposing and re-engineering of language data and tools are crucial. META-SHARE will be an important component of an LT marketplace for researchers and developers, language professionals, as well as for industrial players including SMEs and big enterprises.

3. META-RESEARCH. This line of action carried out innovative Machine Translation (MT) research and built bridges to other fields and disciplines such as Machine Learning and the



META-VISION: Building a community with a shared vision and strategic research agenda

META-SHARE: Building an open resource exchange infrastructure

META-RESEARCH: Building bridges to neighbouring technology fields

Semantic Web community. Central goals were bringing more semantics into MT, optimising the division of labour in hybrid MT, exploiting the context for Translation and preparing an empirical base for MT. Research included, among others, the implementation of prototypes, the preparation of data sets for evaluation purposes and the organisation of internal workshops. This research was meant to advance significantly the state of the art in MT.

More information and further reading:

- About META-NET: <http://www.meta-net.eu/mission>
- META-VISION: <http://www.meta-net.eu/vision>
- META-SHARE: <http://www.meta-net.eu/meta-share>, <http://www.meta-share.eu>
- META-RESEARCH: <http://www.meta-net.eu/meta-research/overview>

2.1.1 META-VISION: Highlights

As a prerequisite for reaching its ambitious goals, the META-VISION arm of the project has started building up a coherent and homogeneous LT community in Europe right from the beginning of the project by bringing together representatives from the highly fragmented and heterogeneous stakeholder groups (researchers, user industries, provider industries, administrators, politicians, integrators etc.). Significant steps towards this goal have been taken through means such as, for example, successfully mobilising ca. 90 external participants for Vision Groups in the three areas Translation and Localisation, Media and Information Services and Interactive Systems and also the META Technology Council with ca. 30 members. Furthermore, META-NET has been engaged in intense dissemination and communication activities in 2010. These include presentations at the FLReNet Forum (February, Spain), LT Days (March, Luxembourg), JIAMCATT 2010 (April, Luxembourg), LREC 2010 (May, Malta), EAMT 2010 (May, France), ICT 2010 (September, Belgium) – according to our estimates we reached more than 2,500 LT professionals and informed them about META-NET’s goals. In addition to the wide visibility provided by successful dissemination and mobilisation activities, META-NET organised three conferences of its own: theMETAnk 2010 was a brainstorming workshop at which about 100 researchers presented and discussed their long-term visions for the field of Human Language Technologies (June 4/5, Berlin). At Translingual Europe 2010 (June 7, Berlin) about 20 presentations were given by researchers and representatives of the provider industries (such as Microsoft, Asia Online, ProMT) and LT/Machine Translation users (European Patent Office, Symantec, EC DGT). The main META-NET event in its first year was META-FORUM 2010 (November 17/18, Brussels, Belgium), at which more than 250 participants were informed about and discussed the initial results of META-VISION, META-SHARE, and META-RESEARCH.

In the second year, META-VISION continued building up a coherent and homogeneous LT community in Europe by bringing together representatives from various stakeholder groups



at various meetings, events and conferences organised by both META-NET and in cooperation with others. META-NET has been engaged in intense dissemination and communication throughout the project’s second year. In 2011, these included (among others) presentations at the W3C Workshop “Content on the Multilingual Web” (April, Italy), FLReNet Forum (May, Italy), Attensity Forum (May, Berlin), Media for All (June, UK), RANLP (September, Bulgaria), the EFNIL Conference (October, UK). In addition to the wide visibility provided by successful dissemination and mobilisation activities, META-NET organised or co-organised several events of its own as well as a number of workshops co-located with other conferences: co-organised TRALAGY 2011 (March,



Paris), a workshop and challenge “Context in Machine Translation” at ICANN 2011 (June, Espoo), the EUROLAN 2011 Summer School (August, Romania), shared task ML4HMT 2011 at the workshop on Using Linguistic Information for Hybrid MT 2011 (September, Spain), and a workshop on “Language Resources, Technologies, and Standards in the Sharing Paradigm” at IJCNLP 2011 (November, Thailand). The main event was, again, META-FORUM 2011 (June 27/28, Budapest, Hungary), which attracted over 300 participants. A satellite event, META-EXHIBITION, featured more than 40 exhibits in total, half of them industry exhibits, the other half posters and demos from collaborating projects. At META-FORUM 2011 delegates met to learn more about the work of the Vision Groups, to discuss the Vision Paper and first SRA draft, and received early editions of the Language White Paper Series. The programme also featured presentations and discussions on META-SHARE and emerging research work in the field.

In the third year, the main focus within META-VISION was on preparing and promoting the Strategic Research Agenda (SRA), for example, by successfully mobilising more than 200 experts from industrial and academic research to participate actively in the discussion and preparation of the SRA. After several iterations and rounds of discussion the document was published online in December 2012. Another important milestone was the publication of the META-NET Language White Paper Series in September 2012. The series consists of 31 volumes reporting on the level of support LT provides for 30 European languages. The corresponding press campaign on occasion of the European Day of Languages (Sep. 26) had a huge impact: it generated more than 600 mentions in the international press, and dozens of television and radio interviews about META-NET’s work and the outcome of the study. Furthermore, META-NET has been engaged in intense dissemination and communication activities. In 2012 these included, among others, presentations at LREC 2012 (May, Istanbul), the Multilingual Web Workshop (June, Dublin), EFNIL 2012 (October, Budapest) and COLING (December, Mumbai). In addition to the wide visibility provided by successful dissemination and mobilisation activities, META-NET organised or co-organised several events: workshop and challenge “Machine Translation and Multimodal Contexts” at ICANN 2012 (September, Lausanne), workshop and shared task ML4HMT 2012 (December, Mumbai) and several workshops at LREC 2012 (May, Istanbul). More than ten dedicated META-NET Roadshow Events were organised by members of our Network of Excellence in countries throughout Europe. The main event of the final year was META-FORUM 2012 (June 20/21, Brussels), which attracted over 250 participants; a satellite expo event featured more than 40 exhibits, many of them from industry and collaborating projects. At META-FORUM 2012 delegates met to learn more about the Strategic Research Agenda (its composition and priority research themes), Europe and its languages (the EFNIL perspective) and Language Technology made in Europe.

More information and further reading:

- Vision Groups and vision process: <http://www.meta-net.eu/vision>
- META Technology Council: <http://www.meta-net.eu/vision/technology-council-members/all>
- META-NET Language White Paper Series: <http://www.meta-net.eu/whitepapers>
- META-NET Strategic Research Agenda for Multilingual Europe: <http://www.meta-net.eu/sra>

2.1.2 META-SHARE: Highlights

This line of action was concerned with the preparation, design and implementation of META-SHARE, the open resource sharing and exchange infrastructure. Important results of the first year were a first, fully functional prototype of the META-SHARE platform, the solid foundations of a metadata schema to be used for the formal description of Language Resources and Technologies and a collection of functional requirements for the system that was distilled out of interviews with a high number of relevant users. Moreover, the landscape of language resources licensing has been explored thoroughly and, with the help of legal experts, a first set of licensing templates has been adopted and prepared. META-SHARE favours and aligns itself with the open data and open source movement, especially the Creative Commons and Free Open Source Software initiatives. At the same time, META-SHARE was put into the

overall HLT context by delineating its borders and relations to other currently running initiatives as well as drafting and signing collaboration agreements with other EU-funded projects. In parallel, META-SHARE was presented and discussed at relevant events (such as FLaReNet Forum, LT Days, LREC 2010, Baltic HLT 2010), and, naturally, META-FORUM 2010 in Brussels.

In 2011 META-SHARE developed on all fronts, including crystallising such finer points as: the language resources metadata model which describes the resources shared, the functionalities of the platform itself, the legal framework, and the basic operational procedures of the infrastructure. Versions 1.0 and 1.1 of the META-SHARE software have been implemented and rolled out to the META-NET network, setting up 13 repositories and populating them with language related datasets and processing tools. These developments follow a revised work-plan to accommodate the needs of the extended META-NET network. Three working groups – on metadata, IPR and legal issues – and the technical infrastructure were established pulling together experts and relevant stakeholders from the wider community.

The working groups formulated initial proposals and specifications for each respective area and discussed and validated them in a series of meetings and workshops. A workshop dedicated to presenting, discussing and offering a hands-on experience to META-SHARE was organised in Athens in October 2011. Work on the metadata model that was initiated in the first year, had been completed, catering for all resource and media types of language data and tools. Mapping tools from existing schemas to the META-SHARE schema have been developed,



focusing first on creating a mapping from the ELRA catalogue to META-SHARE. META-SHARE v1.0 was demonstrated at META-FORUM 2011. This version together with detailed documentation and a user manual was released in September 2011 and tested by the META-NET network. A release version 1.1 was later added to accommodate the requirements of the ICT-PSP projects, in view of their first population round in November 2011. By the end of November 2011, 13 distinct META-SHARE repositories had been installed, with a total of 1277 resources. META-SHARE was presented at the FLaReNet Forum, IJCNLP 2011, LTC 2011, EFNIL 2011, a wide range of workshops, national and regional events, and naturally META-FORUM 2011.

In 2012, META-SHARE focused on refining and widely deploying the language resources (LRs) metadata model, the functionalities of the platform, the legal framework and the basic operational procedures of the infrastructure, in tandem with the implementation of versions 2.0 and v3.0 of the META-SHARE software, their releases to the META-NET network and the greater public, the setup of over 20 repositories and their population with language related datasets and processing tools. The metadata model has been updated to v3.0 accommodating adjustments on specific components and elements following extensive user feedback. META-SHARE versions 2.0 and 3.0 were implemented, including several intermediate iterations as follows: v2.0 was released in March 2012, followed by an improved v2.1 and a v2.9 beta pre-release that was demonstrated at META-FORUM 2012. Constantly in contact with the META-SHARE user base in order to get feedback, additional or improved features of core functionality were implemented and the underlying distributed framework was completed. A user management system was created and integrated, including different user roles and privileges tailored to LRs documentation, curation and sharing. For all major versions, tools were created for easily migrating an existing META-SHARE node to the latest version. META-SHARE v3.0 was released in October 2012 and demonstrated at various national LT

Final Report

Days at the end of 2012 and in early 2013. By the end of January 2013, over twenty distinct META-SHARE repositories had been installed, with a total of over 1,900 LR. Three dedicated helpdesks have been in operation since October 2011 to support LR providers and consumers in all their META-SHARE related activities. META-SHARE has been presented in relevant fora, such as LREC 2012, a wide range of workshops and national as well as regional events, including META-FORUM 2012 and the 2nd META-NET Network Meeting in early 2013. In the period from end January 2013 until the time of writing, META-SHARE has expanded featuring 34 organisation members running 27 repositories and offering more than 2050 LR packages and tools.

More information and further reading:

- META-SHARE: <http://www.meta-share.eu>
- Architecture: <http://www.meta-net.eu/meta-share/architecture>
- Metadata model: <http://www.meta-net.eu/meta-share/metadata-schema>
- Licenses and legal framework: <http://www.meta-net.eu/meta-share/licenses>
- Memorandum of Understanding: <http://www.meta-net.eu/meta-share/memorandum-of-understanding>
- META-TRUST AISBL: <http://www.meta-trust.eu>

2.1.3 META-RESEARCH: Highlights

In the first year of the project, META-RESEARCH was concerned with collecting data, preparing data sets and language resources, compiling inventories of tools and methods, and organising workshops as well as advanced training events for its staff members. Among the major outcomes are the systematic identification of issues in Machine Translation in which semantics has shown potential to positively impact the state of the art, recommendations for approaching the problem of integrating semantic information in MT, and a list of tools and resources that could be employed for this purpose. A new language resource for MT, the Annotated Hybrid Sample MT Corpus, has also been bootstrapped in the first year. It provides data for the language pairs English–German, English–Spanish and English–Czech. A third important outcome is software for the collection of multilingual hidden-web corpora. The tool clusters news articles from different languages discussing the same topic or event and clusters pages identified as being translations of each other.

In the second year, META-RESEARCH made further progress in four key areas, “Semantics in MT”, “Hybrid MT”, “Context in MT”, and “Novel use of large data resources”. This work has produced several high quality publications, the most visible ones being those presented at WMT 2011 in Edinburgh. Moreover, META-RESEARCH has completed several crucial datasets in the area of semantically annotated parallel and multilingual resources: the parallel Czech-English treebank and large automatically annotated parallel corpus (100M words). These datasets have been released under the Creative Commons license and are available through META-SHARE. Work continued to develop hybrid methods in translation. Intensive effort has been devoted to the preparation of the ML4HMT Challenge. This was a first for this area of research in bringing such a challenge to the Machine Learning community. Work has also continued on the preparation of the Context in Machine Translation Challenge, organised jointly with ICANN 2012. Furthermore, work on hierarchical phrase-based translation models enriched with syntactic and dependency features has continued. Results have been presented at the IWSLT workshop in December 2011.

In the third and final year, META-RESEARCH followed the course and directions of work in the four key areas mentioned above. There have again been about 70 publications in 2012 and early 2013. The most visible ones appeared at the WMT 2012 workshop, at COLING 2012 and ICANN 2012. META-RESEARCH also contributed several datasets, mostly related to the two challenges organised by several META-RESEARCH sites: ML4HMT organised in 2011/2012 which resulted in several papers presented at the ML4HMT workshop at COLING 2012, and the second one, organised as part of the ICANN 2012 conference which draws the Machine Learning community. Work has also continued on the use of semantics in Machine Translation, and in the area of applying heavy ML methods to the task of deep parsing and/or translation using such deep analysis. The hybrid MT work has been accepted for

presentation at the AAAI 2013 Spring Symposium conference in San Francisco, another successful step towards building bridges toward the AI and ML communities.

More information and further reading:

- Overview: <http://www.meta-net.eu/meta-research/overview>
- Publications: <http://www.meta-net.eu/meta-research/publications>
- Events: <http://www.meta-net.eu/meta-research/events>
- Training materials: <http://www.meta-net.eu/meta-research/training>

2.2 META-RESEARCH: Building Bridges to other Fields

The goal of pillar 1 was to build bridges to other fields of RTD with the aim of utilising these bridges for pushing the frontiers in Machine Translation technology. The targeted progress was planned to be achieved through the import of valuable new techniques, through the development of new methods and insights through continued cross-area cooperation and through developing and practising a systematic collaborative mode of research based on sharing of existing and new resources, of evaluation methods and of methods for extending technologies to other languages and applications. Like the other two pillars, pillar 1 consists of five work packages, four of them (WP1 to WP4) being research work packages devoted to four different “lines of attack” to the outstanding issues in machine translation, and WP5 being a package aimed at organising open challenges to initiate and to assess progress.

WP1 (“Bringing more semantics into translation”) started with mapping the field and producing an overview of semantic methods in current MT systems. The partners involved also continued an own line of research in this area, mostly in preparing resources and performing initial experiments with anaphora resolution and named entities, both of which have significant impact on current problems in MT. In year 2, WP1 produced several high quality publications, the most visible ones being those presented at WMT 2011 in Edinburgh. Moreover, several crucial datasets in the area of semantically annotated parallel and multilingual resources were prepared: the parallel Czech-English treebank and the large automatically annotated parallel corpus (100M words). These datasets have been released under the CC license and are available through META-SHARE. In year 3, WP1 concentrated on seven areas of different test cases for exploring the use of semantics in MT. Moreover, several datasets and software have been completed and made publicly available through META-SHARE. Each of the seven test cases resulted in one or more publications.

WP2 (“Optimising division of labour in hybrid MT”) first concentrated on data preparation, with four languages processed by six different hybrid systems and on the preparation and running of the workshop on hybrid systems, held in Barcelona in October 2010. In year 2, WP2 has continued to develop hybrid methods in translation. Intensive effort has been devoted to the preparation of the ML4HMT Challenge. There are also several publications in this domain regarding hybrid systems. Based on the experience with developing the challenge, it is however not clear how successful the engagement of the Machine Learning community in this enterprise will be. In the final year, WP2 has continued concentrating on hybrid methods in translation. The second ML4HMT Challenge and an associated workshop at COLING 2012 in Mumbai were organised.

WP3 (“Exploiting the context for translation”) first concentrated on the development of the Challenge on using context in MT for the ICANN 2011 conference. After a restructuring of the management of this work package in year 1, work has continued on the preparation of the Challenge, to be organised jointly with ICANN 2012; the task was to use contextual knowledge in selecting the best translation from an n -best list of candidates. Both WP2 and WP3 produced many publications, also thanks to the submissions to the challenges organised by these work packages. Data from the challenges were published and are now available to the community for the replication and further advancement of the technologies developed. WP4 (“Empirical base for machine translation”) has the work divided into two tracks: collecting data for large-scale machine learning experiments, and the actual work on these experiments. A software prototype for the collection of parallel and comparable corpora was built for large-scale crawling of the web to obtain data usable in these experiments. In year 2,

WP4 has continued its highly progressive work on hierarchical phrase-based translation models enriched with syntactic and dependency features. Results were presented at the IWSLT 2011 workshop. In the final year, WP4 has continued to concentrate on such hierarchical phrase-based translation models enriched with syntactic and dependency features. There has also been a significant number of scientific papers published in WP4.

The goal of WP5 was to help organise the challenges. Since the work has been mostly carried out by and within the work packages responsible for the challenges, there was considerably less work in WP5 than originally foreseen.

More information and further reading:

- Overview: <http://www.meta-net.eu/meta-research/overview>
- Publications: <http://www.meta-net.eu/meta-research/publications>
- Events: <http://www.meta-net.eu/meta-research/events>

2.2.1 WP 1: Bringing more Semantics into Translation

Objectives: This work package tested the hypothesis that semantics can help increase accuracy of MT systems. In order to do so, the three partners involved first made an inventory of semantic tools suitable for inclusion into current MT systems, and then selected one direction of exploration each. The selected systems and semantic tools were then integrated and the contribution of semantics finally evaluated.

2.2.1.1 Task 1.1: Inventory of Semantic Tools and Resources for MT

The goal of Task 1.1 was to survey existing tools and resources suitable to investigate new methods to integrating semantic knowledge into MT. This was achieved by mainly surveying recent literature and by setting up a web page in the intranet in which to organise the inventory of semantic resources and tools. Major outcomes are the clear identification of issues in MT in which semantics has shown potential to positively impact the state of the art, some recommendations and lessons learned about how to approach the problem of integrating semantic information in MT, and a list of almost ready-to-use tools and resources that could be employed to investigate the case studies planned for Task 1.2. We expect the inventory to be updated on a regular basis. The two major outcomes of this work package are a report titled “Reviewing Recent Efforts on Integrating Semantics into MT” and an open web-repository collecting bibliographic references, language resources and tools related to the integration of semantics into statistical MT.

2.2.1.2 Task 1.2: Test cases

The seven test cases are described in detail in deliverable D1.2, “Report on test cases and their evaluation and the resulting publications”. Some of the test cases have been relatively simple and focused on one feature or phenomenon, others (such as the TectoMT system, developed in the course of the whole project) relatively complex. All test cases have been evaluated in the context of an MT system and a language pair for which resources (mainly textual parallel data and appropriate annotations) existed. Some have actually been aimed at a novel metric, where adequacy would be the major component to fully evaluate the addition of semantics into MT. There has been a total of ten publications on these topics. In the case of the TectoMT system, several datasets have been made publicly available, which have been created during the whole timespan of the project and supported also by other projects.

2.2.1.2.1 Bilingual Latent Semantic Modelling

A bilingual latent semantic topic model was further developed to adapt the language model of a phrase-based SMT engine. Adaptation has been based on small contexts of the source sentences that may not necessarily reflect an entire domain or genre. We also developed a fast alternative to Minimum Discrimination Information-based language model adaptation for SMT. This is an alternative to computing a normalization term that requires computing full model probabilities (including back-off probabilities) for all n-grams. The approach has

been tested on the IWSLT 2012 TED talk translation task and showed that Lazy MDI provides comparable language model adaptation performance to classic MDI.

2.2.1.2.2 Data Collection for Adequacy-Oriented MT Evaluation

T4ME has supported the subjective evaluation of the Workshop on Spoken Language Translation (IWSLT) 2011, with the aim of collecting a large sample of subjective evaluations of MT outputs. Subjective rankings were crowd-sourced through CrowdFlowers. In total, pairwise comparisons for four translation tasks were collected, for a total of 30,000 pairwise judgements, which constitute a significant dataset for MT evaluation experiments. During 2012, the collected data were used to investigate novel ranking methods for SMT system evaluation, based on different tournament structures and the collected human judgments were packaged and released to the scientific community.

2.2.1.2.3 MT Adequacy Evaluation without References

Following the project carried out during the MT Marathon 2011, in 2012 we targeted three major challenges for automatic machine translation evaluation: avoiding the use of reference translations, assessing the target sentence against the source, and targeting the adequacy dimension of the problem. The method, solely based on bilingual features extracted from the MT input and output, reported high correlation with human judgements.

2.2.1.2.4 Hybrid Language Models for MT Style Adaptation

The objective of this activity was to improve language modelling by combining syntactic and lexical semantic information. During 2012, this approach was further improved, integrated in the IRSTLM toolkit, documented on the web, and published.

2.2.1.2.5 Modified Distortion Matrices for Phrase-Based SMT

A novel method has been developed to suggest long word reorderings to a phrase-based SMT decoder. The proposed techniques were tested on Arabic-English and German-English using well-known SMT benchmarks.

2.2.1.2.6 Improved Word Alignment Models

IBM schemes estimate bilinear probabilities by combining weighted co-occurrence counts. These have the correct dimension, but do not measure word correlation. Using dimensional analysis, we found another combination – with an additional Dice coefficient factor – of the right dimension that does measure word correlation. Our experiments indicate that alignment probabilities adopt Student's t distributions with predictable parameters.

2.2.1.2.7 Using Approximation of Semantic Features and other Features in SMT

We have continued work mainly on the semantically-heavy English-Czech TectoMT machine translation system, the results of which have been tested during the WMT 2012 competition and reported at an ACL workshop in Montreal. Also, work continued on appropriate feature extraction at the semantic level. Data published in 2012 related to English-Czech MT; four new/extended tools and/or datasets have been published in 2012.

More information and further reading:

- The TectoMT system with description: <http://ufal.mff.cuni.cz/tectomt>
- Browser for the PCEDT 2.0 parallel treebank: <http://ufal.mff.cuni.cz/pcedt2.0>

2.2.2 WP 2: Optimising the Division of Labour in Hybrid MT

Objectives: The two main objectives were to provide a systematic investigation and exploration of possible choices in Hybrid MT, in order to provide optimal support for Hybrid MT design using sophisticated machine-learning technologies, and to build bridges to the ML community to systematically and jointly explore the choice space for Hybrid MT.

2.2.2.1 Summary and Overview of Activities

The partners involved in WP2 worked tightly together in a very cooperative manner to tackle in depth one of the most cited issues in MT, i.e., hybrid approaches to machine translation. It appears, as the results of this WP have confirmed, to be a much more difficult and broader issue than originally anticipated.

The results of this work package are described in detail in deliverable D2.3. Here, we only summarise the activities. WP 2 focused on Machine Learning for High Quality Machine Translation, in particular Hybrid Machine Translation, including system combination methods and approaches that exploit metadata annotations provided by machine translation systems, such as translation probabilities for SMT systems or syntactic/morphological (etc.) information provided to syntax-based or syntax-enhanced MT systems (such as RBMT and HPB-SMT). WP 2 research was structured around the ML4HMT “Applying Machine Learning to Optimise the Division of Labour in Hybrid MT” workshop and the “Optimising the Division of Labour in Hybrid MT” shared task series (both run by WP 2), including data preparation (Spanish → English, Chinese → English), running the WS and ST, competing in the ST, and conducting the ST evaluation. The second WS and shared task ML4HMT-2012 took place at COLING-2012 in Mumbai, India, and is documented as a COLING-2012 Workshop Proceedings publication. WP 2 produced multiple research papers (WS proceedings, conference papers and system description papers) on data preparation for multi-engine MT challenges, software support for human evaluation of MT, as well as novel methods for choice selection for hybrid MT, enlarging the hypothesis space in Minimum Bayes Risk decoding in system combination based MT, exploiting rich alignment information in multi-engine MT, sentence level quality estimation for MT combination, domain adaptation and joint binarised feature vectors for multi-engine MT.

WP 2 built on the observation that alternative approaches to Machine Translation (MT) such as Statistical Machine Translation (SMT) and Rule-Based Machine Translation (RBMT) have different and often complementary strengths (and weaknesses). The objective of WP 2 was to systematically explore hybrid approaches combining the best outputs or capabilities from different MT approaches within a combination-based approach, construed as a multi-engine (MEMT) based approach, or as combining different approaches within a single system (e.g., enhancing a hand-crafted RBMT system with statistics to select translation output from a set of possible outputs). The key idea was to do this systematically in the setting of a shared task, where a training, development and test data-set is provided that, in addition to the source and target strings, includes the output translations of a number of MT systems (representing main MT paradigms such as PB-SMT, RBMT, HPB-SMT, etc.) as well as metadata information from those MT systems for each of the translations and the segments making up longer translations. The metadata annotations include, e.g., phrase translation probabilities and alignment information for the PB-SMT systems and syntactic/morphological or transfer tree information for the RBMT systems. The key idea (and challenge) was to use this information (surface text + rich metadata information) with sophisticated Machine Learning (ML) approaches to explore the space of possibilities in hybrid MT. As far as we are aware, this is the first data set of its kind reported in the literature combining surface text, MT outputs and rich and diverse metadata information. WP 2 organised two workshops/shared tasks (WS/ST) on “Applying Machine Learning to Optimise the Division of Labour in Hybrid MT” (ML4HMT-11 and -12) in Barcelona (Spain) and Mumbai (India), co-located with the International Workshop on Using Linguistic Information in Hybrid Machine Translation (LIHMT-2011) and COLING 2012. WP 2 published a total of 29 research papers (including ML4HMT WS/ST Proceedings and a large number of novel approaches exploring the ML4HMT data sets. This number does not include papers by external non-WP 2 teams participating in the ML4HMT WS/STs.

In summary, however, the research shows that it is extremely challenging to fully exploit the rich metadata information provided with the ML4HMT data sets as the metadata are extremely diverse and often difficult to compare and relate across systems. Furthermore, one of the objectives of WP 2 was to build bridges to the ML community. Apart from a few limited

exceptions, this was not achieved, as the ML4HMT WS/STs did resonate with the MT community but not the ML community.

2.2.2.2 Resources

The data preparation was completed and the data are available for download. The corpus has twice the size (2,000 sentence pairs with translations from five MT systems and meta-data annotations) than originally foreseen.

2.2.2.3 Software

Apart from the technical research work, WP 2 supported the development of the Appraise MT human evaluation software platform.

More information and further reading:

- Machine Learning for Hybrid MT workshop and shared task: <http://www.dfki.de/ml4hmt/>
- Appraise: <https://github.com/cfedermann/Appraise>

2.2.3 WP 3: Exploiting the Context for Translation

Objectives: This work package aimed at exploring the possibilities for using wider, larger, and/or non-traditional “context” in which an utterance is translated. Since it was not expected to develop sufficient data for large-scale experiments, this WP aimed at one challenge (by developing a novel data set that can be used for both training and evaluation). Two workshops for presentation of discoveries in the use of context in MT were to be organised.

After initial delays and a change of management, this WP made very good progress and provided good results. In the second year the first workshop was organised and used to present and advertise the challenge. In the third year the WP has continued to develop methods for taking context into account, e.g., domain-dependence, domain classification and translation model adaptation; modelling context as an additional dimension in vector space; modelling context through reranking methods, word-sense disambiguation, and document-level translation models. Context-based reranking methods applied to the challenge data did not lead to improvements in translations as measured by BLEU. It is not clear whether this was due to properties of the challenge data, of the tested methods, or of the evaluation metric.

