Cracking the Language Barrier for a Multilingual Europe

Georg Rehm
META-NET Network Manager – CRACKER Coordinator
DFKI, Germany
georg.rehm@dfki.de

EFNIL Conference 2015
Helsinki, Finland, October 08, 2015

META-NET has received funding from the EU’s Horizon 2020 research and innovation programme through the contract CRACKER (grant agreement no.: 645357). Formerly co-funded by FP7 and ICT PSP through the contracts T4ME (grant agreement no.: 249119), CESAR (grant agreement no.: 271022), METANET4U (grant agreement no.: 270893) and META-NORD (grant agreement no.: 270899).
Outline

1. META-NET – recap and overview
2. CRACKER
3. META-FORUM 2015 and Riga Summit 2015
4. Cracking the Language Barrier
5. Connecting Europe Facility – Automated Translation (CEF AT)

http://www.meta-net.eu
META-NET and META

- **META-NET**
  - 60 research centres in 34 countries (via four EU-funded projects: T4