The first planned workshop was dedicated to the discussion and announcement of the challenge. It was held in Finland in June 2011, attached to ICANN 2011. Challenge data was prepared with contributions by the three WP 3 partners. The challenge schedule was spread over 2011 and 2012 to give more time to external participants to prepare systems. The second workshop, which closed the challenge, announced its results and included presentations of the participants, was part of ICANN 2012. Both workshops took place in venues attended by the Machine Learning community, since a key goal of WP3 was to attract the attention of ML researchers to statistical MT topics.

There have been a number of publications produced by WP 3. Also, a lot of effort has been devoted to the organisation of the challenge and its two workshops. A number of resources have been created. The SRILM Toolkit has been extended to classify documents according to domain, it is available under the SRILM Research Community License.

2.2.3.1 The Context in Machine Translation Challenge

The Context in MT Challenge involved the publication and advertisement of the challenge and the preparation of the datasets. Given n -best lists produced by statistical machine translation systems, the challenge task was to select the best translation candidate for each source segment with the help of contextual information.

Data preparation required to select and to preprocess parallel data, to prepare domain descriptors and MT system outputs. The selected corpus consisted of an English-Finnish subset and a Greek-French subset of JRC-Acquis. Parallel data from JRC-Acquis was sentence-split, cleaned-up and tokenised. We developed several MT systems for the French-Greek and

Finnish-English language pairs and prepared the corresponding lists of n -best hypotheses to be used in the challenge. Additional n -best lists were prepared with the help of other META-NET partners. Altogether we now have 100-best lists produced by statistical machine translation systems. To represent the domain context, we prepared EUROVOC document descriptors for each source segment. We assembled the whole data, including other segments of each source text, to be used as source text context. Annotated training corpora were made available to the challenge participants.

2.2.3.2 Domain Modelling as a Dimension of Context for Translation

We investigated the use of domain classification based on domain-specific language models to take the domain context into account for MT.

An algorithm for partitioning JRC-Acquis was designed and implemented. A number of techniques were developed to make use of information that had to be extracted by the context. The approach of using EUROVOC's wealth of information provided us with very satisfactory results. This approach uses a simple supervised classification algorithm. We tried out a multitude of algorithms to classify an incoming document into the translation system. Naive-Bayes approaches led us to good – but not perfect – results. We trained several models, one for each predefined domain. Then, we calculated the perplexities of each new incoming document against each of these models. The assumption was that the one that would yield the lowest perplexity would be the most suitable one. This assumption was proven correct in all of our test cases (100%). Language models were 4-gram models.

A translation model was built for each domain. These were used to translate incoming documents. When a new document enters the system, it is classified and translated. This process entails choosing the most suitable translation model(s). We planned to use existing techniques for combining the models, and if needed, develop new ones. The processing pipeline does the following: it takes a plain text file as input, classifies it, translates it with the corresponding translation model and then evaluates whether this translation was the best possible, given the predefined domain models.

2.2.3.3 Context-Based, Near-Synonym Lexical Choice

The lexical choice task involved filling in gaps in a text with words that best fit the context. The task can improve many applications involving natural language generation, for instance, statistical machine translation, question-answering, summarisation, text simplification, and terminology adaptation. We examined to what degree extensive linguistic analysis of the context could be utilised in a near-synonym lexical choice task. Supervised and unsupervised classifiers were trained with the analysed morphological, semantic, syntactic, and extralinguistic features. A large set of linguistic features is not required in the selection task, but the features need to be relevant. Classification accuracy of unsupervised models is greatly reduced by the information unrelated to the task even though the algorithms can easily handle large numbers of features. However, supervised models, which carry out feature extraction implicitly, can ignore unrelated information, but the algorithms can have computational problems with a large number of features. Semi-supervised learning or supervised feature selection with a lightweight classifier can be seen as a solution to alleviate these problems, but it requires labelled training data. Automatic feature selection outperformed manual work as it produced a smaller subset of the features than the manually selected subset, and produced more accurate classifiers.

2.2.3.4 Resources Created and Published

The following resources have been collected and/or annotated and processed. They will be made available under a Creative Commons licence in 2013: Greek-English portion of JRC-Acquis partitioned into domains; Domain-adapted translation models for partitioned JRC-Acquis; Domain-adapted language models for partitioned JRC-Acquis.

2.2.4 WP 4: Empirical Base for Machine Translation

Objectives: The goals of this work package were to address traditional and innovative multilingual corpora collection techniques, to address innovative machine learning for machine translation methodological issues and to address multilingual information retrieval and multilingual question answering.

WP 4 prepared multilingual data and also several tools and technologies for tasks related to multilingual applications, including MT and multilingual IR.

2.2.4.1 Task 4.1: Multilingual data collection

The main deliverable was D4.1, “Software for the collection of multilingual hidden-web corpora”. The system consists of two modules: The first module collects comparable texts from news articles by exploiting information from news aggregators, RSS feeds and crawling news sites. The system clusters news articles from different languages discussing the same topic or event. The second module is for the identification of translated pages from multilingual web sites. The system clusters pages identified as being translations of each other. In a follow-up step, the identification of multilingual articles discussing the same topic was implemented, i.e., a solution for producing parallel corpora from multilingual websites located on a single web server. High precision for the returned parallel documents was required while recall was of secondary importance.

2.2.4.2 Task 4.2: Machine learning methodological issues

The WP4 partners worked on methods that allow bilingual phrases to be trained explicitly using forced alignments on the training data. That is in contrast to the state-of-the-art methods where bilingual phrases are only extracted but not trained. We also extended our hierarchical system to include syntactic dependencies to be included in the log-linear model combination as an additional syntactic feature model. Two types of machine learning techniques were used. First, structured learning (MMR, Sinhue, CRF) and second, language models for smoothing using semantically enriched text.

We extended a hierarchical phrase-based machine translation system to use the information of dependency trees of the target language. We studied two versions to employ a dependency language model: by reranking on n -best lists and by direct integration into the translation process. We achieved improvement on the NIST Chinese-English task in terms of BLEU score from 32.0 to 32.6 using reranking and to 32.9 using direct integration.

We also investigated two new methods for hierarchical phrase-based machine translation. The first extension introduces discriminative lexicalised reordering into hierarchical machine translation. We compared different feature sets for the discriminative reordering model and investigated combinations with three types of non-lexicalised reordering rules which are added to the hierarchical grammar in order to allow for more reordering flexibility during decoding. We achieved improvements in translation quality on a large-scale Chinese-to-English translation task. Furthermore, we developed a new approach to learn phrase pairs for hierarchical machine translation. On the total corpus of training data, we perform two-stage parsing to learn better phrase probabilities. In addition, phrase pairs with a low score are discarded. This results in a smaller translation model and improves translation quality. Some of these methods are implemented in the open-source translation toolkit Jane 2.1.

2.2.4.3 Task 4.3: Analytic based multilingual methods

Task 4.3 is about analytic methods for multilingual information retrieval. We experimented with a strategy to build a general-domain “encyclopedia” of concepts described by a collection of multilingual labels, populated by mining Wikipedia articles. We also developed a web application for editing generic multilingual language resources in a collaborative way by exploiting external resources such as Wikipedia, OmegaWiki and SKOS thesauri.

We worked on a method to automatically identify linguistic contexts which contain possible causes of emotions or emotional states in Italian, combining relevant linguistic patterns and

an incremental repository of common sense knowledge on emotional states and emotion eliciting situations. The results are satisfying and support the validity of the methodology proposed. We also produced a repository of event nouns with associated weighted polarity values. In particular, we are able to improve SentiWordNet values assigning to event nouns scores compatible with their connotations. In terms of multilingual sentiment detection, the possibility to discriminate between objective and subjective expressions contributes to the identification of a document's semantic orientation. Subjectivity word sense disambiguation helps in the task automatically to determine which word senses in a corpus are being used subjectively and which are being used objectively.

We also worked on developing a scalable representation for statistical cross-lingual text mining for multiple languages, targeting up to 100 languages. The central data structure developed within this effort is language neutral document representation which allows the mapping of documents from multiple languages into a statistical intermediate representation allowing comparing documents (cross-lingual similarity function), document classification into a single classification schema and document clustering. Special emphasis was put (a) on the scalability of the approach, (b) decreasing the cost of extending language neutral document representation to a new language, and (c) allowing cross-linguality for languages with small comparable corpora. Introducing the concept of the "hub language" on top of Wikipedia comparable corpora for the languages with more than 10,000 documents enabled linear extra cost in terms of adding a new language in the schema, and to connect languages with small or zero cross-lingual explicit links.

More information and further reading:

- Jane software extension: <http://www-i6.informatik.rwth-aachen.de/jane/>
- [A Cocktail of Deep Syntactic Features for Hierarchical Machine Translation](#)
- [Soft String-to-Dependency Hierarchical Machine Translation](#)
- [Forced Derivations for Hierarchical Machine Translation](#)

2.2.5 WP 5: Novel MT Methodology: the Challenges

Objectives: The goals of this work package were to go beyond the current state of the art in MT technology and to motivate targeted, exploratory research in areas where insufficient progress has been made recently.

WP 5 was the reserve work package for supporting shared tasks. As opposed to the original plans, WP 5 was downsized in terms of support funds following the update of the Description of Work in year 1. The challenges, which took part in year 3, i.e., the "Context in Machine Translation Challenge and Workshop" and the "Applying Machine Learning Techniques to Optimise the Division of Labour in Hybrid MT Challenge and Workshop" are described in detail in deliverable D5.1. The document contains details on the data produced by these challenges, which are now publicly available, and on the papers (total of 19 papers resulting from these two challenges) and results of these challenges. Here, we only briefly introduce them.

2.2.5.1 Context in Machine Translation Challenge and Workshop

Statistical Machine Translation is generally addressed as a sentence-translation task, i.e., a task in which each sentence in a text is translated independently of the rest of the text. The goal of the "Context in MT Challenge" was to study how the context in which a sentence is found can be used in the SMT process to help improve the translation of this sentence. This was approached by preparing a dataset where sentences are provided in context and formulating the task as a reranking task based on the outputs of several existing SMT systems. Given n-best lists produced by SMT systems, the challenge task was to select the best translation candidate for each source segment with the help of contextual information, comprising textual context and domain descriptors. The challenge results were negative in the sense that we did not evidence improvements in MT when using context as prepared in our datasets, and in the lack of external participants. It is not clear whether this was due to properties of the

challenge data, of the tested methods, or of the evaluation metric. We draw lessons in the conclusion of the respective deliverable, with recommendations for future similar attempts.

2.2.5.2 Applying Machine Learning Techniques to Optimise the Division of Labour in Hybrid MT Challenge and Workshop

The workshops and shared tasks are summarised as follows: Can hybrid MT and system combination techniques benefit from extra information (linguistically motivated, decoding, runtime, confidence scores, alignment information or other meta-data) from the systems involved? The objective of the shared tasks is to test whether it is possible to incorporate highly heterogeneous “semantic” information in terms of run-time and alignment information from component MT systems into hybrid MT or system combination based approaches using Machine Learning algorithms, leaving the choice of which metadata information to use (or indeed whether to use the metadata information at all) to the participants of the shared tasks. In an open track, participants were able to use additional metadata information not provided by the shared task. Many submissions availed of this possibility. As a baseline we used the single best translation among translation outputs. The baseline systems did not use any run-time “semantic” metadata provided by any of the component MT systems in the annotated data-set.

More information and further reading:

- Context in Machine Translation 2011: <http://www.cis.hut.fi/icann2011/con-txt-mt11/>
- Machine Translation and Multimodal Contexts 2012: <http://research.ics.aalto.fi/cog/MTMC2012/>
- Machine Learning for Hybrid MT workshop and shared task: <http://www.dfki.de/ml4hmt/>

2.3 META-SHARE: Building an Open Resource Infrastructure

The goal of Pillar 2 was to design, implement and put in use an open resource sharing infrastructure for language data and tools that support research, development and evaluation in the Human Language Technologies domain. This infrastructure, now widely known as META-SHARE, was successfully established as a network of distributed LRs repositories, devoted to the sustainable sharing, dissemination and increased access to LRs in a global scale.

Pillar 2 was structured in four technical work packages, WPs 6-9, and an ancillary WP 10 for support actions. WP 6 and WP 9 took care of preparing, designing, populating and providing services through META-SHARE, while WPs 7 and 8 took care of the technical specification and implementation of the infrastructure.

In the three years of the funded period of META-NET, META-SHARE went from rough conceptual ideas to an implemented, fully-fledged distributed infrastructure, ready for production work. Among the main tasks were a requirements study, the specification and implementation of the infrastructure as well as its constant testing and refinement, the design and implementation of the metadata model underlying all provider and consumer oriented functionalities, the preparation and testing a coherent legal framework for data sharing, the installation of META-SHARE nodes and their connection to the META-SHARE network collection and uploading of language resources, the setup and running of helpdesks and a web-based user forum to assist META-SHARE users in all dimensions of META-SHARE operations. During all these stages, members of the wider META-NET network, especially partners of the four consortia CESAR, METANET4U, META-NORD and T4ME, have been testing and providing feedback. The software code of META-SHARE is available as open source from github under a BSD licence. In mid 2013, META-SHARE is fully operational with more than 30 organisation members and 20 (27 at the time of writing) distributed repositories. More than 2,050 language resources are currently available through META-SHARE. The infrastructure was presented at dozens of national and international scientific meetings and conferences. In addition to the ongoing administration and maintenance work regarding operating the META-SHARE infrastructure we are currently in the process of integrating META-NET’s legal entity META-TRUST AISBL, an international non-profit organisation under Belgian law, into the legal framework of META-SHARE.

More information and further reading:

Final Report

- META-SHARE: <http://www.meta-share.eu>
- Metadata model: <http://www.meta-net.eu/meta-share/metadata-schema>
- Architecture: <http://www.meta-net.eu/meta-share/architecture>
- Charter: <http://www.meta-net.eu/meta-share/charter>
- Memorandum of Understanding: <http://www.meta-net.eu/meta-share/memorandum-of-understanding>
- Licenses and legal framework: <http://www.meta-net.eu/meta-share/licenses>
- META-SHARE software code: <https://github.com/metashare/META-SHARE/>
- User Forum: <http://www.meta-share.org/portal/forum/questions/show/all/newest/all>
- META-TRUST AISBL: <http://www.meta-trust.eu>

2.3.1 WP 6: Infrastructure Preparation and Design

Objectives: The goal of this work package was to design the open, distributed META-SHARE infrastructure in terms of its basic principles, services and functionalities, together with the associated quality and operational requirements, legal framework, governance and business models, as well as sustainability and development plans. The results fed the work in WPs 7, 8 and 9 where the actual infrastructure was specified, implemented, populated, deployed and assessed.

2.3.1.1 Task 6.1: Management of the “Building an Open Resource Infrastructure” Pillar of T4ME – Working groups setup

Task 6.1 took care of the overall monitoring and internal management of the pillar and provided the general operational setup and work plan. Among the work carried out was defining the basic principles of META-SHARE, agreeing on priorities in terms of datasets, tools and technologies, providing input for the technical work in work packages 7 and 8, and kick-starting the development of an early prototype version 0 of the META-SHARE software using an educated guess approach as to potential user requirements, while in parallel running a user requirements elicitation study. This task was also responsible for coordinating the work between T4ME and its sister projects CESAR, METANET4U and META-NORD, funded through the ICT-PSP programme. Three working groups on metadata, legal issues and implementation work were established with representatives of the above mentioned four EU-funded projects and the wider community. The working groups were heavily involved in the preparation, definition and design of the metadata specifications, in the completion of the legal framework, and in providing new versions of the infrastructure. In addition, three helpdesks were set up, on metadata-based LR/LT documentation, on legal issues and IPR clearance and on technical support for platform installation and use.

2.3.1.2 Task 6.2: Mobilisation of the Community

The objective of Task 6.2 was to promote META-SHARE and create consensus around it. Many initiatives have been undertaken for the mobilisation process and for an early involvement of a large community, both in its active involvement in the discussion of the basic principles and issues of META-SHARE, and in promotion and dissemination of the META-SHARE infrastructure. META-SHARE was discussed and presented at a multitude of national and international meetings, workshops, conferences, exhibitions and projects villages. While a full list of presentations can be found at the end of this report, two conference series shall be highlighted here: intense dissemination activities for and around META-SHARE have taken place at LREC 2010 (Malta) and LREC 2012 (Istanbul) as well as at META-FORUM 2010 (Brussels), META-FORUM 2011 (Budapest) and META-FORUM 2012 (Brussels) including software demos, tutorials, papers, presentations and workshops.

2.3.1.3 Task 6.3: META-SHARE design, organisation and service model

Task 6.3 defined a proper governance, organisation and service model for META-SHARE. Several design options were outlined, considering various pros and cons. Following the design adopted, META-SHARE was specified as a network of distributed repositories of lan-

guage resources, including language data and basic language processing tools. Language resources are described according to the META-SHARE metadata schema (see WP 7). Actual resources and their metadata reside in local or hosting (non-local) repositories set up and maintained by network members. Each repository undertakes the responsibility to export metadata records and allow their harvesting. A certain subset of network members undertakes the role of managing nodes committed to offering core and user support services. Managing nodes harvest metadata from all network nodes, synchronise and share them in a p2p-like fashion. Central network servers (hosted by managing nodes) point to local repositories for browsing and downloading actual resources. A Memorandum of Understanding was prepared, version 2.0 of the MoU was released in January 2013 as a result of wide consensus and thorough discussions between the consortium, the legal experts group and the community. The MoU specifies all the details related to META-SHARE membership, lists its current members and lists all core and user support services offered. It describes the META-SHARE governance structure catering for specific provisions as to the roles of the Network and Managing Nodes and the details of operation within META-SHARE. In addition, the Terms of Service for registered and non-registered META-SHARE users were published. Finally, given the firm intent of the European language technology community, as assembled and organised in META-NET, to use and to maintain META-SHARE beyond the period of funding of the T4ME project, a sustainability plan for the period up until the end of 2014 has been prepared, agreed upon and released. The sustainability plan describes tasks, responsibilities and committed effort to maintain the repositories, the inventories of metadata and the technical, legal and user support services throughout the period 2013-2014. The plan is supported and implemented by META-NET/META-SHARE members through their own investment or through their participation in LT-related R&D research and development projects.

2.3.1.4 Task 6.4: Connection of META-SHARE with existing repositories, other projects, initiatives and institutions

Task 6.4 grouped together actions of liaison with relevant organisations and existing or emerging initiatives, with the aim of establishing agreements for the sharing and re-use of readily or soon-to-be-made available LR through META-SHARE. In this task we established connections with more than 40 EU-funded projects and larger initiatives such as, for example, CLARIN, FLaReNet, Multilingual Web, Panacea, TTC, Monnet and Accurat. With almost all of these projects collaboration agreements were drafted and signed. These agreements specify the mode of collaboration and cooperation as well as the expected outcomes. Of primary importance for META-SHARE is that many of these initiatives volunteered or agreed right away to make any language resources and technologies, emerging through their projects, available through the META-SHARE infrastructure. In addition, in the framework of these international fora, several concertation actions were organized including i.a. common workshops with the CLARIN Research Infrastructure on metadata modeling and use of public sector information for LT, bridges with initiatives like the LRE Map and the Language Library as well as the adoption of persistent identification schemes (e.g. PIDs, ISLRN) for language resources.

2.3.1.5 Task 6.5: IPR and Legal issues

All legal issues relating to licensing terms and conditions for the use of language resources were taken care of in Task 6.5. Initially, a workshop was organised at LREC 2010, in which legal experts from ELRA, LDC, Creative Commons, as well as researchers reviewed current practices in data sharing, identified trends and priorities, and elaborated on the diversity of legal protections in Europe. Clear trends towards open data and software licensing were identified. Following this evidence and discussions within META-NET and the wider community at numerous international events, a Charter (the Language Resources Sharing Charter) advocating the adoption of openness for language resources was produced. With the help of legal experts and after several rounds of discussion within the enlarged META-NET network of excellence and also the wider LT community, a four-tier licensing scheme for lan-

guage resource sharing and distribution were produced, that are aligned with the standardised options offered by Creative Commons (e.g., attribution – non commercial use – share alike) and the Free Open Source Software initiatives. The META-SHARE model licensing scheme accommodates comments, requirements, and specific needs expressed by the community but also during the negotiation of specific resources on the part of META-NET network members. Three layers of templates were produced for licensing of language data, plus related legal documents and one-page simplified guides detailing their use and suitability for specific licensing conditions; all documents are publicly available. For licensing of language processing tools (i.e., software), existing standard licenses (GPL, LGPL, BSD, Apache, etc.) already used by the providers are adopted. Licensing options and access rights are formalised in a dedicated mandatory component of the metadata model. Currently we are in the process of including META-TRUST AISBL into the rather complex legal framework as a trust centre.

2.3.1.6 Task 6.6: META-SHARE Constitution, Management Structure

Here we surveyed the different business models that exist in the area of LR production and exchange as well as other data and content aggregation initiatives, social networks, analysed the potential involvement of the European Commission and national governments and tried to identify revenue streams and cost structures leading to a sustainable strategy within a certain period of time. Work for this task included a review of the models of 14 main entities active in this area: ELRA, LDC, NICT (ALAGIN and Language Grid), TAUS/TDA, TST-Centrale, CLARIN, Appen, BAS/BASSS, CRN-CNRTL/CRDO, GSK, OLAC, Speech Ocean and Chinese LDC. As reported in Task 6.3, META-SHARE is organized as a, community-supported, network of distributed repositories. The different types of membership and their respective rights and obligations, the services and the conditions under which these are offered by different types of repositories, the light governance structure as well as the very easy ways to join META-SHARE are presented in the MoU v2.0, available from the metanet.eu website.

More information and further reading:

- META-SHARE: <http://www.meta-share.eu>
- Architecture: <http://www.meta-net.eu/meta-share/architecture>
- Charter: <http://www.meta-net.eu/meta-share/charter>
- Licenses and legal framework: <http://www.meta-net.eu/meta-share/licenses>
- Memorandum of Understanding: <http://www.meta-net.eu/meta-share/memorandum-of-understanding>
- META-TRUST AISBL: <http://www.meta-trust.eu>
- Collaborations: <http://www.meta-net.eu/collaborations>

2.3.2 WP 7: Infrastructure Functional and Technical Specification

Objectives: WP7 provided the user requirements, functional specifications and design of META-SHARE at the service level as well as specifications for the description of LRs and LTs to be offered through META-SHARE. Envisaged services target academic and commercial LR/LT producers, aggregators as well as integrators and users. The resource-level specification surveyed and specified types of LR/LT to be made available through META-SHARE and described a powerful and yet flexible, agreeable and currently broadly used metadata model.

2.3.2.1 Task 7.1: Service-level specification

This task was devoted to the user requirements, functional specifications and the design of META-SHARE at the service level, together with their evaluation. The specifications were passed to and implemented in WP 8. User requirements were collected through interviews with representatives of the stakeholder groups involved (for example, consumers/researchers/users, providers, aggregators and data centres). The interviewees also filled in a questionnaire about their profiles. The recorded interviews were transcribed, structured, distilled and summarised in terms of general desiderata, accessing META-SHARE, registra-

tion, searching, uploading etc. This analysis led to the specification of the major features of META-SHARE, together with the anticipation of major problems to solve or risks involved. In collaboration with the META-SHARE metadata working group, user requirements for resource description and search were checked and taken into account in the schema specification. Cooperation with the implementation team (WP 8) was ensured throughout the specification-implementation cycle, providing feedback on incremental versions of the platform.

2.3.2.2 Task 7.2: Resource-level specification

This task surveyed and specified the types of LR and LT to be included in the infrastructure, developed a metadata schema and identified mappings between different schemas.

First, an in-depth survey of existing metadata schemas allowed us to see specific elements currently in use or recommended by standards as well as requirements and preferences on the part of LR distributors. The overview took into consideration the most widely used metadata schemas in the relevant areas: ISO 12620 (ISO DCR), CES/XCES, TEI, OLAC, IM-DI, ENABLER, BAMDES, DCMI, ELRA Catalogue and Universal Catalogue, LREC map, LDC catalogue, CLARIN's Virtual Language Observatory, Natural Language Software Registry and LT World. This comparison revealed multiple similarities and differences. These results were used for the initial formulation of the META-SHARE LR and LT typology and the metadata schema. These metadata descriptions constitute the means by which users will identify the resources they are looking for. The schema forms an integral part of the META-SHARE search and retrieval mechanism, with a subset of its elements serving as the access points to the catalogue.

After several iterations and many rounds of discussion, the metadata schema was finalised. It includes elements and mechanisms to describe written and spoken corpora, multimodal corpora, sign language corpora, lexicons, language processing tools and services, conceptual resources, language descriptions. An external group of metadata experts from Europe, the USA and Asia was involved at multiple stages of this finalisation process in order to guarantee broad consensus and applicability of the schema. The schema was presented in multiple conference papers. It includes mechanisms for a minimal and maximal schema as well as mandatory, recommended, optional components and elements and also a full LR and media typology. The ICT-PSP projects helped significantly in the finalisation and validation of the metadata schema when they prepared their LRs/LTs for inclusion in META-SHARE. Version 3 of the metadata model, formally realized as an XML Schema specification, is implemented in a corresponding metadata editor in META-SHARE's web interface and documented in a user manual as well as in a section of the META-SHARE portal devoted to the documentation of the model, with detailed presentation of all components and elements. For mapping purposes, anchors to the Dublin Core and the ISOcat DCR have already been included in the documentation of the schema, and mapping tools have been created for conversion from and to the ELRA schema, the OLAC schema and CMDI.

More information and further reading:

- META-SHARE: <http://www.meta-share.eu>
- Metadata model: <http://www.meta-net.eu/meta-share/metadata-schema>
- User Forum : <http://www.meta-share.org/portal/forum/questions/show/all/>
- Knowledge base : <http://www.meta-share.org/portal/knowledgebase/home>
- Architecture: <http://www.meta-net.eu/meta-share/architecture>

2.3.3 WP 8: Infrastructure Implementation

Objectives: The goal of this work package was to implement and to deliver the META-SHARE infrastructure and associated services as described and specified in WP 7.

First planning efforts were undertaken at the pillar 2 kick-off meeting in Athens, Greece, in March 2010, followed by a dedicated technical meeting at LREC 2010 in Malta. META-SHARE is implemented as a network of distributed repositories. The system comprises a set of managing (core) nodes that harvest and replicate, synchronise, metadata from the distrib-

uted LR repositories and their inventories, effectively building a fail-safe p2p-like network.. Each of the managing nodes allows users to browse the inventory, query, access and download language resources that can either be served from the local inventory or from the digital repository at a network node. META-SHARE was implemented using the Django framework which builds upon the Python programming language. The code is available from github under the open source BSD license.

The development team used Scrum as an agile software development approach. Intermediate goals were agreed for “sprints” of two to four weeks, at the end of which the team aimed to deliver “increments of potentially shippable functionality”. The adoption of this approach was a turning point for the team’s efficiency: quantity, quality and compatibility of the output produced by the different parties have all improved substantially since this decision was taken. A number of best-practice software engineering methods were also adopted: the use of github as the version control repository; unit tests and continuous integration for automatically verifying that a certain functionality, once it is achieved, is maintained; collaborative code review and refactoring. Several versions were released:

- META-SHARE v0: demonstrated at META-FORUM 2010
- META-SHARE v1.0: demonstrated at META-FORUM 2011; released in Sept. 2011
- META-SHARE v1.1: released to the META-NET Network, November 2011
- META-SHARE v2.0: first open-source release, March 2012
- META-SHARE v2.1: much improved release from the v2.x series
- META-SHARE v2.9: pre-release of v3.0, July 2012
- META-SHARE v3.0: released in October 2012; demonstrated at various events

2.3.3.1 Task 8.1-2: Development platform and implementation of core functionality

The core functionality of META-SHARE includes an integrated web application and repository based on Django, implementing the META-SHARE metadata model including loss-less import from XML and export to XML. Redesigns of the metadata model could be reflected in the META-SHARE implementation right away through a generic solution which generates the Django object-relation model automatically from the XML Schema that specified the metadata model (see WP 7). Among the core functionalities are also synchronisation of metadata records within the distributed framework of managing and non-managing nodes, and registration, authentication and authorization for user management. During synchronisation the actual downloadable resource data are not shared because of potential size and legal restrictions; however, when a user attempts to download a resource on a node which is not the original resource provider, then he is automatically redirected to the original node providing the resource. Additionally, a user management system on the provider side was created and integrated. The system provides different user roles and privileges tailored to the task of sharing resources and collaboratively describing them. For all major versions of META-SHARE tools were created for easily migrating an existing META-SHARE node and its data to the latest release version.

2.3.3.2 Task 8.3: Implementation of functionality for consumers

Consumer-oriented functionality was implemented so that consumers get a user-friendly and powerful user experience. This includes browsing through the whole inventory of language resources, filtering them by a number of criteria, searching either in a simple Google-like form or using a more detailed advanced, faceted search fully supported by the underlying metadata model; and finally, the presentation of a detailed record describing a language resource and, for all openly shared language resources, its download. We also present several statistics concerning the number of views and downloads of a specific resource as well as recommend similar resources to the user, based on the view and download logs.

2.3.3.3 Task 8.4: Implementation of functionality for providers

Functionality for language resource providers includes all aspects required to document and provide resources, especially the metadata editor as well as import and export of XML representations of resources in the metadata format defined in WP 7. While the metadata editor was already functionally complete with META-SHARE version 2.0, a lot of effort was invested over the following months to improve its usability and to eliminate bugs.

In addition to resource descriptions, META-SHARE hosts downloadable resource packages. In the last development period we worked on resource package hosting on META-SHARE nodes. In order to accommodate different resource package sizes, different ways for uploading such packages to a META-SHARE node have been provided.

2.3.3.4 Task 8.5: Deployment and support

Throughout the development cycles of the last project period, the development team was in contact with the META-SHARE user base in order to get feedback on the implementation. This has greatly helped to improve the overall usability of the system.

Especially after major releases, many support requests regarding the setup and maintenance of META-SHARE nodes were answered. In order to efficiently manage such support requests, different ways of contacting the development team were provided: a help desk mailing list, an online user forum and the GitHub issue tracker. In addition, a manual and an online knowledge base were made available. In order to make META-SHARE more visible for resource seekers, common search engine optimisation (SEO) techniques to the system were applied. Interested node administrators could monitor the success of their META-SHARE node installations using a new interface for adding web analytics software to their nodes.

More information and further reading:

- META-SHARE: <http://www.meta-share.eu>
- META-SHARE software package : <https://github.com/metashare/META-SHARE/>
- User Forum : <http://www.meta-share.org/portal/forum/questions/show/all/>
- Knowledge Base : <http://www.meta-share.org/portal/knowledgebase/>
- Feedback on github : <https://github.com/metashare/META-SHARE/issues>
- Documentation on github : <https://github.com/metashare/META-SHARE/tree/master/misc/docs>

2.3.4 WP 9: Pilot services – Assessment and evaluation

Specific services help providers to populate META-SHARE with their language resources. They are specified in WP 7, implemented in WP 8 and deployed in WP 9. Services were deployed for providers and consumers. Providers can add resources in a local or in a hosting (non-local) repository. Providers can manage their resources (e.g., editing, importing, adding LRs to the networked repositories, etc.) following the rights they have been assigned. Consumers can browse and search resources in repositories, including most of the basic functionalities (view, download, etc.).

In addition, three helpdesks were set up to provide support. They cater for the legal, metadata and technical issues involved in setting up the repositories and making available well documented LRs. They have been widely used and mainly helped the PSP partners to install their own repositories and import resources.

2.3.4.1 Task 9.1: Pilot services

Specific services help providers to populate META-SHARE with their language resources. They are specified in WP 7, implemented in WP 8 and deployed in WP 9. Services were deployed for providers and consumers. Providers can add resources in a local or in a non-local repository. Providers can manage their resources (e.g., editing, importing, adding LRs to the networked repositories, etc.). Consumers can browse and search resources in repositories, including most of the basic functionalities (view, download, etc.).

In addition, three helpdesks were set up to provide support. They cater for the legal, metadata and technical issues involved in setting up the repositories and making available well doc-

umented LRs. They have been widely used and mainly helped the PSP partners to install their own repositories and import resources.

2.3.4.2 Task 9.2: Infrastructure population

First, the META-SHARE core partners set up their own repositories (CNR, DFKI, ELDA, FBK and ATH/ILSP). Then, organisations participating in the three ICT-PSP projects installed new repositories, either one repository per single project or several repositories, one per project partner. As soon as these repositories had been installed, they were populated with resources. For other providers who have chosen not to install their own repository, a hosting repository was installed by ELDA.

At the time of writing, META-SHARE is composed of 27 repositories. It includes seven synchronised Managing Nodes, harvesting between four and six other repositories. The network provides to the users more than 2,050 resources.

2.3.4.3 Task 9.3: Assessment and evaluation

We continuously assessed and monitored META-SHARE, the partners of the ICT-PSP projects were especially helpful in this regard. This approach was directly in line with the providers' perspective of META-SHARE, which is its main objective for this phase of development. Many of their comments and suggestions have been taken into account for improving the infrastructure, most notably the editor for documenting LRs.

More information and further reading:

- META-SHARE: <http://www.meta-share.eu>
- Resources: <http://metashare.dfki.de/repository/search/>

2.3.5 WP 10: META-SHARE Support Actions

Objectives: The goal of WP 10 was to prepare, coordinate, control and report on specialist work carried out within working groups, or by means of external consultations, or through field experts, which underpinned the network's main activities: community building and mobilisation; broadly-based dissemination and awareness efforts; exchanges and consensus building activities; launch, promotion and population of META-SHARE.

2.3.5.1 Task 10.1 Legal Services

A successful workshop was organised at LREC 2010 on Legal Issues for Sharing Language Resources: Constraints and Best Practices. Legal experts from ELRA and LDC, experts from the Creative Commons community, as well as LT researchers reviewed current practices in data sharing, identified trends and priorities, and elaborated on the diversity of legal protections in Europe and the rest of the world. With the help of two legal experts subcontracted by the project – Prodromos Tsiavos and John Hendrik Weitzmann – the Language Resources Sharing Charter, the four-tier licensing framework for datasets and software, guides for the use and implications of the different licenses, and the META-SHARE Memorandum of Understanding were prepared, tested and validated. The legal experts also helped when it came to selecting the best license for a certain language resource and provided consulting and expertise in other issues related to legal questions that came in via email or through the legal helpdesk and the user forum. They also gave crash courses and tutorials on the terminology, purported meaning and entailments of the legal provisions stipulated in the licensing templates.

2.3.5.2 Task 10.2 Technical services

External support was sought for the design of the META-SHARE user interface, user forum and knowledge base, with the goal of providing a simple, functional and user-friendly tool to accommodate the needs of all META-SHARE user types.

More information and further reading:

- Charter: <http://www.meta-net.eu/meta-share/charter>
- Licenses and legal framework: <http://www.meta-net.eu/meta-share/licenses>
- Memorandum of Understanding: <http://www.meta-net.eu/meta-share/memorandum-of-understanding>

2.4 META-VISION: Building an Active Community

The main objective of pillar 3 was building and strengthening a homogeneous multi-stakeholder community that includes actors from research, industries, language communities, public R&D funding and other groups. The final goal was a strategic and open technology alliance that involves researchers, enablers, providers as well as actual and potential users of all technologies that are needed to secure sustainable multilingualism within the European society. This strategic alliance is META, the Multilingual Europe Technology Alliance. For META to be successful, it was and still is important to have support by a large community of relevant public and private stakeholders. In order to identify and to reach this heterogeneous and fragmented group of stakeholders, different strategies were implemented. Pillar 3 was structured into five work packages.

WP 11 examined the current state of the art in HLT with respect to the languages of the EU and the associated goals of META-NET. This WP produced a comprehensive and truthful map of the current state of technology for all relevant languages of the EU and selected regional and third country languages. A key result was the publication of the META-NET Language White Paper Series in September 2012 to which more than 200 experts contributed.

WP 12 focused on the preparation of a major concerted effort geared towards the needed technological base for the European multilingual information society. The task of planning such a joint action was a highly complex endeavour that involved all major groups of actors and stakeholders right from the start. The central instruments were the network itself, the META Technology Council and three Vision Groups. At the end of the process, the META-NET Strategic Research Agenda for Multilingual Europe 2020 (SRA) was prepared, along with first versions of associated roadmaps. The SRA was published online in December 2012 and, as an Open Access document, with the publishing house Springer in January 2013.

WP 13 planned and coordinated all external communication activities. This WP was concerned with drafting and maintaining a communication plan, producing promotion material, initiating collaborations with other research centres and projects, liaising with several industries and the organisation of dissemination events. Among the main results of this WP were the three META-FORUM conferences organised in 2010, 2011 and 2012.

WP 14 dealt with the preparation of teaching and training materials needed for the dissemination of META-NET's goals. To this end, several training events were organised.

WP 15 was dedicated to networking activities involving external competence. An external subcontractor was subcontracted to support META-NET with regard to selected external communication activities, especially concerning identifying and communicating our key messages to stakeholders in Brussels and Luxembourg.

More information and further reading:

- Vision Groups and vision process: <http://www.meta-net.eu/vision>
- META Technology Council: <http://www.meta-net.eu/vision/technology-council-members/all>
- META-NET Language White Paper Series: <http://www.meta-net.eu/whitepapers>
- META-NET Strategic Research Agenda for Multilingual Europe: <http://www.meta-net.eu/sra>
- SRA Roadmaps: <http://www.meta-net.eu/sra/roadmaps>
- META: <http://www.meta-net.eu/meta/about>
- Members of META: <http://www.meta-net.eu/participating-organisations/logos>
- META-FORUM 2010: <http://www.meta-net.eu/events/meta-forum-2010>
- META-FORUM 2011: <http://www.meta-net.eu/events/meta-forum-2011>
- META-FORUM 2012: <http://www.meta-net.eu/events/meta-forum-2012>

2.4.1 WP 11: Charting the Field

Objectives: For the planning process leading to the Strategic Research Agenda and for the mobilisation of the language communities and other stakeholders, a comprehensive and truthful map of the current state of technology for all relevant languages of the EU was required, where the class of relevant languages in this context encompasses at least all official EU languages but preferably also selected regional and third country languages. This map was meant to consist of two parts, a detailed structured inventory of all technologies, including applications and resources that make up the technological base, and a comprehensible description of the state of the technology for each individual language.

2.4.1.1 Task 11.1: The inventory and META Matrix

First, the structure and content of the META Matrix were worked out. The matrix should show the comparative status of various languages regarding LRs/LTs (like the Euromatrix or the BLARK matrix). The META Matrix was partially based and bootstrapped with the content of the first LREC Map, initiated by the community for LREC 2010. The LREC Map represents the LRs used by the authors of LREC 2010 papers, as mentioned by themselves in a questionnaire they had to fill in when submitting papers. This resulted in ca. 2,000 entries, mentioning close to 1,600 different LRs. In addition, a tool was designed that produces the META Matrixes. They address the 23 official languages of the EU, the European non-EU languages (such as Norwegian, Serbian, Croatian) and the regional European languages (such as Basque and Catalan). After bootstrapping the META Matrix, records from other conferences were added, for example, COLING 2010 and EMNLP 2010; agreements with the organisers of other conferences, journals and catalogues were drawn to include their data as well (e.g., IJNLP, ACL-HLT, Interspeech).

When preparing the Language White Paper Series (see below), the META Matrix was especially used to identify the language coverage in terms of data, tools, evaluation and meta-resources (standards, metadata, guidelines), to set priorities for developing the LRs which are missing for certain languages, and also to identify trends in terms of new tools for addressing emerging technologies and innovative applications.

Initially, a set of eight matrixes, including links to the corresponding LRs, was produced for 3 modalities and 160 LR types: spoken language data (7 types) and tools (19 types), written language data (24 types) and tools (78 types), multimodal/multimedia data (8 types) and tools (11 types), evaluation resources (5 types) and meta-resources (guidelines, standards, metadata; 8 types). This initial analysis was later on extended to several regional languages (Asturian, Basque, Catalan, Galician), some non-European languages (Arabic, Hindi, Japanese, Korean, Chinese Mandarin) and to various sign languages.

We produced a second version of the META matrix based on a total of 4,300 records from LREC 2012 and nine other leading conferences. Again, 8 matrices were built. The matrix covers 216 languages. An additional matrix was built for 21 sign languages. The analysis confirms the supremacy of (American) English for all categories (26% overall). However, it appears that LRs for other major languages are increasing, while several previously considered 'minor' or regional languages have benefited from a strong support to recover. Some (e.g., Bulgarian, Estonian, Polish, Slovene) more than tripled the number of resources.

2.4.1.2 Task 11.2: Language White Papers

At least for the 23 official EU member state languages, the consortium wanted to prepare reports that explain "the state of each language with respect to language technology". In the documents, the existing technologies, including applications and resources, should be summarised with respect to better, equally and less resourced languages. They should also discuss shortcomings and gaps. After exploring several different approaches, it was decided to follow the idea to produce longer documents that cover, in addition, a broader set of topics such as, for example, societal aspects, dangers and challenges. The T4ME consortium received support from the three ICT-PSP projects CESAR, METANET4U and META-NORD

who were also tasked with the mission of providing Language White Papers for the languages these three project consortia covered.

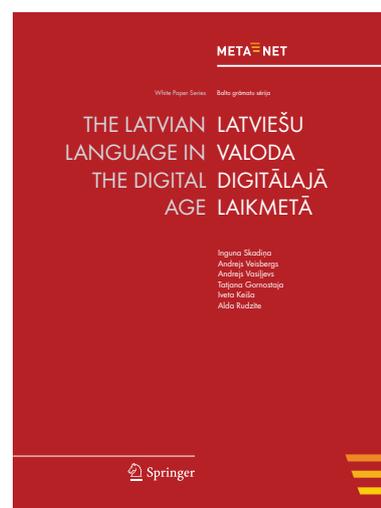
All in all, the process of preparing the language white papers turned out to be much more complex than originally foreseen. It took us more than two years to go from the first ideas, potential document outlines and structures to the 31 final volumes. A total of 209 authors and contributors provided expertise and input to writing the white papers which were meant to be an important regional/national communication instrument for politicians and journalists. First, a structure and generic template was designed and discussed within the T4ME consortium. This template was used to prepare the first five white papers. It included several rather abstract language-independent parts (e.g., the introduction) as well as language-specific parts (e.g., origin and syntax of the language in question). The two main parts of each white paper were the state of the respective language in the digital age and the state of language technology support for the respective language.

The idea was to write the papers in an interesting way that is able to attract journalists and politicians to the general topic of multilingualism and language technology as an important means to overcoming language barriers in Europe. Specifically, the white papers include an assessment of existing technologies and resources, shortcomings and gaps, as well as a cross-language comparison of language technologies and language resources. In order to conduct and finalise this comparison, a meeting was organised in Berlin in October 2011 involving about 90 experts representing all languages covered by the series. It allowed preparing comparative tables for the 30 languages regarding four main areas of language technology: Spoken Language Processing, Written Language Processing, Machine Translation and Language Resources.

After several iterations of revising and refining the structure and contents, authoring teams for all languages involved were assembled and work on the first drafts started. Initially, the white papers were written in English, later on to be translated into the language they report upon. The META-NET White Paper on German was used as the template for all other reports. After finalising the white papers, high-quality paper copies were to be produced in order to ensure wide distribution to non-specialists, journalists and politicians. The “META-NET White Paper Series: Languages in the European Information Society” was published with Springer in September 2012. PDF versions are publicly available on the META-NET website and also through SpringerLink. The very first drafts were publicly unveiled and distributed at META-FORUM 2011 in Budapest (June 27-28, 2011). For the print publication the production pipeline was changed: MS Word was replaced by a LaTeX-based approach, which guaranteed high-quality print and PDF online versions.

The White Paper Series was written for the following 30 European languages (including all 23 EU languages): Basque, Bulgarian, Catalan, Croatian, Czech, Danish, Dutch, English, Estonian, Finnish, French, Galician, German, Greek, Hungarian, Icelandic, Irish, Italian, Latvian, Lithuanian, Maltese, Norwegian, Polish, Portuguese, Romanian, Serbian, Slovak, Slovene, Spanish, and Swedish. META-NET purchased ca. 8,500 print copies and distributed them nationally and internationally to decision makers, journalists, politicians, administrators and other stakeholders in September 2012 on the occasion of the European Day of Languages.

The distribution of the Language White Papers was accompanied by a press release entitled “At Least 21 European Languages in Danger of Digital Extinction”, which received a tremendous echo in the press. More than 600 press articles related to this press release have been counted all around Europe and abroad. This activity was accompanied by a dissemination campaign in many countries, which generated several strong political statements, in some countries at the highest level (Presidential, Prime ministerial and ministerial). This first dissemination campaign paved the way for a second one regarding the Strategic Research Agenda. Circa 100 copies of complete box sets of Language White Papers including the print-



Final Report

ed Strategic Research Agenda (ca. 3,000 pages in total and weighing ca. 8kgs) were sent to very important decision makers in each of the covered countries and at the European/international level.

2.4.1.3 Task 11.3: Report on existing projects and initiatives

Based on related work carried out in the framework of FLReNet, a comprehensive report on existing projects, funding programmes and initiative was prepared. An initial FLReNet report served as the basis of the corresponding META-NET deliverable and report, which analyses its content and expands on the programmes and projects that were identified, focusing on those that address machine translation, multilingual technologies, language resources and infrastructures.

2.4.1.4 Task 11.4: Summaries of detected gaps

Based on the first META-Matrix, the findings made when preparing the Language White Papers and the report on existing projects and initiatives, as well as the Language repository catalogues of ELRA and LDC, the analysis of the Basic Language Resource Kits (BLARK) and Euromatrix+, work was started to detect gaps in the current language technology landscape and coverage per language. The definition of “gaps” including the existence of resources as well as their quality (sufficient size, validation, etc. for data, sufficient performance, evaluation, etc. for tools), the existence of products and of programmes of infrastructural nature was worked out and the sources of information were identified.

More information and further reading:

- META-NET Language White Paper Series: <http://www.meta-net.eu/whitepapers>
- Quotes and testimonials: <http://www.meta-net.eu/whitepapers/all-quotes-and-testimonials>
- Press release: <http://www.meta-net.eu/whitepapers/press-release>
- Press coverage: <http://www.meta-net.eu/whitepapers/press-coverage>

2.4.2 WP 12: Vision and Community Building

Objectives: A central objective of META-NET was the preparation of a major concerted effort towards the creation of the needed technological base for the multilingual European information society. An essential instrument was and is the forging of the strategic technology alliance META (Multilingual Europe Technology Alliance) involving, in addition to the top level RTD centres, the active participation of European LT industry and many private and public stakeholders, including the language communities themselves. This is why the task of planning such a joint action was a highly complex endeavour.

The central instruments were the extended META-NET network, the META Technology Council and the Vision Groups. The vision and planning process consisted of seven steps: 1. Arriving at a mutual understanding: Understanding the particular goals and constraints of research, user industries, provider industries, politics/public administration; 2. Performing an analysis: Survey of the current state, identification of gaps; 3. Creating a shared vision: Convergence on goals and identification of obstacles, definition of major challenges including complex challenges; 4. Investigation of means: investigating possible measures and instruments; 5. Defining the strategies: prioritising instruments and actions, converging on a strategy; 6. Drafting a plan: Draft of a Strategic Research Agenda; 7. Feedback and Revision: in one or several revision loops, reactions from many sources will be taken into account.

For META to be a successful strategic alliance with a high impact, it has been of utmost importance to have a common strategic vision supported by a large community of relevant public and private stakeholders from politics, industry, research, and society in general. The vision and community building process was initially coordinated by the META-NET Standing Committee. One of its main tasks was to assemble three Vision Groups, which were a central instrument within META-NET’s vision and community building efforts. The visions produced by the Vision Groups were consolidated by the META Technology Council and worked out into a Strategic Research Agenda, later on to be presented to regional national as well as

international bodies. The community was involved in this process via META-NET events, participation in public events, and employing an online discussion forum (see Task 12.3).

2.4.2.1 Task 12.1: Vision and Planning Process

The main goal of this task was to assemble and organise three Vision Groups, each composed of ca. 25-30 high-level experts representing selected major areas, challenges and players. As META-NET strives to bring Europe's LT industry not only scientifically, but also economically into a leading position, these groups were primarily composed of distinguished stakeholders from industry and business, supplemented by leading researchers from the respective fields. The groups were assembled in an effort to cluster the relevant and most promising European LT-related industry sectors into reasonably sized working groups. Their main objective was to provide input to the Strategic Research Agenda.

Much effort was put into the identification of candidates and the recruiting process of the three groups. First, there have been extensive discussions and decisions regarding the composition of the groups, selection criteria for internal and external members were defined and the invitation process carried out in the first six months of the project. To cater for the best topical, sector and geographical coverage, the selection was based on research vs. industry proportion and even geographic distribution. In multiple rounds of discussions and suggestions by the Standing Committee and members of the consortium, three lists of candidates were compiled and revised. Lists of approximately 50 candidates per group were agreed upon (including basic, backup and consortium members). A target number for participation in the first round of meetings was set to 8-10 recommended members; taking into account the participation of consortium members, this number is extended to 15 participants approximately. An invitation letter as well as a briefing package explaining META-NET's structure, objectives and expected results was produced.

In 2010 and 2011, three rounds of meetings were successfully completed:

1. Vision Group *Translation and Localisation*: 1st meeting: July 23, 2010, Berlin; 2nd meeting: September 29, 2010, Brussels; 3rd meeting: April 7/8, 2011, Prague
2. Vision Group *Interactive Systems*: 1st meeting: September 10, 2010, Paris; 2nd meeting: October 5, 2010, Prague; 3rd meeting: March 28, 2011, Rotterdam
3. Vision Group *Media and Information Services*: 1st meeting: September 10, 2010, Paris; 2nd meeting: October 15, 2010, Barcelona; 3rd meeting: April 1, 2011, Vienna

In advance of these meetings, the Vision Group members were asked to prepare short statements answering questions such as

- What are attractive, plausible, powerful, challenging, innovative LT-based applications or combinations of applications with use cases that could be realised through massive concerted research, development, and innovation actions?
- Can you think of novel research advances in LT that would be needed to support real breakthroughs of type 1?
- What are expected technological, economic or social developments that have to be considered as prohibitive or enabling factors in the planning of 1 and 2?

Most members who prepared written statements agreed to make them available. The results of the discussions held at these meetings were documented by minutes taken by consortium members. All results were worked into reports that distinguish between domain-specific and domain-independent needs, enabling factors, and visions. The reports, which also document the Vision Group assembly and recruitment process, were handed to the members of the META Technology Council to be further processed into one overarching vision paper and the Strategic Research Agenda. While the results of the first two rounds of meetings were domain-specific vision documents, the third round of meetings was used to create a single, combined document containing the most promising and attractive ideas for technology vi-

sions. The results of these nine meetings were presented at, among others, an ICT 2010 Networking Session in Brussels and at META-FORUM 2010 and META-FORUM 2011.

The META Technology Council met in Venice on May 25, 2011, to adopt and finalise the combined vision paper, “The Future European Multilingual Information Society”. It was decided to prepare a generally understandable and appealing document targeted at politicians and journalists. The document was meant to prepare the reader for more concrete and more detailed technology visions presented in the Strategic Research Agenda SRA later on.

2.4.2.2 Task 12.2: The Strategic Research Agenda for a large-scale joint action and a roadmap for establishing the appropriate instruments

The objective of the META Technology Council was to take as input the domain-specific visions and ideas for innovative applications and research areas produced by the three Vision Groups and to consolidate these visions into a single convincing strategy including roadmaps. This Strategic Research Agenda (SRA) was one of the most central results of META-NET.

First, the Standing Committee assembled the initial META Technology Council. For pragmatic reasons, the first set of council members was chosen to be close to Language Technology in expertise. In early 2011 additional members were invited to follow-up meetings in order to enlarge the council’s outreach, network, and sphere of influence.

The inaugural meeting of the Technology Council took place on November 16, 2010, at Hotel Le Plaza in Brussels, one day before META-FORUM 2010. The council members received the three Vision Group reports via email in advance. At the meeting, the Vision Group results were presented by the conveners of the Vision Groups and then discussed. Gaps were identified and further ideas elaborated. The discussion was documented in minutes and condensed in a presentation. Follow-up meetings of the council took place in Venice on May 25, 2011, in Berlin on September 30, 2011, and in Brussels on June 19, 2012.

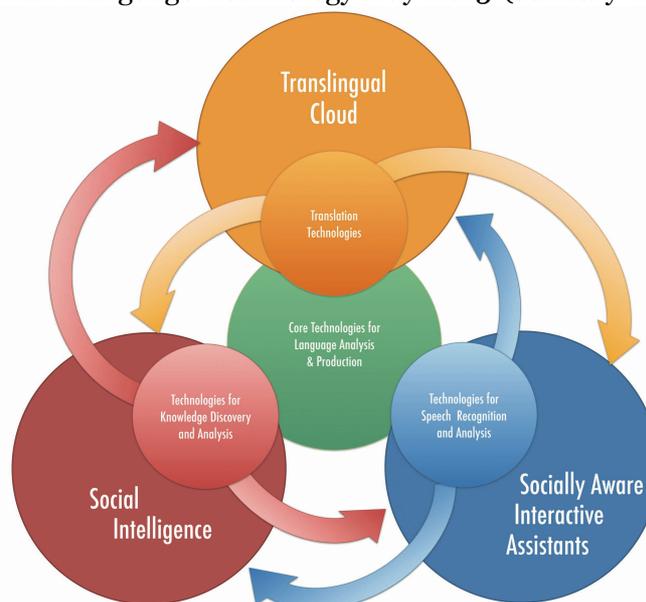
In the third meeting on September 30, 2011, a list of more than a dozen technology visions was extracted from all results of the three Vision Groups. These were further reduced into an even smaller set of lead visions to be included in the SRA. The shortlist of lead technology visions was further discussed and reduced at several smaller meetings and, finally, at the META-NET General Assembly at the end of October 2011 in Berlin. The META-NET visions were also presented by representatives of META-NET at HLT Experts Meetings organised by the EC in the autumn of 2011. After these meetings, the SRA editorial team with support from participants of the HLT Experts Meetings drafted the core of the SRA entitled “LT 2020 Vision and Priority Themes for Language Technology Research in Europe until the Year 2020 – Toward a draft of the META-NET Strategic Research Agenda”, which consolidated the META-NET lead visions into three Priority Research Themes: Translingual Cloud; Social Intelligence and e-Participation; Socially-Aware Interactive Assistants. In the SRA, they build the bridge between societal needs, Language Technology applications, and concrete roadmaps for the organization of research, development and scientific innovation.

To sum up, important milestones in the process towards the SRA include six documents: the three Vision Group reports prepared in 2010, a general Vision Paper (2011), and a Priority Themes Paper (early 2012) in which the visions are specified in a more concrete way. A first draft of the META-NET SRA was presented at META-FORUM 2012 in Brussels in June, while the final version was published online on December 1, 2012 after another round of discussion and feedback (initiated after META-



FORUM 2012) in which we received almost 90 pieces of valuable input and feedback – none of it negative. The SRA was published with Springer in January 2013. The printed version was unveiled and distributed at the German Language Technology Day 2013 (January 24) and at the official META-SHARE and META-NET SRA launch event (January 25) in Berlin. All documents are available online.

The META-NET SRA is not only available online as a PDF document, it is also available as a regular, printed book. We purchased 2,000 copies of the book that we started disseminating in late January 2013. We also purchased 100 complete sets of the META-NET White Paper Series (31 volumes each). For these 100 complete sets we ordered 100 containers through a service provider so that we can either hand over in person, or ship out complete box sets to very important persons or potential supporters.



In order to guarantee maximum visibility and reach, we agreed with Springer that the book would be available under an Open Access license, i.e., a PDF version of the book is available free of charge through SpringerLink. In the respective publishing agreement with Springer the two editors of the SRA transferred all rights to META-NET's legal person, META-TRUST AISBL in Antwerp, Belgium to underline that the SRA is a community-driven effort.

2.4.2.3 Task 12.3: Community Involvement

In order to support the open and broad discussion and vision building process we installed a discussion forum component on the META-NET website. As initial content, the complete Vision Group reports as well as selected results were included for public discussion. To increase the outreach, META-NET set up a Facebook group and a LinkedIn group. Both were also used to announce events and to report ongoing activities.

However, it turned out that general interest in the online discussion on the META-NET website was comparably low despite many different attempts to generate traffic in the discussion forum. This is why the communication strategy was adjusted and the material under development was discussed in various smaller face-to-face meetings, workshops, presentations, skype meetings and phone calls etc. also exploiting the extended possibilities of a larger network. All central results of the META-VISION process including all 31 META-NET Language White Papers, the META-NET Strategic Research Agenda for Multilingual Europe 2020, the individual steps of the vision building process and member lists are available online.

More information and further reading:

- Vision Groups and vision process: <http://www.meta-net.eu/vision>
- META Technology Council: <http://www.meta-net.eu/vision/technology-council-members/all>
- META-NET Strategic Research Agenda for Multilingual Europe: <http://www.meta-net.eu/sra>
- SRA Roadmaps: <http://www.meta-net.eu/sra/roadmaps>
- META: <http://www.meta-net.eu/meta/about>
- Members of META: <http://www.meta-net.eu/participating-organisations/logos>
- META-NET at Facebook: <http://www.facebook.com/META.Alliance>
- META-NET at LinkedIn: <http://www.linkedin.com/groups?mostPopular=&gid=3052556>

2.4.3 WP 13: External Liaisons, Promotion and Mobilisation

Objectives: As a network of excellence seeking a maximal positive impact on the development of our technology area, the success of the project depended heavily on cooperation with many actors in research, economy and society and on powerful instruments for informing and mobilising stakeholders and decision makers on the European and national levels. Besides the relevant R&D communities the main targets of promotion were and are European industry and administration (technology user communities, multilingual technology providers, commercial translation and localisation providers, language/translation Services), member states and language communities (governments, research agencies, language councils and similar bodies), European politics, and administration (European Commission, European Parliament), media and press, as well as the public at large. Hence, the objectives were to create, inform and mobilise a community of relevant stakeholders around META-NET. This was necessary to ensure contribution to the development and acceptance of our goals, research agenda and other recommendations and participation in the community.

2.4.3.1 Task 13.1: Production and Revision of the Communication Plan

A thorough and complex communications plan was produced and being implemented in multiple promotion and mobilisation activities. We also constantly reviewed our activities, channels, media and communication instruments with a view to assessing the efficacy of efforts and the plan. This feedback loop enabled us to improve our efforts in this area and to identify new activities, channels, etc. which can be used to better raise awareness, foster community involvement and engage the wider community.

We also included the members of the three ICT-PSP projects in our communication efforts by assembling a Communication Working Group. Virtual meetings were used to coordinate our national and international communication activities and to better harmonise activities between all META-NET projects. The idea was to produce a push towards using our supporter base as multipliers to convert more stakeholders to supporters and to increase engagement. The communications plan was supported by two supplementary communications plans for the Language White Paper Series and the Strategic Research Agenda PR campaigns. These activities required precise coordination and timing of actions across the entire network and continent. Because of this and the crucial nature of the campaigns they were treated as separate plans, which were drawn up and executed in close conjunction with the coordinators who spearheaded their implementation.

2.4.3.2 Task 13.2: Production of Promotion Material

T4ME (and also the three ICT-PSP projects) produced several promotional materials. First, a multilingual project flier was prepared and translated (the extended network including the ICT-PSP projects have also added to this bank of translated work and a translation of this project literature now exists in each national language and is available for local and international events as necessary). The flier was distributed, in the original and updated form, at community events and conferences targeting various stakeholders throughout the runtime of the project. In addition, several poster designs for conference booths/exhibits were produced as well as stickers and promotional conference banners. We also compiled a number of reusable presentations and other similar materials for members into a communications pack which each member can reuse, translate and adapt as necessary.

In addition, we produced several other promotional materials, among others, four double-sided leaflets covering META-SHARE in general, Open Data, the META-SHARE metadata model and legal aspects. Additional materials and give away items for conferences and events produced include stand up banners, t-shirts, lanyards, mugs, pens, business card holders, laptop sleeves, bookmarks and stickers.

2.4.3.3 Task 13.3: Cooperation with research centres outside the consortium

T4ME, the founding consortium of META-NET, consisted of 13 partners in 10 countries, however, the Network of Excellence was meant to operate at a European level. To this end, the network underwent several rounds of extension. In its current incarnation META-NET consists of 60 research centres in 34 European countries. The biggest extension phase happened after one year when all partners participating in the three ICT-PSP projects CESAR, METANET4U and META-NET became members of META-NET. Furthermore, in that timeframe, the META-NET Statutes and the META-NET Executive Board were established as the network's overarching governance body..

In addition, META-NET entered collaborations with a total of 46 other EU projects such as ACCURAT, CLARIN, FlaReNet, Let's MT, MOLTO, Monnet, Multilingual Web, PANACEA, PASCAL2, PlanetData, Promise and TTC. These projects are from such diverse fields as home automation, in car communications, media mining, social intelligence, and medical information retrieval. Most of these collaboration agreements deal with one or more of the following main topics: making existing and emerging resources and technologies available through META-SHARE; carrying out promotional work together and organising or participating in events; strategic planning activities.

2.4.3.4 Task 13.4: Liaisons to Industries

Industry involvement in META-NET activities was and is very strong. The Vision Groups and the Technology Council feature representatives from a number of large multinational technology companies as well as many European industries, both large groups and also SMEs. Further engagement with industry was achieved through the vision building process where many industry representatives, from both large and small enterprises, have come together with other community experts from research and LT user/consumer groups to help build the SRA. In addition to these information pulling means of liaising with industry META-NET has also taken part in industry conferences such as Localisation World, Attensity Forum, as well as Languages and The Media, and organised joint industry/research events such as Translingual Europe 2010. These events have been successful in spreading the message to large numbers of relevant industry stakeholders from around Europe and the world and have been useful in raising the profile of META-NET's work. Our own META-FORUM conference series (2010, 2011, 2012) also served as a very useful platform for industry liaisons with several sessions dedicated to representatives of the LT user and provider industries. In addition, META-FORUM 2011 hosted META-EXHIBITION. This satellite event attracted over 40 exhibits with approximately 50% industry and 50% research participation from across Europe. We achieved the same number of exhibitions and participation from industry at META-FORUM 2012.

2.4.3.5 Task 13.5: Utilising LT fora for promotion and dissemination

META-NET had a strong representation at a wide variety of LT community events and conferences. The project organised exhibition booths (such as at the EC Villages at LREC 2010 in Malta and LREC 2012 in Istanbul), run networking sessions (such as, for example, at ICT 2010 in Brussels), organised events aiming at gathering LT developers and users (such as Translingual Europe in Berlin and Tralogy in Paris), sponsored events (LREC, IWSLT, EACL, Tralogy), gave keynote addresses (at events such as, for example, HLT: The Baltic Perspective 2010 in Riga, IWSLT 2010 in Paris), workshop talks (Multilingual Web Workshop, Madrid, LREC 2010, FlaReNet Forum) as well as other talks at various events, making META-NET known in Europe and elsewhere. META-NET representatives were invited to address various assemblies, meetings and events beyond the core LT fora, which engage with language communities and consumers of LT and LT products.

2.4.3.6 Task 13.6: The T4ME Prize and the Seal of recognition

The inaugural META Prize and META Seal of Recognition awards were awarded at META-FORUM 2011. These awards recognise outstanding products and services supporting and contributing to the European Multilingual Information Society. The winners were determined by the META Technology Council based on the findings of WP 11, input from the Vision Groups, and nominations from partners within the network. In total there were five winners of the META Prize, one first place and four joint second place awards; and ten products/services which were awarded the META Seal of Recognition. A series of promotional press releases were circulated to all of the winners announcing their success and to promote META-NET in doing so. These were published online and in newsletters etc. by many of the recipients. At META-FORUM 2012, in the second round of awards, there was a single META Prize winner and 11 products and services awarded the Seal of Recognition. Again, a series of promotional press releases were circulated to all of the winners announcing their success and to promote META-NET in doing so. These were published online and in newsletters etc. by many of the recipients including links and references to our work.

2.4.3.7 Task 13.7: Promotion in public administration, politics and public at large

Promotion activities aimed at administration, politics and the public at large included META-NET's presence at multiple conferences and events, such as, for example, ICT 2010 and presentations at the European Federation of National Institutes for Language's Conferences in 2010, 2012 and 2013 (EFNIL) as well as at the conference of the Network for Promotion of Linguistic Diversity (NPLD); EFNIL and NPLD represent language community bodies, language planners and government agencies responsible for every official and many more regional languages. Positive results from this work included attendance at META-FORUM and participation in panel discussions and talks by representatives from these organisations. Our META-FORUM conferences in 2010, 2011 and 2012 were well attended by representatives from administrative bodies and other stakeholders in this group. META-NET, its activities and aims were mentioned at several presentations, discussions and meetings at the EU national level in various environments (Ministries in charge of Research, Culture or Industry, research agencies, parliamentary offices, etc.) in various frameworks (preparation of national programmes, Horizon 2020, research infrastructures, consideration of cultural and regional dimensions, etc.).

The Language White Paper Series and the SRA were crucial documents in promoting our work in public administration, politics and the public at large. In addition to these documents and the associated PR campaigns much work was done on the national level by META-NET partners who organised National LT Roadshow Events and National LT Days. These events served to bolster the LT community in each country/region and also to showcase the strength and opportunities presented by our field to various national administrative representatives and speakers in attendance. This work at the national level was backed up at the European level with similar approaches at, for example, our META-FORUM events. Representatives from various European agencies have attended these events both to inform themselves and also as speakers. This work yielded several meetings with high value stakeholders. The Language White Paper series and SRA were very useful documents in executing this task. Both were published in the reporting period and allowed us to make significant progress in raising awareness and promoting our work particularly in and to the public. A large-scale orchestrated continent-wide PR campaign was rolled out in conjunction with their publication to maximise their impact. The result of this campaign was an overwhelming response from the media and public across the whole continent and around the world. Around 600 items of press coverage in print, online, radio and television resulted from this campaign. News came in from 40+ countries in 35 languages; with every European country covered as well as each language for which there is a white paper. 50+ broadcast interviews (radio/television) with META-NET representatives. Articles outside of Europe are circulating in Brazil, Costa Rica, Cuba, Australia, New Zealand, Russia, US and Canada, among others. As

media coverage grew, many bloggers and comments fora published pieces discussing our work and it went viral, especially on Twitter. This resulted in a lot of additional traffic on our website, of which ca. 65% of hits were new visits. Over 11% of all visits during this period were from Brussels (the greatest number from a single location). Over 78% of the views from Brussels were new viewers. Similar increases were seen in capital cities across Europe.

More information and further reading:

- Collaborations: <http://www.meta-net.eu/collaborations>
- META: <http://www.meta-net.eu/meta/about>
- Members of META: <http://www.meta-net.eu/participating-organisations/logos>
- META-FORUM 2010: <http://www.meta-net.eu/events/meta-forum-2010>
- META-FORUM 2011: <http://www.meta-net.eu/events/meta-forum-2011>
- META-FORUM 2011 Exhibition: <http://www.meta-net.eu/events/meta-forum-2011/exhibition>
- META-FORUM 2012: <http://www.meta-net.eu/events/meta-forum-2012>
- META-FORUM 2012 Exhibition: <http://www.meta-net.eu/events/meta-forum-2012/exhibition>
- META Prize winners: <http://www.meta-net.eu/meta-prize>
- META Seal of Recognition winners: <http://www.meta-net.eu/meta-seal>

2.4.4 WP 14: Training

Objectives: Since scientific and technological progress in machine translation and other cross- and multilingual technologies happens at high speed, much knowledge acquired during university education becomes rapidly outdated through newer developments. Moreover, even many university courses may not always be up-to-date. As language technology research and teaching is not evenly developed in all states and regions of the EU, the gaps between the major centres of progress and other important sites of science and education often widens instead of gradually being closed. Although the META-NET work will not be able to resolve all the gaps and disparities in European language technology education, it contributed to the creation of the competence base needed for the foreseen major joint action through some dedicated measures.

2.4.4.1 Task 14.1: Network Tutorials

The first tutorial was organised by CUNI in Prague, December 13-17, 2010. The course was organised jointly with the CLARA (Marie-Curie ITN) winter school. On the first of the two days, courses included dependency treebanking, with specific accent on the Prague Dependency Treebank, proposition structure as developed around the Penn Treebank. The second day started by FBK's exposition on TimeML followed by a CUNI lecture on the software platform used for handling tree structures for MT and other applications developed at CUNI. The second section of the Proposition and Discourse talks concluded the META-NET training part. In the preceding three days, participants could choose any or all of the following courses, for which they were freely admitted together with the CLARA students: University of Tübingen's course on their approach to treebanking and querying the treebanks, Univ. of Bergen course on tools for automatic analysis of corpora, another Tübingen course on corpus error detection and analysis, and CUNI's course on crowdsourcing using a game environment. The joint organisation allowed for substantial increase of the originally planned three two-hour courses to a total of nine courses with one to three two-hour blocks each, and the participation of over 50 students and researchers.

The second tutorial planned for the autumn of 2012 had to be postponed due to the lack of suitable event to attach it to; it will be organised by CUNI in August 2013 at the Depling conference organized by CUNI. Due to T4ME being over at that time, it will be supported by CUNI's own funds but it will acknowledge META-NET as well.

Moreover, CUNI lead the organisation of another event, the META-RESEARCH Workshop on Advanced Treebanking at LREC 2012 (together with colleagues of the three PSP projects, Koenraad de Smedt, Marko Tadic and Antonio Branco). This workshop gave the opportunity to present results of the META-RESEARCH part, and it was well attended also by researchers outside the META-NET community.

2.4.4.2 Task 14.2: Technology in a Nutshell

This task was concerned with providing short courses or power briefs for decision makers and other influential individuals in industry, administration and language communities. These short courses were meant to rapidly introduce and explain the fundamentals of various areas of LT emphasising the value and impact of greater use of LT. Given the time pressures on such high value individuals we anticipated that it would have been difficult to convince them to attend such courses in person. In order to overcome this hurdle, and prepare the ground for more in-depth material, we produced educational materials that could be disseminated online so that our target audience can educate themselves in their own time.

To date two full-length “nutshell” video modules have been produced covering statistical machine translation and an overview of speech technologies. These were produced as modular animated short lectures. With each nutshell having several kernels meaning each full video lecture is in turn available as shorter subsections of the whole to allow modular learning at a pace that suits the viewer. The full suite of nutshell videos will be promoted in an online awareness and education campaign through META-NET in 2013 thus increasing the impact and longevity of the work beyond the funded phase of the work.

2.4.4.3 Task 14.3: Portal to Course Materials

Training materials were collected following two different directions: materials created within META-NET activities, and materials that META-NET partners developed for their own purposes. Among the materials developed within META-NET, we made available materials developed for the first training course on Advance Treebank Annotation and on the first workshop on Machine Learning and Machine Translation, organised within the framework of WP 2 on hybrid machine translation.

In addition, partners of META-NET are active in their own research and training activities and occasionally produce training materials that can be useful for the whole community. An open call to partners was issued to obtain such materials. The materials collected in this way are the following: video lectures from the International Conference on Advance Translation Technology for Multilingual Europe, Berlin 2010; video lectures from the SMART Workshop, Grenoble 2007 that brought together researchers on Statistical Learning, Machine Translation and Textual Information Access; video lectures from the Statistical Multilingual Analysis for Retrieval and Translation (SMART) Dissemination Workshop, Barcelona 2009; video lectures from the W3C Workshop: Content on the Multilingual Web, Pisa 2011; three tutorials on Text Mining (Marko Grobelnik, Dunja Mladenic: Text Mining and Link Analysis; Marko Grobelnik, Dunja Mladenic, Blas Fortuna: Text Mining and Link Analysis for Web and Semantic Web; Marie-Francine Moens: Text Mining, Information and Fact Extraction); Introduction to Language and Speech Technology, by Jan Odijk (and for one part, Arjan van Hessen); Multiword Expressions in NLP, by Jan Odijk; Acoustic Databases, by Jan Odijk. All materials are available on the META-NET website.

More information and further reading:

- META-RESEARCH tutorial: <http://ufal.mff.cuni.cz/meta-training>
- Training materials: <http://www.meta-net.eu/meta-research/training>

2.4.5 WP 15: Enlisting Outside Competence for Community Mobilisation

Objectives: The objective of this work package was to prepare, coordinate, control and report on specialist work carried out within working groups, or by means of external consultations, or through field experts, which underpins the network’s main activities: community building and mobilisation; broadly-based dissemination and awareness efforts; exchanges and consensus building activities; launch and operation of the Council and Vision Groups; preparation and promulgation of the Strategic Research Agenda and the associated roadmaps; publications and annual conferences, etc.

The goal of the META-VISION pillar of META-NET was to put together a strategic technical alliance of stakeholders from many relevant backgrounds. This objective depended on a number of non-research activities, grouped as subcontracts in this work package. Multiple networking activities were needed to achieve a broad mobilisation of stakeholders. This included meetings at different scales (annual conferences and other European and national-level events; meetings of the Technology Council and Vision Groups, the META-NET General Assembly etc.); language white papers and other publications, both in print and digital; founding META-TRUST AISBL, a non-profit organisation as a legal entity for META-NET under Belgian law; and missions for fact-finding and soliciting support including liaisons with industry, national agencies and EU institutions.

This work package was concerned with the preparation of subcontracted actions, the selection of the appropriate contributors, and the controlling of the work by monitoring progress and assessing results. Within this WP the pillar 3 budget was managed using a budget plan and a procedure for clearance of foreseen costs through the EC.

More information and further reading:

- META-NET events, meetings and conferences: <http://www.meta-net.eu/events>
- META-TRUST AISBL: <http://www.meta-trust.eu>

2.5 Global Work Packages

In addition to the fifteen research-, infrastructure-, strategy- and content-oriented work packages the T4ME project comprised two global work packages that were concerned with its virtual information centre and website (WP 16) as well as project management (WP 17).

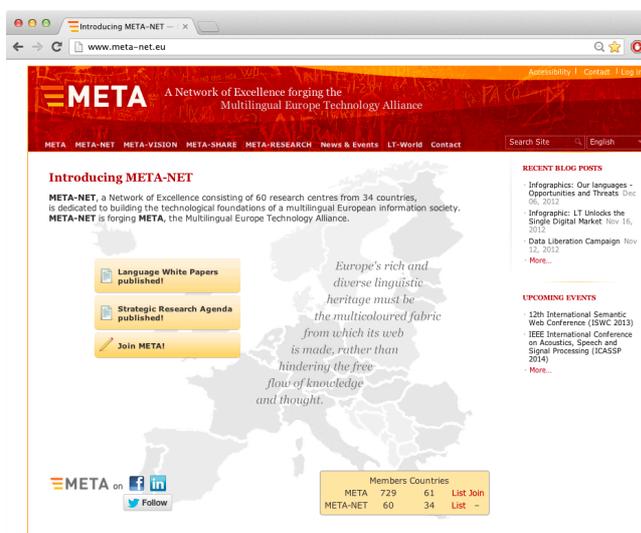
2.5.1 WP 16: Virtual Information Centre

Objectives: In order to become an acknowledged network of excellence META-NET has to accept the function as a central source of information. Internationally recognised information resources cannot be established in a few months. However, through existing information services provided by members of the network, we managed a head start in this area. The META-NET Virtual Information Centre consists of three parts:

1. META-NET website: <http://www.meta-net.eu>
2. LT World – The Knowledge Portal of META: <http://www.lt-world.org>
3. T4ME intranet: <http://t4me.dfki.de>

The T4ME intranet website was set up as a means to support the project's internal activities: management, organisation, preparation, reporting, documentation, communication, and publication of deliverables. While the T4ME intranet provided an infrastructure for the management of information within the project proper, the META-NET website is aimed at informing the public with regard to the Network of Excellence. The META-NET portal was and is supported by the knowledge portal LT World.

Due to space constraints we are unable to report in detail about the specific subtasks carried out in this work package. To sum up the activities: we set up



Co-funded by the 7th Framework Programme and the ICT Policy Support Programme of the European Commission through the contracts T4ME (grant agreement no. 244519), T4ME (grant agreement no. 271022), META-NET (grant agreement no. 270893) and META-VISION (grant agreement no. 270893). The website uses Google Analytics (more details).

the T4ME intranet and the META-NET website and also significantly expanded the already existing information portal LT World. While the T4ME intranet was a fairly straightforward and small website mainly used to store and to exchange files, the META-NET website was of much bigger scope.

The META-NET website is built upon the popular Zope/Plone framework and content management system. Its content is available in more than 20 languages. It includes information on all aspects of the initiative, i.e., META (including list of members), META-NET (members, Executive Board, mission, events, conferences, presentation videos etc.), META-VISION (especially the SRA, Language White Papers and details about the vision process, vision groups, technology council etc.), META-SHARE (for example, information about licenses and the metadata schema), META-RESEARCH, a calendar of events, blog, subsites about the four EU-funded projects etc.

2.5.2 WP 17: Coordination

Objectives: This work package ensured the overall running of the Network of Excellence and its adaptation to new challenges and situations. It was responsible for the communication with the EC Project Officer and peer reviewers, and for the interaction with external partners and contractors if such interaction was not part of another work package.

Just as with WP 16, due to space constraints we are unable to report in detail the multitude of tasks and activities carried out in WP 17. Among these tasks were the organisation of dozens of consortium and board meetings, phone and video conferences, three review meetings at the EC in Luxembourg as well as various additional meetings with representatives of the EC. We also monitored the budget and controlled progress and success with respect to the objectives of the project. General project management tasks also included enabling and supporting communication within the consortium and handling unforeseen issues.

An important aspect of this work package was concerned with collecting, reviewing and packaging the 67 deliverables foreseen in T4ME's Description of Work (V3.7). For the preparation of deliverables the Coordinator provided a document template that was available in the intranet. Many deliverables underwent a consortium-internal review before being submitted to the European Commission and to the external reviewers.

3 Socioeconomic Impact

3.1 Dissemination Strategies

The project's goals and aims are broad reaching, as such the dissemination strategy used to communicate our work needed to address several diverse stakeholder groups. To keep messages and actions consistent and more manageable, the many stakeholder and interest groups which we sought to address, inform and mobilise were grouped into broad stakeholder categories. The categories were

- *The research community:* further subdivided into core LT and neighbouring fields
- *LT Industry:* subdivided into service providers and user/application industries
- *Policy makers:* subdivided into National and European
- *User groups/associations*
- *Public at large*

With the stakeholders grouped together according to commonalities in this way we were better able to exploit common ground between stakeholders to maximise the effects of our efforts. This allowed us to tailor our messages and instruments to better suit the interests and concerns of each group and to time our actions appropriately. For example, a general somewhat sensationalist message accompanying the Language White Paper series launch gained a lot of attention with newspapers, blogs and the public at large – in the wake of this getting

our messages to politicians and policy makers at both national and European level was much more effective as these stakeholders interests (the public at large) were already mobilised.

In order for our messages to be heard, and listened to, beyond the core LT community we adopted a strategy of building and strengthening links into communities of interest beyond “traditional” LT research and LT/LSP industry, e.g., language communities, media, neighbouring fields. Throughout the project our work in this area has, to highlight a few successes, created strong links with EFNIL and NPLD who represent the official and regional language bodies (respectively) for Europe’s language communities; created a number of research data sets, shared tasks and challenges for neighbouring research fields like machine learning; and created strong links with the multilingual media community in Europe.

Once we established links with individuals or organisations in these stakeholder groups our strategy was to mobilise them to get behind our work and help spread our messages further within their own networks. In this regard we had articles published in circulars, our members were invited to give presentations, keynote addresses and take part in conferences and even take part in advisory panels on LT issues for stakeholder communities. This stakeholder mobilisation strategy allowed us to turn individuals and organisations from interested observers into active participants in our work and in some cases into valuable evangelists in a relevant stakeholder community.

Industries represent a significant proportion of our stakeholder groups. Even in a broader sense, the relevance of industry is also significant for other stakeholder groups like politicians and the public at large who want to see that R&D work being done by the academic community is both relevant and has potential for economic impact. A significant effort was made to keep industry involved in our work at various levels throughout all pillars. The vision, strategy and community building aspects of the project all featured significant input from industry visionaries and stakeholders and also the licencing and structure of META-SHARE was specifically tailored to facilitate the easy participation of industries. This all contributed to industry buy-in to our work which was evidenced by their attendance and participation in our events (META-FORUM, META-EXHIBITION and others) and having this industry buy-in and involvement contributed to the appeal of our work when presented to other stakeholders both in industry and beyond.

3.2 Dissemination Activities

3.2.1 Conferences, Presentations and Publications

Lists of conferences, publications and presentations can be seen below in the final section. Some highlights include several workshops and conferences organised by META-NET (most importantly META-FORUM 2010, 2011 and 2012), sponsoring two LREC conferences, a summer school and EACL; an invited talk at the Finnish parliament, several talks at the European Parliament’s conference on Language Technology for Multilingual Applications and keynote addresses at the EFNIL conferences. From this, and the many others in the lists below, the broad reach and impact of the project can be seen reaching all the identified stakeholder groups multiple times and through suitable channels, e.g., high level such as the European Parliament’s conference on Language Technology for Multilingual Applications for addressing high value EU stakeholders and through “traditional” media articles about the LWPs to reach the public which in turn spurred a follow up buzz in social media, blogs etc. allowing us to reach new generation media consumers, who in turn, helped spread the word further by engaging in online discussions and information sharing.

3.2.2 Community Building

META-NET’s work towards the technological foundations of a multilingual European information society around the three lines of action is largely community focused. This is evident in the wide variety of relevant stakeholder groups identified and the numbers of individuals

and organisations concerned. The importance of a strong community cannot be underestimated and as such significant work has been done to build and maintain the META-NET and META community.

The META-NET founding members consisted of 13 partners in 10 countries; the network was extended to include three PSP projects – CESAR, METANET4U and META-NORD – making a total of 44 partners in 31 countries covering all EU countries and languages. More recently the core Network of Excellence has grown further to include additional members to add to the expertise within the network, include new countries and broaden the core META-NET community. As a result the network now consists of 60 members in 34 European countries. In addition, the META Technology Council provides input into the vision process which guides much of META-NET’s work and is comprised of leading members of the LT industry community, along with researchers and some prominent members of META-NET.

In addition to the expansion of the Network of Excellence, community mobilisation efforts have led to a significant growth in the Multilingual Europe Technology Alliance. META now represents 730 organisations in 61 countries around Europe and the world all of whom support the META goals. The research community is well represented in the alliance but there is also very strong industry representation at 34% and governments and non-profit organisations make up 3% and 2% of the membership respectively. The geographical spread of the membership is primarily within Europe and the major European economies. However, META membership extends beyond just Europe, spanning the globe and reaching all continents. As well as membership of the network and META our community of involvement includes our many partner projects. Throughout the project our members have been working to increase collaboration within the field of European LT and to build bridges to neighbouring fields. This work has led to collaboration agreements between META-NET and 46 other European funded projects from such diverse fields as home automation, in-car communications, media mining, social intelligence, and medical information retrieval. The collaboration between these projects and META-NET include resource sharing through META-SHARE, the hosting of META-SHARE nodes, participation in events and sharing expertise and information relevant to META-NET activities and reports, e.g., the SRA.

Our active community spans every European country and beyond. In order to better engage with this community at a local level national contact points from the network were established in each country. Their role was to represent the project in each country and to take on the role of, for example, coordinating the Language White Paper release and approaching the relevant politicians and policy makers in each country. These national contact points were also involved in organising and running various national events around local LT communities (see below).

The project also maintains community links on various social media. These are a lightweight mechanism to reach members of our own community but also to help connect with others in related areas and communities of interest. The potential for information to “go viral” in social media means it can be very effective at spreading our messages beyond our own networks as others pick up the information and share and comment on it themselves. Currently the project has a presence on Facebook (439 followers), LinkedIn (378 members) and Twitter (196 followers). These channels have proven quite effective for spreading information and reaching people who would have been otherwise difficult to target. For further discussion see deliverable D12.4.

3.3 Planned Use of Results After T4ME/META-NET

T4ME/META-NET has never been a typical research project. Right from the outset the project had been designed to be the first funded phase of a long-term endeavour and initiative. With regard to all three lines of action (META-RESEARCH, META-SHARE, META-VISION) there are multiple results that have been used since the project ended and that will be used in the medium to long term.

META-RESEARCH – In this line of action more than 100 research publications have been produced that provide scientific progress to and for their respective sub-fields of language technology. The workshops and shared tasks initiated by META-NET will be continued in the future with the hope of further intensifying the research approaches started by the initiative. Other projects, research groups or individual researchers will continue to use and work on the software developed or extended in META-RESEARCH.

META-SHARE – In the second pillar of META-NET the open resource exchange infrastructure META-SHARE has been developed, along with a legal framework, flexible metadata schema and helpdesks to provide feedback to questions. Additionally, this part of the project helped to establish a sense of sharing in the community, i.e., that it is much more worthwhile for all parties concerned to share language resources and language technologies than to keep them under wraps. Towards the end of the project we developed a sustainability plan in order to continue regular operations and maintenance work on META-SHARE, to continue staffing the different helpdesks and also to extend the META-SHARE network by adding further nodes as well as resources and technologies. The whole work and results around META-SHARE play an important role in the planning of future activities on the level of the European Commission, especially with regard to the two programmes Connecting Europe Facility (CEF) and Horizon 2020. Currently, the European Language Resources Association (ELRA) is switching their internal systems and will use META-SHARE exclusively for distributing language resources and technologies to the community.

META-VISION – The key results and outcomes of the META-VISION line of action are the Language White Paper Series *Europe’s Languages in the Digitale Age* and the META-NET Strategic Research Agenda for Multilingual Europe 2020. We will continue to use the White Paper Series to promote our work and raise awareness for the topics of languages and language technologies and we will continue to use the SRA to promote our goals and strategic research programme. Currently, a first concrete result and success is the fact that “multilingual technologies” are among the five main building blocks of the digital component of CEF. We hope that similar results can soon be reported for the work programme of Horizon 2020. The maintenance and further extension of the initial roadmaps we prepared as companion documents for the SRA will be taken over by three ongoing support actions so that their results and discussion can also be directly used and integrated into the further strategy planning. In terms of community building, we will continue to use the established META-NET brand along with its sub-brands (META-FORUM, META-Prize, META Seal of Recognition etc.) further to enlarge the membership base of the Multilingual Technology Alliance (META) and we will also continue to host smaller and larger events, especially the META-FORUM conference series. In addition we currently plan other activities, among others, these are a META-NET book, additional (digital only) volumes of the White Paper Series (the volume on Welsh is in pre-final state), and new communication campaigns.

4 Deliverables, Publications, Presentations

4.1 Deliverables

1. Burchardt, Aljoscha, Kathrin Eichler, Georg Rehm, and Hans Uszkoreit. *D11.2 Language White Papers*. Public. T4ME Net (META-NET). DFKI, BM, CNR, ATH, CUP, CNRS, UU, AALTO, FBK, DCU, RWTH, JSI, ELDA, March 7, 2012. http://www.meta-net.eu/public_documents/t4me/META-NET-D11.2-Final.pdf.
2. Burchardt, Aljoscha, Georg Rehm, and Felix Sasaki. *D12.1 Vision Paper “European Multilingual Information Society 2020.”* Public. T4ME Net (META-NET). DFKI, CNR, ILSP, CUNI, CNRS, FBK, DCU, March 22, 2011. http://www.meta-net.eu/public_documents/t4me/META-NET-D12.1-Final.pdf
3. Choukri, Khalid, and Joseph Mariani. *D11.4.1 First Summary of Detected Gaps*. Public. T4ME Net (META-NET). ELDA, CNRS, ILSP, March 21, 2011. http://www.meta-net.eu/public_documents/t4me/META-NET-D11.4.1-Final.pdf.
4. Choukri, Khalid, Stelios Piperidis, Podromos Tsiavos, John Hendrik Weitzmann, and Tasos Patrikakos. *D6.1.2 META-SHARE: Licenses, Legal, IPR and Licensing Issues*. Public. T4ME Net (META-NET). ELDA,

- ILSP, external experts, September 8, 2011. http://www.meta-net.eu/public_documents/t4me/META-NET-D6.1.2-Final.pdf.
5. Choukri, Khalid, Stelios Piperidis, Prodromos Tsiavos, Tasos Patrikakos, Maria Gavrilidou, and John H. Weitzmann. *D6.1.3 META-SHARE: Licenses, Legal, IPR and Licensing Issues*. Public. T4ME Net (META-NET). ELDA, ATH, external experts, February 24, 2012. http://www.meta-net.eu/public_documents/t4me/META-NET-D6.1.3-Final.pdf.
 6. Desipri, Elina, Maria Gavrilidou, Penny Labropoulou, Stelios Piperidis, Francesca Frontini, Monica Monachini, Victoria Arranz, Valérie Mapelli, Gil Francopoulo, and Thierry Declerck. *D7.2.3 Documentation and User Manual of the META-SHARE Metadata Model*. Public. T4ME Net (META-NET). DFKI, CNR, ILSP, CNRS, ELDA, November 23, 2011. http://www.meta-net.eu/public_documents/t4me/META-NET-D7.2.3-Final.pdf.
 7. Desipri, Elina, Maria Gavrilidou, Penny Labropoulou, Stelios Piperidis, Francesca Frontini, Monica Monachini, Victoria Arranz, Valérie Mapelli, Gil Francopoulo, and Thierry Declerck. *D7.2.4 Documentation and User Manual of the META-SHARE Metadata Model (final)*. Public. T4ME Net (META-NET). DFKI, CNR, ATH, CNRS, ELDA, February 15, 2012. http://www.meta-net.eu/public_documents/t4me/META-NET-D7.2.4-Final.pdf.
 8. Federico, Marcello, Jan Hajič, and Jan Odijk. *D1.2 Report on Test Cases and Their Evaluation and the Resulting Publications*. Public. T4ME Net (META-NET). FBK, CUNI, UU, February 25, 2013. http://www.meta-net.eu/public_documents/t4me/meta-net-d1.2-final.pdf.
 9. Federmann, Christian, Byron Georgantopoulos, Riccardo del Gratta, Olivier Hamon, Bernardo Magnini, Dimitris Mavroeidis, Stelios Piperidis, Marc Schroeder, and Manuela Speranza. *D7.1.2 Functional and Technical Specification Including the Detailed Implementation Plan of META-SHARE*. Restricted. T4ME Net (META-NET). DFKI, FBK, ILSP, CNR, September 8, 2011.
 10. Federmann, Christian, Byron Georgantopoulos, Riccardo del Gratta, Bernardo Magnini, Dimitris Mavroeidis, Stelios Piperidis, and Manuela Speranza. *D7.1.1 META-SHARE Functional and Technical Specification*. Restricted. T4ME Net (META-NET). DFKI, FBK, ILSP, CNR, March 17, 2011.
 11. Federmann, Christian, Bernd Kiefer, Marc Schröder, Christian Spurk, Jörg Steffen, Ioanna Giannopoulou, Olivier Hamon, Christian Girardi, Salvatore Minutoli, Dimitris Mavroeidis, Xaris Papageorgiou, and Kostas Perifanos. *D8.2 META-SHARE Prototype Version 2*. Public. T4ME Net (META-NET). DFKI, CNR, ATH, FBK, ELDA WP leader, March 8, 2012. http://www.meta-net.eu/public_documents/t4me/META-NET-D8.2-Final.pdf.
 12. Federmann, Christian, Maite Melero, and Josef van Genabith. *D2.2 Optimal Choice Selection in MT*. Public. T4ME Net (META-NET). BM, DCU, DFKI, February 1, 2012. http://www.meta-net.eu/public_documents/t4me/META-NET-D2.2-Final.pdf.
 13. Federmann, Christian, and Marc Schröder. *D8.1.1 META-SHARE V1 Prototype*. Public. T4ME Net (META-NET). DFKI, CNR, ILSP, FBK, ELDA, July 28, 2011. http://www.meta-net.eu/public_documents/t4me/META-NET-D8.1.1-Final.pdf.
 14. Gavrilidou, Maria, Penny Labropoulou, Stelios Piperidis, Manuela Speranza, Monica Monachini, Victoria Arranz, and Gil Francopoulo. *D7.2.1 Specification of Metadata-Based Descriptions for Language Resources and Technologies*. Public. T4ME Net (META-NET). DFKI, CNR, ILSP, CNRS, UU, ELDA, FBK, March 17, 2011. http://www.meta-net.eu/public_documents/t4me/META-NET-D7.2.1-Final.pdf/at_download/file.
 15. Hajič, Jan, Josef van Genabith, Toni Badia, Georg Rehm, Hermann Ney, and Hans Uszkoreit. *D14.2 Report on Education and Training Activities*. Public. T4ME Net (META-NET). DCU, CUNI, DFKI, BM, RWTH, February 22, 2013. http://www.meta-net.eu/public_documents/t4me/META-NET-D14.2-Final.pdf.
 16. Hajič, Jan, Josef van Genabith, Toni Badia, Georg Rehm, and Hans Uszkoreit. *D14.1 Report on Education and Training Activities*. Public. T4ME Net (META-NET). DCU, DFKI, CUNI, BM, March 9, 2012. http://www.meta-net.eu/public_documents/t4me/META-NET-D14.1-Final.pdf.
 17. Hajič, Jan, Pierre Zweigenbaum, Timo Honkela, Dimitris Mavroeidis, François Yvon, Guillaume Wisniewski, Aurélien Max, Patrick Paroubek, Joseph Mariani, Jaakko Väyrynen, Stelios Piperidis, John Judge, Tsuyoshi Okita, and Josef van Genabith. *D5.1 Report on Challenge: Call, Award Recipients, Proposals and Results*. Public. T4ME Net (META-NET), February 28, 2013. http://www.meta-net.eu/public_documents/t4me/META-NET-D5.1-Final.pdf.
 18. Hamon, Olivier. *D9.1.1 META-SHARE Pilot Services*. Public. T4ME Net (META-NET). DFKI, ATH, CNR, ELDA, FBK, CNRS, CUNI, UU, December 27, 2011. http://www.meta-net.eu/public_documents/t4me/META-NET-D9.1.1-Final.pdf.
 19. Hamon, Olivier, and Ioanna Giannopoulou. *D9.1.2 META-SHARE Pilot Services*. Restricted. T4ME Net (META-NET). DFKI, ATH, CNR, ELDA, FBK, CNRS, CUNI, UU, February 21, 2013.

20. Jörg, Brigitte, Alistair Burt, André Herzog, Felix Sasaki, Sebastian Krause, Christian Spurk, Jörg Steffen, and Georg Rehm. *D16.3 Project Website Report*. Restricted. T4ME Net (META-NET). DFKI, March 6, 2013.
21. Jörg, Brigitte, Alistair Burt, André Herzog, Felix Sasaki, and Georg Rehm. *D16.2 Project Website Report*. Restricted. T4ME Net (META-NET). DFKI, March 7, 2012.
22. Jörg, Brigitte, Joachim Daiber, and Alistair Burt. *D16.1 Project Website Report P1*. Restricted. T4ME Net (META-NET). DFKI, March 18, 2011.
23. Judge, John, Aljoscha Burchardt, Margaretha Mazura, Georg Rehm, and Josef van Genabith. *D13.1.1 Communication Plan*. Restricted. T4ME Net (META-NET). DCU, DFKI, EMF, October 24, 2010.
24. Judge, John, Aljoscha Burchardt, Margaretha Mazura, Georg Rehm, Josef van Genabith, and Laura Grehan. *D13.1.2 Revised Communication Plan*. Restricted. T4ME Net (META-NET). DCU, DFKI, September 1, 2011.
25. Judge, John, Aljoscha Burchardt, and Georg Rehm. *D13.7.1 Yearly Mobilisation Report*. Restricted. T4ME Net (META-NET). DFKI, DCU, February 28, 2011.
26. Judge, John, Aljoscha Burchardt, and Georg Rehm. *D13.7.2 Yearly Mobilisation Report*. Restricted. T4ME Net (META-NET). DFKI, DCU, February 16, 2012.
27. Judge, John, Georg Rehm, and Aljoscha Burchardt. *D12.3 Community Involvement*. Public. T4ME Net (META-NET). DCU, DFKI, March 6, 2013. http://www.meta-net.eu/public_documents/t4me/META-NET-D12.3-Final.pdf.
28. Judge, John, Georg Rehm, and Aljoscha Burchardt. *D12.4 External Communication and Community Involvement*. Public. T4ME Net (META-NET). DCU, DFKI, March 6, 2013. http://www.meta-net.eu/public_documents/t4me/META-NET-D12.4-Final.pdf.
29. Judge, John, Georg Rehm, and Aljoscha Burchardt. *D13.4 Industrial Liaison*. Public. T4ME Net (META-NET). DCU, DFKI, March 6, 2013. http://www.meta-net.eu/public_documents/t4me/META-NET-D13.4-Final.pdf.
30. Judge, John, Georg Rehm, and Aljoscha Burchardt. *D13.6 Prize Awards*. Public. T4ME Net (META-NET). DCU, DFKI, March 6, 2013. http://www.meta-net.eu/public_documents/t4me/META-NET-D13.6-Final.pdf.
31. Judge, John, Georg Rehm, and Aljoscha Burchardt. *D13.7.3 Yearly Mobilisation Report*. Public. T4ME Net (META-NET). DCU, DFKI, March 6, 2013. http://www.meta-net.eu/public_documents/t4me/META-NET-D13.7.3-Final.pdf.
32. Koutsombogera, Maria, and Stelios Piperidis. *D10.1 META-SHARE Support Actions: Report on Legal Issues*. Restricted. T4ME Net (META-NET). ATH, February 21, 2013.
33. Koutsombogera, Maria, and Stelios Piperidis. *D10.2 META-SHARE Support Actions: Report on Technical Issues*. Restricted. T4ME Net (META-NET). ATH, February 21, 2013.
34. Koutsombogera, Maria, and Stelios Piperidis. *D13.3 Cooperation with Research Centres*. Restricted. T4ME Net (META-NET). ATH, DCU, DFKI, February 10, 2012.
35. Labropoulou, Penny, Maria Gavrilidou, Elina Desipri, Stelios Piperidis, Francesca Frontini, Monica Monachini, Victoria Arranz, and Gil Francopoulo. *D7.2.2 Specification of Metadata-Based Descriptions for Language Resources and Technologies*. Public. T4ME Net (META-NET). DFKI, CNR, ATH, CNRS, UU, ELDA, FBK, February 15, 2012. http://www.meta-net.eu/public_documents/t4me/META-NET-D7.2.2-Final.pdf.
36. Magnini, Bernardo, Manuela Speranza, Olivier Hamon, Victoria Arranz, Christian Federmann, and Marc Schröder. *D8.1.2 META-SHARE Release Version 1: Installation Instructions*. Public. T4ME Net (META-NET). DFKI, ATH, CNR, ELDA, FBK, September 26, 2011. http://www.meta-net.eu/public_documents/t4me/META-NET-D8.1.2-Final.pdf.
37. Magnini, Bernardo, Manuela Speranza, Olivier Hamon, Penny Labropoulou, and Stelios Piperidis. *D9.2.2 Assessment of META-SHARE Pilot Services*. Public. T4ME Net (META-NET). FBK, ELDA, ILSP, March 16, 2013. http://www.meta-net.eu/public_documents/t4me/META-NET-D9.2.2-Final.pdf.
38. Magnini, Bernardo, Manuela Speranza, Olivier Hamon, Penny Labropoulou, and Stelios Piperidis. *D9.2.1 Assessment of META-SHARE Pilot Services*. Public. T4ME Net (META-NET). FBK, March 8, 2012. http://www.meta-net.eu/public_documents/t4me/META-NET-D9.2.1-Final.pdf.
39. Mariani, Joseph, and John Judge. *D13.5 Utilising Existing LT Fora for Mobilization*. Restricted. T4ME Net (META-NET). CNRS, CUP, DCU, ATH, January 30, 2013.
40. Mariani, Joseph, Khalid Choukri, and Stelios Piperidis. *D6.3.1 META-SHARE: Constitution, Business Model, Business Plan*. Public. T4ME Net (META-NET). CNRS, ELDA, ILSP, February 9, 2011. http://www.meta-net.eu/public_documents/t4me/META-NET-D6.3.1-Final.pdf.

41. Mariani, Joseph, and Gil Francopoulo. *D11.1.1 First Public Version of the META-Matrix*. Public. T4ME Net (META-NET). CNRS, CNR, ELDA, ILSP, March 21, 2011. http://www.meta-net.eu/public_documents/t4me/META-NET-D11.1.1-Final.pdf.
42. Mariani, Joseph, and Gil Francopoulo. *D11.1.2 Second Public Version of the META-Matrix*. Restricted. T4ME Net (META-NET). CNRS, CNR, ELDA, ILSP, September 3, 2012.
43. Melero, Maite, Christian Federmann, and Josef van Genabith. *D2.4.2 ML4HMT 2011 Workshop Report*. Public. T4ME Net (META-NET). BM, DFKI, DCU, December 23, 2011. http://www.meta-net.eu/public_documents/t4me/META-NET-D2.4.2-Final.pdf.
44. Melero, Maite, Christian Federmann, and Josef van Genabith. *D2.4.3 ML4HMT 2012 Workshop Report*. Public. T4ME Net (META-NET). BM, DFKI, DCU, January 21, 2013. http://www.meta-net.eu/public_documents/t4me/META-NET-D2.4.3-Final.pdf.
45. Melero, Maite, and Pavel Pecina. *D2.4.1 Machine Learning Tutorial*. Public. T4ME Net (META-NET). DFKI, DCU, BM, November 3, 2010. http://www.meta-net.eu/public_documents/t4me/META-NET-D2.4.1-Final.pdf.
46. Musillo, Gabriele, and Marcello Federico. *D1.1 Reviewing Recent Efforts on Integrating Semantics into Machine Translation*. Public. T4ME Net (META-NET). FBK, December 8, 2010. http://www.meta-net.eu/public_documents/t4me/T4ME-D1-1.pdf.
47. Okita, Tsuyoshi, and Josef van Genabith. *D2.3 Evaluation Report*. Public. T4ME Net (META-NET). DFKI, DCU, February 23, 2013. http://www.meta-net.eu/public_documents/t4me/META-NET-D2.3-Final.pdf.
48. Pajntar, Boštjan. *D4.1 Software Prototypes for the Collection of Parallel and Comparable Corpora from Hidden-Web, Web2.0 and Multilingual News Sources*. Restricted to other programme participants. T4ME Net (META-NET). JSI, March 8, 2011.
49. Pecina, Pavel, Eleftherios Avramidis, Christian Federmann, Josef van Genabith, Maite Melero, and Marta Ruiz. *D2.1 Annotated Hybrid Sample MT Corpus*. Public. T4ME Net (META-NET). DCU, DFKI, BM, January 21, 2011. http://www.meta-net.eu/public_documents/t4me/meta-net-d2.1-final.pdf.
50. Peitz, Stephan, and Hermann Ney. *D4.2 Improved Statistical Models*. Restricted to other programme participants. T4ME Net (META-NET). RWTH, February 1, 2012.
51. Peitz, Stephan, and Hermann Ney. *D4.4 MT Techniques*. Public. T4ME Net (META-NET). RWTH, February 20, 2013. http://www.meta-net.eu/public_documents/t4me/META-NET-D4.4-Final.pdf.
52. Piperidis, Stelios, Nicoletta Calzolari, and Maria Koutsombogera. *D6.2.1 META-SHARE: Design and Governance*. Public. T4ME Net (META-NET). CNR, ILSP, September 8, 2011. http://www.meta-net.eu/public_documents/t4me/META-NET-D6.2.1-Final.pdf.
53. Piperidis, Stelios, Nicoletta Calzolari, and Maria Koutsombogera. *D6.2.2 META-SHARE: Design, Governance and Business Model*. Public. T4ME Net (META-NET). CNR, ILSP, March 5, 2012. http://www.meta-net.eu/public_documents/t4me/META-NET-D6.2.2-Final.pdf.
54. Piperidis, Stelios, and Maria Koutsombogera. *D6.2.3 META-SHARE: Design, Governance and Sustainability*. Public. T4ME Net (META-NET). ATH, CNR, CNRS, ELDA, March 6, 2013. http://www.meta-net.eu/public_documents/t4me/META-NET-D6.2.3-Final.pdf.
55. Rehm, Georg. *D15.1 Report on Using Outside Competence for Community Mobilisation*. Restricted. T4ME Net (META-NET). DFKI, BM, CNR, ILSP, CUP, CNRS, UU, AALTO, FBK, DCU, RWTH, JSI, ELDA, September 12, 2011.
56. Rehm, Georg, Aljoscha Burchardt, Marcello Federico, Christian Federmann, Josef van Genabith, Byron Georgantopoulos, Marko Grobelnik, Jan Hajic, Brigitte Jörg, John Judge, Maria Koutsombogera, Joseph Mariani, Dimitris Mavroeidis, Bostjan Pajntar, Pavel Pecina, Stelios Piperidis, Hans Uszkoreit. *D17.1 Yearly Report No. 1*. Restricted. T4ME Net (META-NET). DFKI, BM, CNR, ILSP, CUP, CNRS, UU, AALTO, FBK, DCU, RWTH, JSI, ELDA, September 12, 2011.
57. Rehm, Georg, and Hans Uszkoreit. *D12.2.2 Strategic Research Agenda*. Public. T4ME Net (META-NET). DFKI, February 21, 2013. http://www.meta-net.eu/public_documents/t4me/META-NET-D12.2.2-Final.pdf.
58. Rehm, Georg, Aljoscha Burchardt, Marcello Federico, Josef van Genabith, Marko Grobelnik, Jan Hajič, Brigitte Jörg, John Judge, Maria Koutsombogera, Joseph Mariani, Stelios Piperidis, Hans Uszkoreit et al. *D17.2 Yearly Report No. 2*. Restricted. T4ME Net (META-NET). DFKI, BM, CNR, ILSP, CUP, CNRS, UU, AALTO, FBK, DCU, RWTH, JSI, ELDA, June 8, 2012.
59. Rehm, Georg, Marcello Federico, Josef van Genabith, Marko Grobelnik, Jan Hajič, Brigitte Jörg, John Judge, Maria Koutsombogera, Joseph Mariani, Stelios Piperidis, Hans Uszkoreit et al. *D17.3 Yearly Report No. 3*. Restricted. T4ME Net (META-NET). DFKI, BM, CNR, ILSP, CUP, CNRS, UU, AALTO, FBK, DCU, RWTH, JSI, ELDA, June 3, 2013.

Final Report

60. Russo, Irene, Jan Rupnik, and Andrej Muhič. D4.3 New Methods for Multilingual Information Retrieval and Sentiment Detection Based on Language Neutral Document Representation. Public. T4ME Net (META-NET). CNR, JSI, RWTH, February 20, 2013. http://www.meta-net.eu/public_documents/t4me/META-NET-D4.3-Final.pdf.
61. Soria, Claudia, and Maria Koutsombogera. D6.3 META-SHARE Communication and Community Mobilisation Plan. Restricted. T4ME Net (META-NET). CNR, ATH, ELDA, DFKI, February 10, 2012.
62. Soria, Claudia, and Joseph Mariani. D11.3 Report on Existing Projects and Initiatives. Public. T4ME Net (META-NET). CNR, CNRS, March 18, 2011. http://www.meta-net.eu/public_documents/t4me/META-NET-D11.3-Final.pdf.
63. Spurk, Christian, Christian Federmann, Ioanna Giannopoulou, Christian Girardi, Olivier Hamon, Dimitris Mavroeidis, Salvatore Minutoli, Marc Schröder, Jörg Steffen, and Xaris Papageorgiou. D8.3 META-SHARE Version 3. Public. T4ME Net (META-NET). DFKI, CNR, ATH, FBK, ELDA, October 10, 2012. http://www.meta-net.eu/public_documents/t4me/META-NET-D8.3-Final.pdf.
64. Zweigenbaum, Pierre, Timo Honkela, Dimitris Mavroeidis, François Yvon, Guillaume Wisniewski, Aurélien Max, Patrick Paroubek, Joseph Mariani, Jaakko Väyrynen, and Stelios Piperidis. D3.2 Report on Advances in the Use of Context in Machine Translation. Public. T4ME Net (META-NET). CNRS, AALTO, ILSP, February 23, 2013. http://www.meta-net.eu/public_documents/t4me/META-NET-D3.2-Final.pdf.

4.2 Publications

1. Adda, Gilles, Benoît Sagot, Karén Fort, and Joseph Mariani. “Crowdsourcing for Language Resource Development: Critical Analysis of Amazon Mechanical Turk Overpowering Use.” In *Proceedings of the 5th Language & Technology Conference (LTC 2011)*. Poznan, Poland, 2011. http://hal.archives-ouvertes.fr/docs/00/64/81/87/PDF/ltc-56-adda_final.pdf.
2. Allauzen, Alexandre, and François Yvon. “Méthodes statistiques pour la traduction automatique.” In *Modèles statistiques pour l'accès à l'information textuelle*, edited by Éric Gaussier and François Yvon, 271–356. Paris: Hermès, 2011. <http://www.lavoisier.fr/livre/notice.asp?ouvrage=2139540>.
3. Allauzen, Alexandre, and François Yvon. “Textual Information Access.” In *Statistical Methods for Machine Translation*, edited by Éric Gaussier and François Yvon, 223–304. Hoboken: Wiley, 2012. <http://onlinelibrary.wiley.com/doi/10.1002/9781118562796.ch7/summary>.
4. Arcan, Mihael, Paul Buitelaar, and Christian Federmann. “Using Domain-specific and Collaborative Resources for Term Translation.” In *Proceedings of the 6th Workshop on Syntax, Semantics and Structure in Statistical Translation (SSST-6)*, 86–94. Jeju, Republic of Korea: Association for Computational Linguistics (ACL), 2012. <http://www.aclweb.org/anthology-new/W/W12/W12-4210.pdf>.
5. Arcan, Mihael, Christian Federmann, and Paul Buitelaar. “Experiments with Term Translation.” In *Proceedings of the 24th International Conference on Computational Linguistics (COLING 2012)*, 67–82. Mumbai, India, 2012. <http://aclweb.org/anthology-new/C/C12/C12-1005.pdf>.
6. Asooja, Kartik, Jorge Gracia, Nitish Aggarwal, and Asunción Gómez Pérez. “Using Cross-Lingual Explicit Semantic Analysis for Improving Ontology Translation.” In *Proceedings of the Second Shared Task on Applying Machine Learning Techniques to Optimise the Division of Labour in Hybrid Machine Translation (ML4HMT-12)*, 25–36. Mumbai, India, 2012. <http://www.aclweb.org/anthology-new/W/W12/W12-5703.pdf>.
7. Avramidis, Eleftherios. “DFKI System Combination with Sentence Ranking at ML4HMT-2011.” In *Proceedings of the International Workshop on Using Linguistic Information for Hybrid Machine Translation (LIHMT 2011) and of the Shared Task on Applying Machine Learning Techniques to Optimising the Division of Labour in Hybrid Machine Translation (M. Shared Task on Applying Machine Learning Techniques to Optimising the Division of Labour in Hybrid Machine Translation (ML4HMT-11))*, 99–103. Barcelona, Spain: Universitat Politècnica de Catalunya, 2011. <http://ixa2.si.ehu.es/lihmt2011/proceedings.pdf>.
8. Avramidis, Eleftherios, Aljoscha Burchardt, Christian Federmann, Maja Popović, Cindy Tscherwinka, and David Vilar. “Involving Language Professionals in the Evaluation of Machine Translation.” In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012)*, 1127–1130. Istanbul, Turkey: European Language Resources Association (ELRA), 2012. http://www.lrec-conf.org/proceedings/lrec2012/pdf/294_Paper.pdf.
9. Avramidis, Eleftherios, Marta R. Costa-Jussà, Christian Federmann, Maite Melero, Pavel Pecina, and Josef van Genabith. “A Richly Annotated, Multilingual Parallel Corpus for Hybrid Machine Translation.” In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012)*, 2189–2193. Istanbul, Turkey: European Language Resources Association (ELRA), 2012. http://www.lrec-conf.org/proceedings/lrec2012/pdf/444_Paper.pdf.

10. Barrault, Loïc, and Patrik Lambert. "Machine Translation System Combination with MANY for ML4HMT." In *Proceedings of the International Workshop on Using Linguistic Information for Hybrid Machine Translation (LIHMT 2011) and of the Shared Task on Applying Machine Learning Techniques to Optimising the Division of Labour in Hybrid Machine Translation (M. Shared Task on Applying Machine Learning Techniques to Optimising the Division of Labour in Hybrid Machine Translation (ML4HMT-11))*, 89–93. Barcelona, Spain: Universitat Politècnica de Catalunya, 2011. <http://ixa2.si.ehu.es/lihmt2011/proceedings.pdf>.
11. Bisazza, Arianna, and Marcello Federico. "Cutting the Long Tail: Hybrid Language Models for Translation Style Adaptation." In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2012)*, 439–448. Avignon, France: Association for Computational Linguistics (ACL), 2012. <http://aclweb.org/anthology-new/E/E12/E12-1045.pdf>.
12. Bisazza, Arianna, and Marcello Federico. "Modified Distortion Matrices for Phrase-Based Statistical Machine Translation." In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (ACL 2012)*, 478–487. Jeju, Republic of Korea: Association for Computational Linguistics (ACL), 2012. <http://www.aclweb.org/anthology-new/P/P12/P12-1050.pdf>.
13. Bojar, Ondřej, Silvie Cinková, Jan Hajič, Barbora Hladká, Vladislav Kuboň, Jiří Mírovský, Jarmila Panevová, Nino Peterek, Johanka Spoustová, and Zdeněk Žabokrtský. *The Czech Language in the Digital Age / Čeština v digitálním věku*. Edited by Georg Rehm and Hans Uszkoreit. META-NET White Paper Series "Europe's Languages in the Digital Age". Heidelberg et al.: Springer, 2012. <http://www.meta-net.eu/whitepapers/e-book/czech.pdf>.
14. Bojar, Ondřej, Miloš Ercegovčević, Martin Popel, and Omar Zaidan. "A Grain of Salt for the WMT Manual Evaluation." In *Proceedings of the 6th Workshop on Statistical Machine Translation (WMT 2011)*, 1–11. Edinburgh, Scotland: Association for Computational Linguistics (ACL), 2011. <http://statmt.org/wmt11/pdf/WMT01.pdf>.
15. Burchardt, Aljoscha, Markus Egg, Katrin Eichler, Brigitte Krenn, Jörn Kreutel, Annette Leßmöllmann, Georg Rehm, Manfred Stede, Hans Uszkoreit, and Martin Volk. *The German Language in the Digital Age / Die Deutsche Sprache im Digitalen Zeitalter*. Edited by Georg Rehm and Hans Uszkoreit. META-NET White Paper Series "Europe's Languages in the Digital Age". Heidelberg et al.: Springer, 2012. <http://www.meta-net.eu/whitepapers/e-book/german.pdf>.
16. Burchardt, Aljoscha, and Georg Rehm. *Vision Document. Vision Group Translation and Localisation: Results of First Two Meetings*. Berlin, Germany, December 3, 2010. <http://www.meta-net.eu/vision/reports/VisionGroup-TranslationLocalisation-draft.pdf>.
17. Calzolari, Nicoletta, Riccardo Del Gratta, Gil Francopoulo, Joseph Mariani, Francesco Rubino, Irene Russo, and Claudia Soria. "The LRE Map. Harmonising Community Descriptions of Resources." In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012)*, 1084–1089. Istanbul, Turkey, 2012. http://www.lrec-conf.org/proceedings/lrec2012/pdf/769_Paper.pdf.
18. Calzolari, Nicoletta, Riccardo Del Gratta, Francesca Frontini, and Irene Russo. "The Language Library: Many Layers, More Knowledge." In *Proceedings of the Workshop on Language Resources, Technology and Services in the Sharing Paradigm (LRTS)*, 93–97. Chiang Mai, Thailand, 2011. <https://www.aclweb.org/anthology/W/W11/W11-3312.pdf>.
19. Calzolari, Nicoletta, Bernardo Magnini, Claudia Soria, and Manuela Speranza. *The Italian Language in the Digital Age / La Lingua Italiana nell'Era Digitale*. Edited by Georg Rehm and Hans Uszkoreit. META-NET White Paper Series "Europe's Languages in the Digital Age". Heidelberg et al.: Springer, 2012. <http://www.meta-net.eu/whitepapers/e-book/italian.pdf>.
20. Calzolari, Nicoletta, Monica Monachini, and Valeria Quochi. "Interoperability Framework: The FLReNet Action Plan Proposal." In *Proceedings of the Workshop on Language Resources, Technology and Services in the Sharing Paradigm (LRTS)*, 41–49. Chiang Mai, Thailand, 2011. <http://www.aclweb.org/anthology-new/W/W11/W11-3306.pdf>.
21. Caselli, Tommaso, Irene Russo, and Francesco Rubino. "Assigning Connotation Values to Events." In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012)*, 3082–3089. Istanbul, Turkey: European Language Resources Association (ELRA), 2012. http://www.lrec-conf.org/proceedings/lrec2012/pdf/1111_Paper.pdf.
22. Cettolo, Mauro, Nicola Bertoldi, and Marcello Federico. "Methods for Smoothing the Optimizer Instability in SMT." In *Proceedings of the 13th Machine Translation Summit*, 32–39. Xiamen, China, 2011. <http://mt-archive.info/MTS-2011-Cettolo.pdf>.
23. Choukri, Khalid, Jungyeul Park, Olivier Hamon, and Victoria Arranz. "Proposal for the International Standard Language Resource Number (ISLRN)." In *Proceedings of the Workshop on Language Resources, Technology and Services in the Sharing Paradigm (LRTS)*, 75–83. Chiang Mai, Thailand, 2011. <http://www.aclweb.org/anthology-new/W/W11/W11-3310.pdf>.

Final Report

24. Costa-Jussà, Marta R., and Rafael E. Banchs. "Automatic Normalization of Short Texts by Combining Statistical and Rule-based Techniques." *Language Resources and Evaluation. Special Issue Analysis of on Short Texts on the Web* 47, no. 1 (March 2013): 179–193.
25. Costa-Jussà, Marta R., Mireia Farrús, José B. Mariño, and Jose A. R. Fonollosa. "Study and Comparison of Rule-Based and Statistical Catalan-Spanish Machine Translation Systems." *Computing and Informatics* 31, no. 2 (2012): 245–270.
26. Crego, Josep Maria, François Yvon, and José B. Mariño. "N-code: An Open-source Bilingual N-gram SMT Toolkit." *The Prague Bulletin of Mathematical Linguistics (PBML)* 96, no. 1 (October 2011): 49–58.
27. Dušek, Ondrej, Zdeněk Žabokrtský, Martin Popel, Martin Majliš, Michal Novák, and David Mareček. "Formemes in English-Czech Deep Syntactic MT." In *Proceedings of the 7th Workshop on Statistical Machine Translation (WMT 2012)*, 267–274. Montréal, Canada: Association for Computational Linguistics (ACL), 2012. <http://www.aclweb.org/anthology-new/W/W12/W12-3132.pdf>.
28. Farrús, Mireia, Marta R. Costa-Jussà, and Maja Popović. "Study and Correlation Analysis of Linguistic, Perceptual and Automatic Machine Translation Evaluations." *Journal of the American Society for Information Science and Technology (JASIST)* 63, no. 1 (2012): 174–184.
29. Federico, Marcello, Luisa Bentivogli, Michael Paul, and Sebastian Stüker. "Overview of the IWSLT 2011 Evaluation Campaign." In *Proceedings of the 8th International Workshop on Spoken Language Translation (IWSLT 2011)*, 11–27. San Francisco, USA, 2011. <http://mt-archive.info/IWSLT-2011-Federico.pdf>.
30. Federico, Marcello, Sebastian Stüker, Luisa Bentivogli, Michael Paul, Mauro Cettolo, Teresa Herrmann, Jan Niehues, and Giovanni Moretti. "The IWSLT 2011 Evaluation Campaign on Automatic Talk Translation." In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012)*, 3543–3550. Istanbul, Turkey: European Language Resources Association (ELRA), 2012. https://wit3.fbk.eu/archive/2011-01/subeval_files/IWSLT2011-LREC2012.pdf.
31. Federmann, Christian. "Multi-Engine Machine Translation as a Lifelong Machine Learning Problem." In *Lifelong Machine Learning: Papers from the 2013 AAAI Spring Symposium*, 4–8. Stanford, USA: Association for the Advancement of Artificial Intelligence, 2013. <http://www.aaai.org/ocs/index.php/SSS/SSS13/paper/view/5730/5970>.
32. Federmann, Christian. "A Machine-Learning Framework for Hybrid Machine Translation." In *Proceedings of the 35th Annual German Conference on Artificial Intelligence*, edited by Birte Glimm and Antonio Krüger, 37–48. Heidelberg et al.: Springer, 2012. http://link.springer.com/chapter/10.1007/978-3-642-33347-7_4.
33. Federmann, Christian. "Appraise: An Open-Source Toolkit for Manual Evaluation of MT Output." *The Prague Bulletin of Mathematical Linguistics (PBML)* 98 (2012): 25–35.
34. Federmann, Christian. "Can Machine Learning Algorithms Improve Phrase Selection in Hybrid Machine Translation?" In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2012)*, 113–118. Avignon, France: Association for Computational Linguistics (ACL), 2012. <http://www.aclweb.org/anthology-new/W/W12/W12-0115.pdf>.
35. Federmann, Christian. "Hybrid Machine Translation Using Joint, Binarised Feature Vectors." In *Proceedings of the 10th Biennial Conference of the Association for Machine Translation in the Americas (AMTA 2012)*. San Diego, USA, 2012. <http://amta2012.amtaweb.org/AMTA2012Files/papers/126.pdf>.
36. Federmann, Christian. "System Combination Using Joint, Binarised Feature Vectors." In *Proceedings of the Second Shared Task on Applying Machine Learning Techniques to Optimise the Division of Labour in Hybrid Machine Translation (ML4HMT-12)*, 77–84. Mumbai, India, 2012. <http://www.aclweb.org/anthology-new/W/W12/W12-5708.pdf>.
37. Federmann, Christian. "Results from the ML4HMT Shared Task on Applying Machine Learning Techniques to Optimise the Division of Labour in Hybrid MT." In *Proceedings of the International Workshop on Using Linguistic Information for Hybrid Machine Translation (LIHMT 2011) and of the Shared Task on Applying Machine Learning Techniques to Optimise the Division of Labour in Hybrid Machine Translation (ML4HMT-11)*, 110–117. Barcelona, Spain: Universitat Politècnica de Catalunya, 2011. <http://ixa2.si.ehu.es/lihmt2011/proceedings.pdf>.
38. Federmann, Christian, Eleftherios Avramidis, Marta R. Costa-Jussà, Maite Melero, Pavel Pecina, and Josef van Genabith. "The ML4HMT Workshop on Optimising the Division of Labour in Hybrid Machine Translation." In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012)*, 3430–3435. Istanbul, Turkey: European Language Resources Association (ELRA), 2012. http://www.lrec-conf.org/proceedings/lrec2012/pdf/996_Paper.pdf.
39. Federmann, Christian, Yu Chen, and Ri Wang. "DFKI System Combination Using Syntactic Information at ML4HMT-2011." In *Proceedings of the International Workshop on Using Linguistic Information for Hybrid Machine Translation (LIHMT 2011) and of the Shared Task on Applying Machine Learning Techniques*

- to *Optimise the Division of Labour in Hybrid Machine Translation (ML4HMT-11)*, 104–109. Barcelona, Spain: Universitat Politècnica de Catalunya, 2011. <http://ixa2.si.ehu.es/lihmt2011/proceedings.pdf>.
40. Federmann, Christian, Ioanna Giannopoulou, Christian Girardi, Olivier Hamon, Dimitris Mavroeidis, Salvatore Minutoli, and Marc Schröder. “META-SHARE V2: An Open Network of Repositories for Language Resources Including Data and Tools.” In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012)*, 3300–3303. Istanbul, Turkey: European Language Resources Association (ELRA), 2012. http://www.lrec-conf.org/proceedings/lrec2012/pdf/816_Paper.pdf.
 41. Federmann, Christian, Maite Melero, Pavel Pecina, and Josef van Genabith. “Towards Optimal Choice Selection for Improved Hybrid Machine Translation.” *The Prague Bulletin of Mathematical Linguistics (PBML)* 97 (April 2012): 5–22.
 42. Federmann, Christian, Tsuyoshi Okita, Maite Melero, Marta R. Costa-Jussà, Toni Badia, and Josef van Genabith. “Results from the ML4HMT-12 Shared Task on Applying Machine Learning Techniques to Optimise the Division of Labour in Hybrid Machine Translation.” In *Proceedings of the Second Shared Task on Applying Machine Learning Techniques to Optimise the Division of Labour in Hybrid Machine Translation (ML4HMT-12)*, 85–90. Mumbai, India, 2012. <http://www.aclweb.org/anthology-new/W/W12/W12-5709.pdf>.
 43. Freitag, Markus, Gregor Leusch, Joern Wuebker, Stephan Peitz, Hermann Ney, Teresa Herrmann, Jan Niehues, et al. “Joint WMT Submission of the QUAERO Project.” In *Proceedings of the 6th Workshop on Statistical Machine Translation (WMT 2011)*, 358–364. Edinburgh, Scotland: Association for Computational Linguistics (ACL), 2011. <http://www.statmt.org/wmt11/WMT-2011.pdf>.
 44. Gahbiche-Braham, Souhir, Hélène Bonneau-Maynard, Thomas Lavergne, and François Yvon. “Joint Segmentation and POS Tagging for Arabic Using a CRF-based Classifier.” In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012), SemRel 2012 Workshop*. Istanbul, Turkey: European Language Resources Association (ELRA), 2012. <http://oniros.org/papers/Gahbiche2012lrec.pdf>.
 45. Gahbiche-Braham, Souhir, Hélène Bonneau-Maynard, and François Yvon. “Two Ways to Use a Noisy Parallel News Corpus for Improving Statistical Machine Translation.” In *Proceedings of the 4th Workshop on Building and Using Comparable Corpora: Comparable Corpora and the Web*, 44–51. Portland, USA: Association for Computational Linguistics (ACL), 2011. <http://aclweb.org/anthology-new/W/W11/W11-1207.pdf>.
 46. Gavrilidou, Maria, Maria Koutsombogera, Anastasios Patrikakos, and Stelios Piperidis. *The Greek Language in the Digital Age / □ □□□□□□□ □□□□□□ □□□□ □□□□□□ □□□□□*. Edited by Georg Rehm and Hans Uszkoreit. META-NET White Paper Series “Europe’s Languages in the Digital Age”. Heidelberg et al.: Springer, 2012. <http://www.meta-net.eu/whitepapers/e-book/greek.pdf>.
 47. Gavrilidou, Maria, Penny Labropoulou, Elina Desipri, Stelios Piperidis, Haris Papegeorgiou, Monica Monachini, Francesca Frontini, et al. “The META-SHARE Metadata Schema for the Description of Language Resources.” In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012)*, 1090–1097. Istanbul, Turkey: European Language Resources Association (ELRA), 2012. http://www.lrec-conf.org/proceedings/lrec2012/pdf/998_Paper.pdf.
 48. Gavrilidou, Maria, Penny Labropoulou, and Stelios Piperidis. “A Proposal for a Metadata Model for Language Resources.” In *Selected Papers of the 10th ICGL*, 284–293. Komotini, Greece: Democritus University of Thrace, 2011. http://www.icgl.gr/files/English/19.Gavrilidou_et_al_10ICGL_pp.284-293.pdf.
 49. Gavrilidou, Maria, Penny Labropoulou, Stelios Piperidis, Monica Monachini, Gil Francopoulo, Victoria Aranz, and Valérie Mapelli. “A Metadata Schema for the Description of Language Resources.” In *Proceedings of the Workshop on Language Resources, Technology and Services in the Sharing Paradigm (LRTS)*, 84–92. Chiang Mai, Thailand, 2011. <http://www.aclweb.org/anthology-new/W/W11/W11-3311.pdf>.
 50. Genabith, Josef van, Toni Badia, Christian Federmann, Maite Melero, Marta R. Costa-Jussà, and Tsuyoshi Okita, eds. *Proceedings of the Second Workshop on Applying Machine Learning Techniques to Optimise the Division of Labour in Hybrid MT (ML4HMT-12)*. Mumbai, 2012. <http://www.dfki.de/ml4hmt/>.
 51. Gratta, Riccardo Del, Francesca Frontini, Francesco Rubino, Irene Russo, and Nicoletta Calzolari. “The Language Library: Supporting Community Effort for Collective Resource Production.” In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012)*, 43–49. Istanbul, Turkey: European Language Resources Association (ELRA), 2012. http://www.lrec-conf.org/proceedings/lrec2012/pdf/396_Paper.pdf.
 52. Hardmeier, Christian, and Marcello Federico. “Modelling Pronominal Anaphora in Statistical Machine Translation.” In *Proceedings of the 7th International Workshop on Spoken Language Translation (IWSLT 2010)*. Paris, France, 2010. <http://www.diva-portal.org/smash/get/diva2:420761/FULLTEXT01>
 53. Hardmeier, Christian, Jörg Tiedemann, Markus Saers, Marcello Federico, and Mathur Prashant. “The Up-sala-FBK Systems at WMT 2011.” In *Proceedings of the 6th Workshop on Statistical Machine Translation*

- (WMT 2011), 372–378. Edinburgh, Scotland: Association for Computational Linguistics (ACL), 2011. <http://www.aclweb.org/anthology-new/W/W11/W11-2144.pdf>.
54. Henderson, James, and Filip Jurčiček. “Data-Driven Methods for Spoken Language Understanding.” In *Data-driven Methods for Adaptive Spoken Dialogue Systems*, edited by Oliver Lemon and Olivier Pietquin, 177:19–38. Computational Learning for Conversational Interfaces IX. New York et al.: Springer, 2012. http://link.springer.com/chapter/10.1007%2F978-1-4614-4803-7_3.
 55. Holub, Martin, Vincent Kríž, Silvie Cinková, and Eckhard Bick. “Tailored Feature Extraction for Lexical Disambiguation of English Verbs Based on Corpus Pattern Analysis.” In *Proceedings of the 24th International Conference on Computational Linguistics (COLING2012), Technical Papers*, 1195–1210. Mumbai, India: Association for Computational Linguistics (ACL), 2012. <http://aclweb.org/anthology-new/C/C12/C12-1073.pdf>.
 56. Honkela, Timo, Włodzisław Duch, Mark A. Girolami, and Samuel Kaski, eds. *Artificial Neural Networks and Machine Learning - ICANN 2011 - 21st International Conference on Artificial Neural Networks, Espoo, Finland, June 14-17, 2011, Proceedings, Part I*. Vol. 1. 2 vols. Lecture Notes in Computer Science 6791. Heidelberg et al.: Springer, 2011. <http://link.springer.com/book/10.1007/978-3-642-21735-7/page/1>.
 57. Honkela, Timo, Włodzisław Duch, Mark A. Girolami, and Samuel Kaski, eds. *Artificial Neural Networks and Machine Learning - ICANN 2011 - 21st International Conference on Artificial Neural Networks, Espoo, Finland, June 14-17, 2011, Proceedings, Part II*. Vol. 2. 2 vols. Lecture Notes in Computer Science 6792. Heidelberg et al.: Springer, 2011. <http://www.springer.com/computer/ai/book/978-3-642-21737-1>.
 58. Honkela, Timo, Juha Raitio, Krista Lagus, Ilari T. Nieminen, Nina Honkela, and Mika Pantzar. “Subjects on Objects in Contexts: Using GICA Method to Quantify Epistemological Subjectivity.” In *Proceedings of the International Joint Conference on Neural Networks (IJCNN 2012)*, 2875–2883. Brisbane, Australia, 2012. <http://users.ics.aalto.fi/tho/online-papers/GICA2012.pdf>.
 59. Huck, Matthias, Stephan Peitz, Markus Freitag, and Hermann Ney. “Discriminative Reordering Extensions for Hierarchical Phrase-Based Machine Translation.” In *Proceedings of 16th Annual Conference of the European Association for Machine Translation (EAMT 2012)*, 313–320. Trento, Italy, 2012. <http://www-i6.informatik.rwth-aachen.de/publications/download/776/HuckMatthiasPeitzStephanFreitagMarkusNeyHermann--DiscriminativeReorderingExtensionsforHierarchicalPhrase-BasedMachineTranslation--2012.pdf>.
 60. Hunsicker, Sabine, Chen Yu, and Christian Federmann. “Machine Learning for Hybrid Machine Translation.” In *Proceedings of the 7th Workshop on Statistical Machine Translation (WMT 2012)*, 312–316. Montréal, Canada: Association for Computational Linguistics (ACL), 2012. <http://www.aclweb.org/anthology-new/W/W12/W12-3138.pdf>.
 61. Jörg, Brigitte, Jan Dvořák, and Thomas Vestdam. “Streamlining the CERIF XML Data Exchange Format Towards CERIF 2.0.” In *Proceedings of the 11th International Conference on Current Research Information Systems (CRIS 2012)*, 221–230. Prague, Czech Republic, 2012. http://www.eurocris.org/Uploads/Web%20pages/CRIS%202012%20-%20Prague/CRIS2012_24_full_paper.pdf.
 62. Jörg, Brigitte, Thorsten Höllrigl, and Miguel-Angel Sicilia. “Entities and Identities in Research Information Systems.” In *Proceedings of the 11th International Conference on Current Research Information Systems (CRIS 2012)*, 184–195. Prague, Czech Republic, 2012. http://www.eurocris.org/Uploads/Web%20pages/CRIS%202012%20-%20Prague/CRIS2012_20_full_paper.pdf.
 63. Jörg, Brigitte, Ivan Ruiz-Rube, Miguel-Angel Sicilia, Jan Dvořák, Keith Jeffery, Thorsten Höllrigl, Henrik S. Rasmussen, Andreas Engfer, Thomas Vestdam, and Elena Garcia Barriocanal. “Connecting Closed World Research Information Systems through the Linked Open Data Web.” *International Journal of Software Engineering and Knowledge Engineering* 22, no. 03 (May 2012): 345–364.
 64. Jörg, Brigitte, Hans Uszkoreit, and Alastair Burt. “LT World: Ontology and Reference Information Portal.” In *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC 2010)*, 1002–1006. Valletta, Malta: European Language Resources Association (ELRA), 2010. http://www.lrec-conf.org/proceedings/lrec2010/pdf/714_Paper.pdf.
 65. Judge, John, Ailbhe Ní Chasaide, Rose Ní Dhubhda, Kevin P. Scannell, and Elaine Uí Dhonnchadha. *The Irish Language in the Digital Age / An Ghaeilge sa Ré Dhigiteach*. Edited by Georg Rehm and Hans Uszkoreit. META-NET White Paper Series “Europe’s Languages in the Digital Age”. Heidelberg et al.: Springer, 2012. <http://www.meta-net.eu/whitepapers/e-book/irish.pdf>.
 66. Koutsombogera, Maria, and Stelios Piperidis. *Vision Document. Vision Group Media and Information Services: Results of First Two Meetings*. Athens, Greece, November 10, 2010. <http://www.meta-net.eu/vision/reports/VisionGroup-MediaInformationServices-draft.pdf>.

67. Krek, Simon. *The Slovene Language in the Digital Age / Slovenski jezik v digitalni dobi*. Edited by Georg Rehm and Hans Uszkoreit. META-NET White Paper Series “Europe’s Languages in the Digital Age”. Heidelberg et al.: Springer, 2012. <http://www.meta-net.eu/whitepapers/e-book/slovene.pdf>.
68. Lardilleux, Adrien, Yves Lepage, and François Yvon. “The Contribution of Low Frequencies to Multilingual Sub-sentential Alignment: a Differential Associative Approach.” *International Journal of Advanced Intelligence* 3, no. 2 (July 2011): 189–217.
69. Lardilleux, Adrien, François Yvon, and Yves Lepage. “Hierarchical Sub-sentential Alignment with Anymalign.” In *Proceedings of 16th Annual Conference of the European Association for Machine Translation (EAMT 2012)*, 279–286. Trento, Italy: European Association for Machine Translation, 2012. <http://hnk.ffzg.hr/bibl/eamt2012/html/Papers/61.pdf>.
70. Lardilleux, Adrien, François Yvon, and Yves Lepage. “Généralisation de l’alignement sous-phrastique par échantillonnage.” In *Actes de la 18e Conférence sur le Traitement Automatique des Langues Naturelles (TALN 2011)*. Montpellier, France, 2011. http://hal.archives-ouvertes.fr/docs/00/60/64/00/PDF/Lardilleux_taln11_submission_14.pdf.
71. Lavergne, Thomas, Josep Maria Crego, Alexandre Allauzen, and François Yvon. “From N-gram-based to CRF-based Translation Models.” In *Proceedings of the 6th Workshop on Statistical Machine Translation (WMT 2011)*, 542–553. Edinburgh, Scotland: Association for Computational Linguistics (ACL), 2011. <http://www.statmt.org/wmt11/WMT-2011.pdf>.
72. Le, Hai-Son, Alexandre Allauzen, and François Yvon. “Continuous Space Translation Models with Neural Networks.” In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 39–48. Montréal, Canada: Association for Computational Linguistics (ACL), 2012. <http://www.aclweb.org/anthology-new/N/N12/N12-1005.pdf>.
73. Le, Hai-Son, Alexandre Allauzen, and François Yvon. “Measuring the Influence of Long Range Dependencies with Neural Network Language Models.” In *NAACL-HLT 2012 Workshop: Will We Ever Really Replace the N-gram Model? On the Future of Language Modeling for HLT*, 1–10. Montréal, Canada: Association for Computational Linguistics (ACL), 2012. <http://aclweb.org/anthology-new/W/W12/W12-2701.pdf>.
74. Le, Hai-Son, Thomas Lavergne, Alexandre Allauzen, Marianna Apidianaki, Li Gong, Aurélien Max, Artem Sokolov, Guillaume Wisniewski, and François Yvon. “LIMSIS @ WMT’12.” In *Proceedings of the 7th Workshop on Statistical Machine Translation (WMT 2012)*, 330–337. Montréal, Canada: Association for Computational Linguistics (ACL), 2012. <http://www.statmt.org/wmt12/pdf/WMT41.pdf>.
75. Li, Maoxi, and Mingwen Wang. “Confusion Network Based System Combination for Chinese Translation Output: Word-Level or Character-Level?” In *Proceedings of the Second Shared Task on Applying Machine Learning Techniques to Optimise the Division of Labour in Hybrid Machine Translation (ML4HMT-12)*, 17–24. Mumbai, India, 2012. <http://www.aclweb.org/anthology-new/W/W12/W12-5702.pdf>.
76. Liin, Krista, Kadri Muischnek, Kaili Müürisep, and Kadri Vider. *The Estonian Language in the Digital Age / Eesti keel digiajastul*. Edited by Georg Rehm and Hans Uszkoreit. META-NET White Paper Series “Europe’s Languages in the Digital Age”. Heidelberg et al.: Springer, 2012. <http://www.meta-net.eu/whitepapers/e-book/estonian.pdf>.
77. Lindh-Knuutila, Tiina, Jaakko Väyrynen, and Timo Honkela. “Semantic Analysis in Word Vector Spaces with ICA and Feature Selection.” In *Proceedings of the 11th Conference on Natural Language Processing (KONVENS2012)*, 98–107. Vienna, Austria, 2012. http://www.oegai.at/konvens2012/proceedings/15_lindhknuutila120/15_lindhknuutila120.pdf.
78. Malmi, Eric, Juha Raitio, Oskar Kohonen, Krista Lagus, and Timo Honkela. “Identifying Anomalous Social Contexts from Mobile Proximity Data Using Binomial Mixture Models.” In *Advances in Intelligent Data Analysis XI*, edited by Jaakko Hollmén, Frank Klawonn, and Allan Tucker, 195–206. 11th International Symposium. Heidelberg et al.: Springer, 2012. http://link.springer.com/chapter/10.1007%2F978-3-642-34156-4_19.
79. Mareček, David, Martin Popel, and Zdeněk Žabokrtský. “Maximum Entropy Translation Model in Dependency-Based MT Framework.” In *Proceedings of the Joint 5th Workshop on Statistical Machine Translation and MetricsMATR*, 207–212. Uppsala, Sweden: Association for Computational Linguistics (ACL), 2010. <http://www.statmt.org/wmt10/pdf/WMT31.pdf>.
80. Mariani, Joseph. “Les technologies de la langue en soutien au multilinguisme.” *Net.lang. Réussir le cyberspace multilingue* 65 (2012): 149–170.
81. Mariani, Joseph, and Bernardo Magnini. *Vision Document. Vision Group Interactive Systems: Results of First Two Meetings*. Paris, France, December 28, 2010. <http://www.meta-net.eu/vision/reports/VisionGroup-InteractiveSystems-draft.pdf>.
82. Mariani, Joseph, Patrick Paroubek, Gil Francopoulo, Aurélien Max, François Yvon, and Pierre Zweigenbaum. *The French Language in the Digital Age / La langue française à l’Ère du numérique*. Edited by Georg Rehm

- and Hans Uszkoreit. META-NET White Paper Series “Europe’s Languages in the Digital Age”. Heidelberg et al.: Springer, 2012. <http://www.meta-net.eu/whitepapers/e-book/french.pdf>.
83. Mehdad, Yashar, Matteo Negri, and Marcello Federico. “Match Without a Referee: Evaluating MT Adequacy Without Reference Translations.” In *Proceedings of the 7th Workshop on Statistical Machine Translation (WMT 2012)*, 171–180. Montréal, Canada: Association for Computational Linguistics (ACL), 2012. <http://www.aclweb.org/anthology-new/W/W12/W12-3122.pdf>.
 84. Mehmood, Yasir, and Timo Honkela. “Comparing Talks, Realities and Concerns over the Climate Change: Comparing Texts with Numerical and Categorical Data.” In *Proceedings of Conference on Language and Technology 2012 (CLT 2012)*, 21–27. Lahore, Pakistan: Society for Natural Language Processing, 2012. <http://www.cle.org.pk/clt12/pdf/4.pdf>.
 85. Melero, Maite, Toni Badia, and Asunción Moreno. *The Spanish Language in the Digital Age / La lengua española en la era digital*. Edited by Georg Rehm and Hans Uszkoreit. META-NET White Paper Series “Europe’s Languages in the Digital Age”. Heidelberg et al.: Springer, 2012. <http://www.meta-net.eu/whitepapers/e-book/spanish.pdf>.
 86. Meyer, Thomas, and Andrej Popescu-Belis. “Using Sense-labeled Discourse Connectives for Statistical for Statistical Machine Translation.” In *Proceedings of the EACL 2012 Joint ESIRMT-HyTra Workshop (Exploiting Synergies Between IR and MT, and Hybrid Approaches to MT)*, 129–138. Avignon, France: Association for Computational Linguistics (ACL), 2012. <http://aclweb.org/anthology-new/W/W12/W12-0117.pdf>.
 87. Meyer, Thomas, Andrej Popescu-Belis, and Najeh Hajlaoui. “Machine Translation of Labeled Discourse Connectives.” In *Proceedings of the 10th Biennial Conference of the Association for Machine Translation in the Americas (AMTA 2012)*. San Diego, USA, 2012. <http://amta2012.amtaweb.org/AMTA2012Files/papers/119.pdf>.
 88. Moreno, Asunción, Núria Bel, Maite Melero, Carmen García Mateo, Inmaculada Hernández, Sergio Oller, Aljoscha Burchardt, Kathrin Eichler, Georg Rehm, and Hans Uszkoreit. “META-NET. Official Languages of Spain in the Digital Age.” In *Proceedings of VII Jornadas En Tecnología Del Habla and III Iberian SLTech Workshop (IberSPEECH 2012)*, 352–361. Madrid, Spain, 2012. http://iberspeech2012.ii.uam.es/IberSPEECH2012_OnlineProceedings.pdf.
 89. Muhič, Andrej, Jan Rupnik, and Primož Škraba. “Cross-lingual Document Similarity.” In *Proceedings of the 34th International Conference on Information Technology Interfaces (ITI 2012)*, 387–392. Ljubljana, Slovenia, 2012. http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=6308038&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxppls%2Fabs_all.jsp%3Farnumber%3D6308038.
 90. Nikoulina, Vassilina, Agnes Sandor, and Marc Dymetman. “Hybrid Adaptation of Named Entity Recognition for Statistical Machine Translation.” In *Proceedings of the Second Shared Task on Applying Machine Learning Techniques to Optimise the Division of Labour in Hybrid Machine Translation (ML4HMT-12)*, 1–16. Mumbai, India, 2012. <http://www.aclweb.org/anthology-new/W/W12/W12-5701.pdf>.
 91. Odijk, Jan. *The Dutch Language in the Digital Age / Het Nederlands in het Digitale Tijdperk*. Edited by Georg Rehm and Hans Uszkoreit. META-NET White Paper Series “Europe’s Languages in the Digital Age”. Heidelberg et al.: Springer, 2012. <http://www.meta-net.eu/whitepapers/e-book/dutch.pdf>.
 92. Okita, Tsuyoshi. “Neural Probabilistic Language Model for System Combination.” In *Proceedings of the Second Shared Task on Applying Machine Learning Techniques to Optimise the Division of Labour in Hybrid Machine Translation (ML4HMT-12)*, 65–76. Mumbai, India, 2012. <http://www.aclweb.org/anthology-new/W/W12/W12-5707.pdf>.
 93. Okita, Tsuyoshi, and Josef van Genabith. “Minimum Bayes Risk Decoding with Enlarged Hypothesis Space in System Combination.” In *Computational Linguistics and Intelligent Text Processing: 13th International Conference (CICLing 2012)*, 40–51. New Delhi, India: Springer, 2012. <http://dl.acm.org/citation.cfm?id=2238696.2238702>.
 94. Okita, Tsuyoshi, and Josef van Genabith. “DCU Confusion Network-based System Combination for ML4HMT.” In *Proceedings of the International Workshop on Using Linguistic Information for Hybrid Machine Translation (LIHMT 2011) and of the Shared Task on Applying Machine Learning Techniques to Optimise the Division of Labour in Hybrid Machine Translation (ML4HMT-11)*, 93–98. Barcelona, Spain: Universitat Politècnica de Catalunya, 2011. <http://ixa2.si.ehu.es/lihmt2011/proceedings.pdf>.
 95. Okita, Tsuyoshi, Raphaël Rubino, and Josef van Genabith. “Sentence-level Quality Estimation for MT System Combination.” In *Proceedings of the Second Shared Task on Applying Machine Learning Techniques to Optimise the Division of Labour in Hybrid Machine Translation (ML4HMT-12)*, 55–64. Mumbai, India, 2012. <http://www.aclweb.org/anthology-new/W/W12/W12-5706.pdf>.
 96. Okita, Tsuyoshi, Antonio Toral, and Josef van Genabith. “Topic Modeling-based Domain Adaptation for System Combination.” In *Proceedings of the Second Shared Task on Applying Machine Learning Techni-*

- ques to Optimise the Division of Labour in Hybrid Machine Translation (ML4HMT-12), 45–54. Mumbai, India, 2012. <http://www.aclweb.org/anthology-new/W/W12/W12-5705.pdf>.
97. Paukkeri, Mari-Sanna, Alberto Pérez García-Plaza, Víctor Fresno, Raquel Martínez Unanue, and Timo Honkela. “Learning a Taxonomy from a Set of Text Documents.” *Applied Soft Computing* 12, no. 3 (March 2012): 1138–1148.
 98. Paukkeri, Mari-Sanna, Ilkka Kivimäki, Santosh Tirunagari, Erkki Oja, and Timo Honkela. “Effect of Dimensionality Reduction on Different Distance Measures in Document Clustering.” In *Neural Information Processing*, edited by Bao-Liang Lu, Liqing Zhang, and James Kwok, 167–176. 18th International Conference, ICONIP 2011, Shanghai, China, November 13–17, 2011, Proceedings, Part III. Heidelberg et al.: Springer, 2011. http://link.springer.com/chapter/10.1007%2F978-3-642-24965-5_19#page-1.
 99. Paukkeri, Mari-Sanna, Marja Ollikainen, and Timo Honkela. “Assessing User-specific Difficulty of Documents.” *Information Processing & Management* 49, no. 1 (January 2013): 198–212.
 100. Paukkeri, Mari-Sanna, Jaakko Väyrynen, and Antti Arppe. “Exploring Extensive Linguistic Feature Sets in Near-synonym Lexical Choice.” In *Computational Linguistics and Intelligent Text Processing: 13th International Conference (CICLing 2012)*, edited by Alexander Gelbukh, 1–12. New Delhi, India: Springer, 2012. http://link.springer.com/content/pdf/10.1007%2F978-3-642-28601-8_1.pdf.
 101. Pedersen, Bolette Sandford, Jürgen Wedekind, Steen Bøhm-Andersen, Peter J. Henrichsen, Sanne Hoffensetz-Andresen, Sabine Kirchmeier-Andersen, Jens Kjærum, et al. *The Danish Language in the Digital Age / Det danske sprog i den digitale tidsalder*. Edited by Georg Rehm and Hans Uszkoreit. META-NET White Paper Series “Europe’s Languages in the Digital Age”. Heidelberg et al.: Springer, 2012. <http://www.meta-net.eu/whitepapers/e-book/danish.pdf>.
 102. Peitz, Stephan, Arne Mauser, Joern Wuebker, and Hermann Ney. “Forced Derivations for Hierarchical Machine Translation.” In *Proceedings of the 24th International Conference on Computational Linguistics (COLING 2012), Papers*, 933–942. Mumbai, India, 2012. <http://aclweb.org/anthology-new/C/C12/C12-2091.pdf>.
 103. Peter, Jan-Thorsten, Matthias Huck, Hermann Ney, and Daniel Stein. “Soft String-to-Dependency Hierarchical Machine Translation.” In *Proceedings of the 8th International Workshop on Spoken Language Translation (IWSLT 2011)*, 246–253. San Francisco, USA, 2011. http://iwslt2011.org/doku.php?id=o8_proceedings.
 104. Piperidis, Stelios. “The META-SHARE Language Resources Sharing Infrastructure: Principles, Challenges, Solutions.” In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012)*, 36–42. Istanbul, Turkey: European Language Resources Association (ELRA), 2012. http://www.lrec-conf.org/proceedings/lrec2012/pdf/1086_Paper.pdf.
 105. Raitio, Juha, Tapani Raiko, and Timo Honkela. “Hybrid Bilinear and Trilinear Models for Exploratory Analysis of Three-way Poisson Counts.” In *Proceedings of the 22nd International Conference on Artificial Neural Networks and Machine Learning (ICANN 2012)*, edited by Alessandro E. Villa, Włodzisław Duch, Péter Érdi, Francesco Masuli, and Günther Palm, 475–482. Heidelberg et al.: Springer, 2012. http://link.springer.com/chapter/10.1007%2F978-3-642-33266-1_59.
 106. Rehm, Georg. “The META-NET Strategic Research Agenda for Language Technology in Europe.” *Translation: Computation, Corpora, Cognition* 2, no. 1. Special Issue on Language Technology for Multilingual Europe (2013): in print.
 107. Rehm, Georg. “META-NET: A Research and Technology Strategy for Multilingual Europe.” In *Jazykovedné Štúdie □. 31./Linguistic Studies No. 31. Proceedings of the Conference Rozvoj Jazykových Technológií a Zdrojov Na Slovensku a Vo Svete (10 Rokov Slovenského Národného korpusu)/Development of the Human Language Technologies and Resources in Slovakia and in the World (10 Years of the Slovak National Corpus)*, edited by Katarína Gajdošová and Adriána Žáková, in print. Bratislava: Veda, 2013.
 108. Rehm, Georg. “Sprachgrenzen sind Marktgrenzen – Technologien für ein sprachübergreifendes Informations- und Wissensmanagement.” *Wissensmanagement* 4 (June 2013): 14–16.
 109. Rehm, Georg, Aljoscha Burchardt, and Felix Sasaki. “Das mehrsprachige Europa – eine Herausforderung für die Sprachtechnologie.” *Libreas* 18 (2011): 30–37.
 110. Rehm, Georg, and Hans Uszkoreit. “An Extended Summary of the META-NET Strategic Research Agenda for Multilingual Europe 2020.” *Contributions to the Annual Conference 2012 of EFNIL in Budapest* (2013): in print.
 111. Rehm, Georg, and Hans Uszkoreit, eds. *The META-NET Strategic Research Agenda for Multilingual Europe 2020*. Heidelberg et al.: Springer, 2013. <http://link.springer.com/content/pdf/bfm%3A978-3-642-36349-8%2F1.pdf>.

112. Rehm, Georg, and Hans Uszkoreit. "Language Technology for Multilingual Europe: Current Status and Future Prospects." In *Language Education in Creating a Multilingual Europe. Contributions to the Annual Conference 2011 of EFNIL in London*, edited by Gerhard Stickel and Michael Carrier, 94:145–159. Frankfurt/Main: Lang, 2012. <http://www.peterlang.com/index.cfm?event=cmp.ccc.seitenstruktur.detailseiten&seitentyp=produkt&pk=71772&concordeid=263990>.
113. Rehm, Georg, and Hans Uszkoreit. "META-NET Strategic Research Agenda for Multilingual Europe 2020: An Extended Summary." Edited by Dan Tufiş, Vasile Rus, and Corina Forascu. *Towards Multilingual Europe 2020: A Romanian Perspective*. Editura Academiei Române (2012): 9–33.
114. Rehm, Georg, and Hans Uszkoreit, eds. *The META-NET White Paper Series – Europe's Languages in the Digital Age. 31 Volumes (ca. 3,000 Pages) Reporting on the Level of Support through Language Technology for 30 European Languages*. Heidelberg et al.: Springer, 2012. <http://www.meta-net.eu/whitepapers/overview>.
115. Rehm, Georg, and Hans Uszkoreit. "Multilingual Europe: A Challenge for Language Tech." *MultiLingual* (May 2011): 51–52.
116. Rögnvaldsson, Eiríkur, Kristín M. Jóhannsdóttir, Sigrún Helgadóttir, and Steinþór Steingrímsson. *The Icelandic Language in the Digital Age / Íslensk tunga á stafrænni öld*. Edited by Georg Rehm and Hans Uszkoreit. META-NET White Paper Series "Europe's Languages in the Digital Age". Heidelberg et al.: Springer, 2012. <http://www.meta-net.eu/whitepapers/e-book/icelandic.pdf>.
117. Ruiz, Nick, Arianna Bisazza, Fabio Brugnara, Daniele Falavigna, Diego Giuliani, Suhel Jaber, Roberto Gretter, and Marcello Federico. "FBK @ IWSLT 2011." In *Proceedings of the 8th International Workshop on Spoken Language Translation (IWSLT 2011)*, 86–93. San Francisco, USA, 2011. http://iwslt2011.org/doku.php?id=o8_proceedings.
118. Ruiz, Nick, and Marcello Federico. "MDI Adaptation for the Lazy: Avoiding Normalization in LM Adaptation for Lecture Translation." In *Proceedings of the 9th International Workshop on Spoken Language Translation (IWSLT 2012)*, 244–251. Hong Kong, China, 2012. <http://hltc.cs.ust.hk/iwslt/index.php/program.html>.
119. Ruiz, Nick, and Marcello Federico. "Topic Adaptation for Lecture Translation through Bilingual Latent Semantic Models." In *Proceedings of the 6th Workshop on Statistical Machine Translation (WMT 2011)*, 294–302. Edinburgh, Scotland: Association for Computational Linguistics (ACL), 2011. <http://aclweb.org/anthology-new/W/W11/W11-2133.pdf>.
120. Rupnik, Jan, Andrej Muhič, and Primož Škraba. "Cross-Lingual Document Retrieval through Hub Languages." In *xLiTe: Cross-Lingual Technologies, NIPS 2012 Workshop*, 1–5. Lake Tahoe, USA, 2012. <http://km.aifb.kit.edu/ws/xlite/xLiTe-Paper-AndrejMuhic.pdf>.
121. Rupnik, Jan, Andrej Muhič, and Primož Škraba. "Multilingual Document Retrieval through Hub Languages." In *Conference on Data Mining and Data Warehouses (SiKDD 2012)*, 1–4. Ljubljana, Slovenia, 2012. http://ailab.ijs.si/dunja/SiKDD2012/Papers/Rupnik_Multilingual.pdf.
122. Rupnik, Jan, Andrej Muhič, and Primož Škraba. "Low-Rank Approximations for Large, Multi-lingual Data." In *Proceedings of Neural Information Processing Systems (NIPS 2011)*. Sierra Nevada, Spain, 2011. http://ailab.ijs.si/primož_skraba/papers/nips_full.pdf.
123. Russo, Irene, Tommaso Caselli, Francesco Rubino, Ester Boldrini, and Patricio Martínez-Barco. "EMOCause: An Easy-adaptable Approach to Emotion Cause Contexts." In *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA 2.011)*, 153–160. Portland, USA: Association for Computational Linguistics (ACL), 2011.
124. Samuelsson, Christer. "HAL: Challenging Three Key Aspects of IBM-style Statistical Machine Translation." In *Proceedings of the 10th Biennial Conference of the Association for Machine Translation in the Americas (AMTA 2012)*. San Diego, USA, 2012. <http://amta2012.amtaweb.org/AMTA2012Files/papers/103.pdf>.
125. Skadiņa, Inguna, Andrejs Veisbergs, Andrejs Vasiljevs, Tatjana Gornostaja, Iveta Keiša, and Alda Rudzīte. *The Latvian Language in the Digital Age / Latviešu valoda digitālajā laikmetā*. Edited by Georg Rehm and Hans Uszkoreit. META-NET White Paper Series "Europe's Languages in the Digital Age". Heidelberg et al.: Springer, 2012. <http://www.meta-net.eu/whitepapers/e-book/latvian.pdf>.
126. Smedt, Koenraad De, Gunn I. Lyse, Anje Müller Gjesdal, and Gyri S. Losnegaard. *The Norwegian Language in the Digital Age (Bokmål Version) / Norsk i den digitale tidsalderen (bokmålsversjon)*. Edited by Georg Rehm and Hans Uszkoreit. META-NET White Paper Series "Europe's Languages in the Digital Age". Heidelberg et al.: Springer, 2012. <http://www.meta-net.eu/whitepapers/e-book/norwegian-bokmaal.pdf>.
127. Smedt, Koenraad De, Gunn I. Lyse, Anje Müller Gjesdal, and Gyri S. Losnegaard. *The Norwegian Language in the Digital Age (Nynorsk Version) / Norsk i den digitale tidsalderen (nynorskversjon)*. Edited by Georg Rehm and Hans Uszkoreit. META-NET White Paper Series "Europe's Languages in the Digital Age". Heidelberg et al.: Springer, 2012. <http://www.meta-net.eu/whitepapers/e-book/norwegian-nynorsk.pdf>.

128. Sokolov, Artem, Guillaume Wisniewski, and François Yvon. “Computing Lattice BLEU Oracle Scores for Machine Translation.” In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2012)*, 120–129. Avignon, France: Association for Computational Linguistics (ACL), 2012. <http://aclweb.org/anthology-new/E/E12/E12-1013.pdf>.
129. Sokolov, Artem, Guillaume Wisniewski, and François Yvon. “Non-Linear N-best Reranking with Few Features.” In *Proceedings of the 10th Biennial Conference of the Association for Machine Translation in the Americas (AMTA 2012)*. San Diego, USA, 2012. <http://amta2012.amtaweb.org/AMTA2012Files/papers/116.pdf>.
130. Sokolov, Artem, and François Yvon. “Minimum Error Rate Training Semiring.” In *Proceedings of the 15th International Conference of the European Association for Machine Translation (EAMT 2011)*, 241–248. Leuven, Belgium: European Association for Machine Translation, 2011. <http://www.ccl.kuleuven.be/EAMT2011/proceedings/pdf/eamt2011proceedings.pdf>.
131. Stein, Daniel, Stephan Peitz, David Vilar, and Hermann Ney. “A Cocktail of Deep Syntactic Features for Hierarchical Machine Translation.” In *Proceedings of the 9th Conference of the Association for Machine Translation in the Americas (AMTA 2010)*. Denver, USA, 2010. <http://www-i6.informatik.rwth-aachen.de/publications/download/674/Stein-AMTA%202010-2010.pdf>.
132. Tomeh, Nadi, Alexandre Allauzen, Thomas Lavergne, and François Yvon. “Designing an Improved Discriminative Word Aligner.” In *Proceedings of the 12th International Conference on Intelligent Text Processing and Computational Linguistics*. Tokyo, Japan, 2011. <http://oniros.org/papers/tomeh2011cicling.pdf>.
133. Tomeh, Nadi, Alexandre Allauzen, and François Yvon. “Discriminative Weighted Alignment Matrices For Statistical Machine Translation.” In *Proceedings of the 15th International Conference of the European Association for Machine Translation (EAMT 2011)*, 305–312. Leuven, Belgium: European Association for Machine Translation, 2011. <http://www.ccl.kuleuven.be/EAMT2011/proceedings/pdf/eamt2011proceedings.pdf>.
134. Virpioja, Sami, Mari-Sanna Paukkeri, Abishek Tripathi, Tiina Lindh-Knuutila, and Krista Lagus. “Evaluating Vector Space Models with Canonical Correlation Analysis.” *Natural Language Engineering* 18, no. 03 (2012): 399–436.
135. Weitzmann, John Hendrik, Georg Rehm, and Hans Uszkoreit. “Licensing and Sharing Language Resources: An Approach Inspired by Creative Commons and Open Science Data Movements.” In *Proceedings of 7th International Language Resources and Evaluation Conference (LREC 2010), Workshop on Legal Issues for Sharing Language Resources: Constraints and Best Practices*. Valletta, Malta: European Language Resources Association (ELRA), 2010.
136. Wu, Xiaofeng, Tsuyoshi Okita, Josef van Genabith, and Qun Liu. “System Combination with Extra Alignment Information.” In *Proceedings of the Second Shared Task on Applying Machine Learning Techniques to Optimise the Division of Labour in Hybrid Machine Translation (ML4HMT-12)*, 37–44. Mumbai, India, 2012. <http://www.aclweb.org/anthology-new/W/W12/W12-5704.pdf>.
137. Wuebker, Joern, Matthias Huck, Stephan Peitz, Malte Nuhn, Markus Freitag, Jan-Thorsten Peter, Saab Mansour, and Hermann Ney. “Jane 2: Source Phrase-based and Hierarchical Statistical Machine Translation.” In *Proceedings of the 24th International Conference on Computational Linguistics (COLING 2012), Demonstration Papers*, 483–492. Mumbai, India, 2012. <http://aclweb.org/anthology-new/C/C12/C12-3061.pdf>.
138. Zeman, Daniel. “CUNI: Feature Selection and Error Analysis of a Transition-Based Parser.” In *Proceedings of the Workshop on Machine Translation and Parsing in Indian Languages (MTPIL-2012)*, 143–148. Mumbai, India: Association for Computational Linguistics (ACL), 2012. <http://www.aclweb.org/anthology/W/W12/W12-5614.pdf>.
139. Zhuang, Yong, Guillaume Wisniewski, and François Yvon. “Non-Linear Models for Confidence Estimation.” In *Proceedings of the 7th Workshop on Statistical Machine Translation (WMT 2012)*, 157–162. Montréal, Canada: Association for Computational Linguistics (ACL), 2012. <http://aclweb.org/anthology-new/W/W12/W12-3120.pdf>.

4.3 Presentations

1. Adda, Gilles, and Joseph Mariani. “Language Resources and Amazon Mechanical Turk: Ethical, Legal, and Other Issues.” Workshop Talk presented at the 7th International Conference on Language Resources and Evaluation (LREC 2010), Workshop on Legal Issues for Language Resources, Valletta, Malta, May 17, 2010. <http://www.mt-archive.info/LREC-2010-Adda.pdf>.
2. Burchardt, Aljoscha. “Technological Support of AV Media Translation Including Subtitling and Dubbing.” Panel presented at Languages and the Media 2010, Berlin, Germany, October 7, 2010. <http://www.meta-net.eu/events/languages-and-the-media-2010-panel>.

Final Report

3. Calzolari, Nicoletta. “Large Scale Annotation Efforts.” Presentation at the METAnk 2010, Berlin, Germany, June 4, 2010. http://videlectures.net/translingeu2010_calzolari_resources/.
4. Choukri, Khalid. “ELRA & META-SHARE The European Language Resources Association and The New Paradigm/trend of Data Sharing.” Presentation at the META-NET Network Meeting and General Assembly, Berlin, Germany, January 25, 2013. <http://www.meta-net.eu/events/meta-net-ga-2013/talks/khalidchoukri.pdf>.
5. Eichler, Kathrin, Aljoscha Burchardt, Georg Rehm, Hans Uszkoreit, Jan Odijk, and Joseph Mariani. “META-NET Language White Paper Series: Current State of Play, Tables and Clusters, Arriving at a Consensus, Discussion, Timeplan.” Presentation at the META-NET Network Meeting and General Assembly. Current State of Play and Next Steps, Berlin, Germany, October 21, 2011. <http://www.meta-net.eu/events/meta-net-ga-2011/talks/KathrinEichler.pdf>.
6. Gavrilidou, Maria, Penny Labropoulou, Elina Despri, and Stelios Piperidis. “Preliminary Proposal for a Metadata Schema for the Description of Language Resources.” presented at the 7th International Conference on Language Resources and Evaluation (LREC 2010), Workshop on Language Resource and Language Technology Standards – state of the art, emerging needs, and future developments, Valletta, Malta, May 17, 2010.
7. Genabith, Josef van. “Translingual Cloud.” Presentation at Deutscher Sprachtechnologietag 2013. Sprachtechnologie heute und morgen: Perspektiven und Strategien, Berlin, Germany, January 24, 2013. <http://www.meta-net.eu/events/german-lt-day-2013/talks/josefvangenabith.pdf>.
8. Genabith, Josef van. “Preliminary Findings of the Vision Group ‘Translation and Localisation’.” Presentation at META Technology Council, Brussels, Belgium, November 16, 2010. <http://www.meta-net.eu/vision/reports/visiongroup-translationlocalisation-council-2010-presentation.pdf>.
9. Genabith, Josef van. “An Overview of Multilingual LT Research in Europe.” Presentation at META-FORUM 2011 Solutions for Multilingual Europe, Budapest, Hungary, June 27, 2011. <http://www.meta-net.eu/events/meta-forum-2011/programme>.
10. Genabith, Josef van. “Evaluation, Localization and Open Source Tools in EuroMatrixPlus.” Presentation at Translingual Europe 2010, Berlin, Germany, June 7, 2010. <http://www.meta-net.eu/events/tle-2010>.
11. Genabith, Josef van, and John Judge. “The Communication Working Group – Current State of Play, Issues, Next Steps, Events, Activities.” Presentation at the META-NET Network Meeting and General Assembly. Current State of Play and Next Steps, Berlin, Germany, October 21, 2011. <http://www.meta-net.eu/events/meta-net-ga-2011/talks/JohnJudge.pdf>.
12. Grobelnik, Marko. “Priority Theme 2: Social Intelligence and e-Participation.” Presentation at the META-NET Network Meeting and General Assembly, Berlin, Germany, January 25, 2013. <http://www.meta-net.eu/events/meta-net-ga-2013/talks/roadmap-markogrobelnik.pdf>.
13. Grobelnik, Marko. “Priority Theme 2: Social Intelligence and e-Participation.” Presentation at META-FORUM 2012 A Strategy for Multilingual Europe, Brussels, Belgium, June 20, 2012. <http://www.meta-net.eu/events/meta-forum-2012/programme>.
14. Hajič, Jan. “Strategic Research Agenda for Multilingual Europe 2020, Priority Theme 1: The Translingual Cloud.” Presentation at the META-NET Network Meeting and General Assembly, Berlin, Germany, January 25, 2013. <http://www.meta-net.eu/events/meta-net-ga-2013/talks/roadmap-janhajic.pdf>.
15. Hajič, Jan. “Deep Linguistic Information in Hybrid Machine Translation.” Invited Talk presented at the Second Workshop on Applying Machine Learning Techniques to Optimise the Division of Labour in Hybrid MT (ML4HMT-12), Mumbai, India, December 9, 2012. <https://github.com/cfedermann/ML4HMT-2012/blob/master/slides/ML4HMT-2012-jan-hajic-invited-talk-slides.ppt?raw=true>.
16. Hajič, Jan. “META-RESEARCH: Leading-Edge Machine Translation Research for Multilingual Europe.” Presentation at META-FORUM 2011 Solutions for Multilingual Europe, Budapest, Hungary, June 27, 2011. <http://www.meta-net.eu/events/meta-forum-2011/programme>.
17. Hajič, Jan. “Resources for Adding Semantics to Machine Translations.” Invited Talk presented at the 7th International Workshop on Spoken Language Translation (IWSLT 2010), Paris, France, December 3, 2010.
18. Hajič, Jan. “Building Bridges Using Innovative Approaches in Machine Translation.” Presentation at META-FORUM 2010 Challenges for Multilingual Europe, Brussels, Belgium, November 17, 2010. http://www.meta-net.eu/events/meta-forum-2010/slides/META-FORUM2010_Hajic.pdf.
19. Hajič, Jan, Qun Liu, Hans Uszkoreit, and Josef van Genabith. “ML4HMT-12 Panel.” Panel presented at the Second Workshop on Applying Machine Learning Techniques to Optimise the Division of Labour in Hybrid MT (ML4HMT-12), Mumbai, India, December 9, 2012. <https://github.com/cfedermann/ML4HMT-2012/blob/master/slides/ML4HMT-2012-panel-discussion-slides.pdf?raw=true>.

20. Honkela, Timo. "Multimodality Grounded Translation by Humans and Machines." Presentation at Tralogy II Futures in technologies for translation, Paris, France, January 18, 2013. http://webcast.in2p3.fr/videos-multimodality_grounded_translation_by_humans_and_machines.
21. Honkela, Timo. "The Analysis of Different Languages in Scientific Research and Publishing." Invited Talk to Finish Parliament on Language question in research: English versus national languages?, Helsinki, Finland, November 17, 2010.
22. Honkela, Timo. "The Use of Multilingual Language Technologies for Promoting Successful Scientific Communication Across Language Borders." Invited Talk to Finish Parliament on Language question in research: English versus national languages?, Helsinki, Finland, November 17, 2010.
23. Honkela, Timo, Oskar Kohonen, Jorma Laaksonen, Krista Lagus, Klaus Förger, Mats Sjöberg, Tapio Takala, Harri Valpola, and Paul Wagner. "Learning Semantics of Movement." Presentation at the 25th Annual Conference on Neural Information Processing Systems (NIPS 2011), Sierra Nevada, Spain, December 17, 2011. <http://media.nips.cc/Conferences/2011/NIPS-Workshops-Book-2011.pdf>.
24. Mariani, Joseph. "Priority Theme 3: Socially Aware Interactive Assistant." Presentation at the META-NET Network Meeting and General Assembly, Berlin, Germany, January 25, 2013. <http://www.meta-net.eu/events/meta-net-ga-2013/talks/roadmap-josephmariani.pdf>.
25. Mariani, Joseph. "An Historical Perspective on Language Resources and Evaluation in Europe." Presentation at LDC 20th Anniversary Workshop, Philadelphia, USA, September 6, 2012. http://www ldc.upenn.edu/About/FoLR_LDC20/mariani.pdf.
26. Mariani, Joseph. "The Future European Multilingual Information Society." Presentation at META-FORUM 2011 Solutions for Multilingual Europe, Budapest, Hungary, June 28, 2011. <http://www.meta-net.eu/events/meta-forum-2011/programme>.
27. Mariani, Joseph. "Research Priority Theme 3: Socially Aware Interactive Assistants." Presentation at META-FORUM 2012 A Strategy for Multilingual Europe, Brussels, Belgium, June 20, 2012. <http://www.meta-net.eu/events/meta-forum-2012/programme>.
28. Mariani, Joseph. "European Activities on Language Resources." Workshop Talk presented at Oriental-Cocosda 2010, Kathmandu, Nepal, November 24, 2010. http://desceco.org/O-COCOSDA2010/program_oc2010.pdf.
29. Mariani, Joseph. "Preliminary Findings of the 'Interactive Systems' Vision Group." Presentation at the META Technology Council, Brussels, Belgium, November 16, 2010. <http://www.meta-net.eu/vision/reports/visiongroup-interactivesystems-council-2010-presentation.pdf>.
30. Mariani, Joseph. "Activities in Language Resources and Evaluation in Europe." Presentation at the Interact Presidential Summit, Moffett Field, USA, November 8, 2010.
31. Mariani, Joseph. "Speech and Language Technologies Frameworks in the European Research Area." Seminar presented at the Multisound Seminar, Istanbul, Turkey, June 16, 2010.
32. Mariani, Joseph, and Gil Francopoulo. "Language Matrices and the Language Resource Impact Factor." Workshop Talk presented at the Parole Workshop, Lisbon, Spain, October 18, 2012.
33. Mariani, Joseph, Patrick Paroubek, Gil Francopoulo, Aurélien Max, François Yvon, and Pierre Zweigenbaum. "The White Paper on the French language in the Digital Age." Presentation at the III International Symposium on Multilingualism in Cyberspace (SIMC), Paris, France, November 21, 2012. <http://www.maayajo.org/IMG/SIMC/Mariani-SIMC2012LWPJMariani.ppt>.
34. Open, Stephan. "Social Intelligence und e-Participation." Presentation at Deutscher Sprachtechnologietag 2013. Sprachtechnologie heute und morgen: Perspektiven und Strategien, Berlin, Germany, January 24, 2013. <http://www.meta-net.eu/events/german-lt-day-2013/talks/stephanoopen.pdf>.
35. Piperidis, Stelios. "META-SHARE: An Overview." Presentation at the META-NET Network Meeting and General Assembly, Berlin, Germany, January 25, 2013. <http://www.meta-net.eu/events/meta-net-ga-2013/talks/steliospiperidis.pdf>.
36. Piperidis, Stelios. "Core Language Resources and Technologies." Presentation at the META-NET Network Meeting and General Assembly, Berlin, Germany, January 25, 2013. <http://www.meta-net.eu/events/meta-net-ga-2013/talks/roadmap-steliospiperidis.pdf>.
37. Piperidis, Stelios. "META-SHARE – Current State of Play and Next Steps." Presentation at the META-NET Network Meeting and General Assembly. Current State of Play and Next Steps, Berlin, Germany, October 21, 2011. <http://www.meta-net.eu/events/meta-net-ga-2011/talks/SteliosPiperidis.pdf>.
38. Piperidis, Stelios. "META-SHARE: An Open Resource Exchange Infrastructure for Stimulating Research and Innovation." Presentation at META-FORUM 2011 Solutions for Multilingual Europe, Budapest, Hungary, June 27, 2011. <http://www.meta-net.eu/events/meta-forum-2011/programme>.

39. Piperidis, Stelios. "Overview, Current State, Towards Version 3 of META-SHARE." Presentation at META-FORUM 2012 A Strategy for Multilingual Europe, Brussels, Belgium, June 21, 2012. <http://www.meta-net.eu/events/meta-forum-2012/programme>.
40. Piperidis, Stelios. "META-SHARE: The Open Resource Exchange Facility." Presentation at META-FORUM 2010 Challenges for Multilingual Europe, Brussels, Belgium, November 17, 2010. http://www.meta-net.eu/events/meta-forum-2010/slides/META-FORUM2010_Piperidis.pdf.
41. Piperidis, Stelios. "Results of Vision Group 'Media and Information Services'." Presentation at the META Technology Council, Brussels, Belgium, November 16, 2010. <http://www.meta-net.eu/vision/reports/visiongroup-mediainformationservices-council-2010-presentation.pdf>.
42. Piperidis, Stelios. "META-SHARE." Workshop Talk presented at COCODA/WRITe/FLaReNet Workshop on Language Technology Issues for International Cooperation, Valletta, Malta, May 22, 2010.
43. Rehm, Georg. "Update on the META-NET Strategic Research Agenda for Multilingual Europe 2020: Final Version and Next Steps." Poster presented at the MultilingualWeb Workshop, Rome, Italy, March 12, 2013. <http://www.multilingualweb.eu/en/documents/rome-workshop/rome-program>.
44. Rehm, Georg. "Language Technology for Multilingual Europe: Current State and Future Plans." Presentation at European Languages in the Age of Technology – quo vadis?, Vilnius, Lithuania, November 14, 2012.
45. Rehm, Georg. "META – Overview of Members, Gaps, Recruiting Additional Members." Presentation at the META-NET Network Meeting and General Assembly. Current State of Play and Next Steps, Berlin, Germany, October 21, 2011. <http://www.meta-net.eu/events/meta-net-ga-2011/talks/GeorgRehm-META.pdf>.
46. Rehm, Georg. "META-NET and LT Innovate: Introduction and Presentation of Partnership." Presentation at META-FORUM 2012 A Strategy for Multilingual Europe, Brussels, Belgium, June 20, 2012. <http://www.meta-net.eu/events/meta-forum-2012/programme>.
47. Rehm, Georg. "The META-NET Strategic Research Agenda and Linked Open Data." Presentation at the MultilingualWeb Workshop, Dublin, Ireland, June 11, 2012. <http://www.w3.org/International/multilingualweb/dublin/slides/15-rehm.pdf>.
48. Rehm, Georg. "META-NET and META-SHARE: An Overview." Keynote presented at HLT: The Baltic Perspective, Riga, Latvia, October 7, 2010.
49. Samuelsson, Christer. "Parametric Distributions for Statistical Machine Translation: What Do Alignment Probabilities Actually Look Like." Seminar, Amsterdam, Netherlands, February 1, 2012.
50. Sasaki, Felix. "Language Resources, Language Technologies, Text Mining, the Semantic Web: How Interoperability of Machines Can Help Humans in the Multilingual Web." Workshop presented at the W3C Workshop: The Multilingual Web - Where Are We?, Madrid, Spain, October 27, 2010. <http://www.w3.org/International/multilingualweb/madrid/slides/sasaki.pdf>.
51. Sasaki, Felix. "Metadata (and More...) for the Ubiquitous Multilingual Web." Presentation at the METank 2010, Berlin, Germany, June 4, 2010. http://videlectures.net/translingeu2010_sasaki_meta/.
52. Steinbiss, Volker. "Interactive Systems to Break the Communication Barriers." Presentation at Language Technology for a Multilingual Europe - Networking Session at ICT 2010, Brussels, Belgium, September 28, 2010.
53. Uszkoreit, Hans. "Die deutsche Sprache im digitalen Zeitalter: Wo stehen wir?" Presentation at Deutscher Sprachtechnologietag 2013. Sprachtechnologie heute und morgen: Perspektiven und Strategien, Berlin, Germany, January 24, 2013. <http://www.meta-net.eu/events/german-lt-day-2013/talks/hansuszkoreit.pdf>.
54. Uszkoreit, Hans. "Die META-NET SRA: Die strategische Forschungsagenda für ein mehrsprachiges Europa 2020." Presentation at Deutscher Sprachtechnologietag 2013. Sprachtechnologie heute und morgen: Perspektiven und Strategien, Berlin, Germany, January 24, 2013. <http://www.meta-net.eu/events/german-lt-day-2013/talks/hansuszkoreit.pdf>.
55. Uszkoreit, Hans. "European Language Technology: Where Do We Stand – in a Nutshell." Presentation at META-FORUM 2012 A Strategy for Multilingual Europe, Brussels, Belgium, June 20, 2012. <http://www.meta-net.eu/events/meta-forum-2012/programme>.
56. Uszkoreit, Hans. "The META-NET Strategic Research Agenda: Overview, Preparation, Dissemination." Presentation at META-FORUM 2012 A Strategy for Multilingual Europe, Brussels, Belgium, June 20, 2012. <http://www.meta-net.eu/events/meta-forum-2012/programme>.
57. Uszkoreit, Hans, and Georg Rehm. "META-NET – Progress of the Initiative." Presentation at META-NET Network Meeting and General Assembly. Current State of Play and Next Steps, Berlin, Germany, October 21, 2011. <http://www.meta-net.eu/events/meta-net-ga-2011/talks/HansUszkoreit.pdf>.

58. Uszkoreit, Hans. “The Strategic Research Agenda for Multilingual Europe – Summary of Discussions and Plan for Completion.” Presentation at META-FORUM 2011 Solutions for Multilingual Europe, Budapest, Hungary, June 28, 2011. <http://www.meta-net.eu/events/meta-forum-2011/programme>.
59. Uszkoreit, Hans. “META, META-NET and META-FORUM.” Presentation at META-FORUM 2011 Solutions for Multilingual Europe, Budapest, Hungary, June 27, 2011. <http://www.meta-net.eu/events/meta-forum-2011/programme>.
60. Uszkoreit, Hans. “META-NET: Towards the Multilingual Europe Technology Alliance.” Invited Talk presented at the Meeting of the “LT Berlin” working group, Berlin, Germany, December 9, 2010.
61. Uszkoreit, Hans. “META-NET and META-FORUM: An Overview.” Presentation at META-FORUM 2010 Challenges for Multilingual Europe, Brussels, Belgium, November 17, 2010. http://www.meta-net.eu/events/meta-forum-2010/slides/META-FORUM2010_Uszkoreit_Welcome.pdf.
62. Uszkoreit, Hans. “Towards Visions for Language Technology in Multilingual Europe.” Presentation at META-FORUM 2010 Challenges for Multilingual Europe, Brussels, Belgium, November 17, 2010. http://www.meta-net.eu/events/meta-forum-2010/slides/meta-forum2010_uszkoreit_visiongroups.pdf.
63. Uszkoreit, Hans. “META-NET: Towards the Multilingual Europe Technology Alliance.” Invited Talk presented at Translingual Europe 2010, Berlin, Germany, June 7, 2010. <http://www.meta-net.eu/events/tle-2010>.
64. Uszkoreit, Hans. “META-NET: Towards the Multilingual Europe Technology Alliance.” Presentation at Translingual Europe 2010, Berlin, Germany, June 7, 2010. <http://www.meta-net.eu/events/tle-2010>.
65. Uszkoreit, Hans. “META-NET.” Sponsor Talk presented at the 7th International Conference on Language Resources and Evaluation (LREC 2010), Valletta, Malta, May 19, 2010.
66. Uszkoreit, Hans. “META-NET Building META-SHARE an Open Resource Exchange and Sharing Facility.” Invited Talk presented at the Language Technology Days, Luxembourg, Luxembourg, March 23, 2010. <http://cordis.europa.eu/fp7/ict/language-technologies/docs/s1-uszkoreit.pdf>.
67. Uszkoreit, Hans. “META-NET/T4ME.” Invited Talk presented at Language Resources for the Future - The Future of Language Resources, Barcelona, Spain, February 11, 2010.
68. Waibel, Alex. “Preliminary Findings of the Vision Group ‘Interactive Systems’.” Presentation at META-FORUM 2010 Challenges for Multilingual Europe, Brussels, Belgium, November 17, 2010. http://www.meta-net.eu/events/meta-forum-2010/slides/META-FORUM2010_Waibel.pdf.
69. Waibel, Alexander. “Socially Aware Interactive Assistants.” Presentation at Deutscher Sprachtechnologietag 2013. Sprachtechnologie heute und morgen: Perspektiven und Strategien, Berlin, Germany, January 24, 2013. <http://www.meta-net.eu/events/german-lt-day-2013/talks/alexwaibel.pdf>.
70. Weitzmann, John Hendrik, and Prodromos Tsiavos. “Language Resources and Legal Issues: Problems and Solutions for Basic and Industrial Research.” Presentation at META-FORUM 2010 Challenges for Multilingual Europe, Brussels Belgium, November 18, 2010. http://www.meta-net.eu/events/meta-forum-2010/slides/META-FORUM2010_Weitzmann-Tsiavos.pdf.
71. Yvon, François. “Optimisation et Évaluation des systèmes de traduction statistique.” Presentation at Journées DGA, Paris, France, July 5, 2011. <http://perso.limsi.fr/yvon/publications/sources/Yvon12algorithmic.pdf>.

4.4 Conferences, Workshops, Meetings

4.4.1 Conferences Organised by META-NET

1. theMETAnk, Berlin, Germany, June 4/5, 2010
2. Translingual Europe 2010, Hotel Maritim, Berlin, Germany, June 7, 2010
3. META-FORUM 2010, Brussels, Belgium, November 17/18, 2010
4. META-FORUM 2011, Budapest, Hungary, June 27/28, 2011
5. META-FORUM 2012, Brussels, Belgium, June 20/21, 2012

4.4.2 National Conferences and Roadshow Events

1. CESAR Roadshow Event, Sofia, Bulgaria, May 2, 2012
2. CESAR Roadshow Event, Bratislava, Slovakia, June 7/8, 2012

3. Multilinguality in Europe, Sofia, Bulgaria, September 26, 2012
4. LT4RD – Language Technologies for R&D Conference, Bucharest, Romania, 26-27 September, 2012
5. HLT Days 2012, Warsaw, Poland, 26-27 September, 2012
6. National workshop as a part of Conference of National Programme of Estonian Language Technology (NPELT), Tartu, Estonia, 2-3 October, 2012
7. HLT Baltic Conference, Tartu, Estonia, 4-5 October, 2012
8. Language Technology Conference, Ljubljana, Slovenia, 8-9 October, 2012
9. Workshop on LT and Innovation, Oslo, Norway, October 15, 2012
10. Parole Workshop, Lisbon, Portugal, 18-19 October, 2012
11. CESAR Roadshow Event, Belgrade, Serbia, 29 October, 2012
12. META-NORD/META-NET Sprogteknologisk, Copenhagen, Denmark, October 31, 2012
13. META-NORD and FIN-CLARIN workshop, Helsinki, Finland, November 2, 2012
14. European Languages in the Age of Technology: quo vadis?, Vilnius, Lithuania, 14 November, 2012
15. The Portuguese Language on the Digital Age Workshop, Lisbon, Portugal, 16 November, 2012
16. Workshop Estado de las lenguas oficiales habladas en España en la era digital Madrid, Spain, 20-23 November, 2012
17. Workshop Digital Survival of Maltese: building a roadmap, Valletta, Malta, 24 November, 2012
18. CESAR Roadshow Event, Zagreb, Croatia, November 30, 2012
19. CESAR Roadshow Event, Budapest, Hungary, December 3, 2012
20. German Language Technology Day, Berlin, Germany, 24 January, 2013

4.4.3 Meetings and Workshops Organised by META-NET

1. Workshop Language Resources: From Storyboard to Sustainability and LR Lifecycle Management at LREC 2010, Valleta, Malta, May 23 2010
2. Workshop on Legal issues for Language Resources at LREC 2010, Valleta, Malta, May 17 2010
3. Localization World - Panel discussion: Pricing in MT, Hotel Maritim, Berlin, Germany, June 8/9 2010
4. Vision Group Text Translation and Localisation 1st meeting, DFKI, Berlin, Germany, July 23 2010
5. Vision Group Media and Information Services 1st meeting, Paris, France, September 10 2010
6. Vision Group Interactive Systems 1st meeting, Paris, France, September 10, 2010
7. ICT 2010 - Networking Session Language Technology for a Multilingual Europe, Brussels, Belgium, September 26 2010
8. Vision Group Text Translation and Localisation 2nd meeting, Hotel Le Plaza, Brussels, Belgium, September 29 2010
9. Vision Group Interactive Systems 2nd meeting, Prague, Czech Republic, October 5 2010
10. Languages and the Media 2010 – Panel Presentation: Technological Support of AV Media Translation including Subtitling and Dubbing, Berlin, Germany, October 7, 2010
11. LISA Forum Europe, Budapest, Hungary, October 13 2010
12. Vision Group Media and Information Services 2nd meeting, Barcelona, Spain, October 15 2010
13. The International Workshop on Spoken Language Translation (IWSLT), Paris, France, December 2/3 2010
14. META-SHARE Meeting, Athens, Greece, December 14-16 2010
15. META-NET Training Event, Prague, Czech Republic, December 16-17 2010
16. Vision Group Interactive Systems 3rd meeting, Rotterdam, The Netherlands, March 28 2011
17. Vision Group Media and Information Services 3rd meeting, Vienna, Austria, April 1, 2011
18. Vision Group Translation and Localisation 3rd meeting, Prague, Czech Republic, April 7/8, 2011

Final Report

19. META-NET Workshop and Challenge at ICANN 2011: Context in Machine Translation, Espoo, Finland, June 14 2011
20. Media for All – Panel on Future Technology, London, UK, June 29 - July 1, 2011
21. Workshop Language Technology for a Multilingual Europe at GSCL 2011, Hamburg, Germany, September 27, 2011
22. Workshop on IPR and Metadata by META-NORD, Helsinki, Finland, September 30 2011
23. META-NET Network Meeting and General Assembly, Berlin, Germany, October 21/22 2011
24. IJCNLP 2011 - Workshop on Language Resources, Technologies and Standards in the Sharing Paradigm, Chiang Mai, Thailand, November 9-13 2011
25. ML4HMT-11 Workshop, Barcelona, Spain, November 19 2011
26. LTC 2011, Poznan, Poland, November, 25-27 2011
27. LREC 2012 Tutorial: Using the META-SHARE model implementation for describing and documenting language resources, Istanbul, Turkey, May 21 2012
28. LREC 2012 Workshop on IPR issues, Istanbul, Turkey, May 21, 2012
29. LREC 2012 Workshop Persistent identification for Language Resources, Istanbul, Turkey, May 21 2012
30. META-RESEARCH Workshop on Advanced Treebanking, Istanbul, Turkey, May 22 2012
31. LREC 2012 Workshop Exploring and Exploiting Official Publications, Istanbul, Turkey, May 27 2012
32. LREC 2012 Workshop Reinforcing International Collaboration on LRE, Istanbul, Turkey, May 27 2012
33. HLT Workshop, Tallinn, Estonia, June 5 2012
34. META-NET Workshop and Challenge at ICANN 2012: Context in Machine Translation, Lausanne, Switzerland, 10 September 2012
35. Second Workshop on Applying Machine Learning Techniques to Optimise the Division of Labour in Hybrid MT (ML4HMT-12), Mumbai, India, December 9 2012
36. META-NET Network Meeting and General Assembly, Berlin, Germany, January 25 2013

4.4.4 Conferences in which META-NET Played a Major Role

1. FLAReNet meeting Language Resources for the Future – The Future of Language Resources, Barcelona, Spain, February 11/12 2010
2. Language Technology Days 2010, Luxembourg, March 22/23 2010
3. LREC 2010, Valletta, Malta, May 19-21 2010
4. Forum TAL, Rome, Italy, October 28/29 2010
5. Language Technology for Multilingual Applications, Luxembourg, January 27 2011
6. Meeting of Representatives of European Language Councils, Copenhagen, Denmark, March 08 2011
7. TRALOGY, Paris, France, March 3/4 2011
8. FLAReNet 3rd Forum Language Resources in the Sharing Age – The Strategic Research Agenda, Venice, Italy, May 26-27, 2011
9. EUROLAN 2011 Summer School, Cluj-Napoca, Romania, August 28-September 4 2011
10. EACL 2012, Avignon, France, April 23-27 2012
11. LREC 2012, Istanbul, Turkey, May 21-27, 2012
12. ICT Fair, Lisbon, Portugal, 21-22 November, 2012
13. SIMC 2012 – International Symposium on Multilingualism in the Cyberspace, Paris, France, 21-23 November 2012
14. Translating and the Computer conference, London, UK, 29-30 November 2012
15. IWSDS'12, Paris, France, 29-30 November 2012
16. TRALOGY II, Paris, France, 17-18 January 2013

4.4.5 Invited Keynote Speeches by META-NET Members

1. HLT: The Baltic Perspective, Riga, Latvia, October 7/8 2010
2. EFNIL 2010, Thessaloniki, Greece, November 3 2010
3. Language question in research: English versus national languages?, Finnish Parliament, Helsinki, Finland, November 17 2010
4. CICling 2011, Tokyo, Japan, February 22 2011
5. Japanese Workshop for Machine Translation, Tokyo, Japan, February 23 2011
6. Attensity Forum 2011, Berlin, Germany, May 6 2011
7. RANLP 2011, Hissar, Bulgaria, September 12-14 2011
8. GSCL 2011: Multilingual Resources and Multilingual Applications, Hamburg, Germany, Sept. 28-30 2011
9. NPLD Assembly, Eskilstuna, Sweden, October 26 2011
10. EFNIL 2011, London, UK, October 26 2011
11. CHAT 2012 Workshop, Madrid, Spain, June 22 2012
12. EFNIL Conference 2012, Budapest, Hungary, October 24-26 2012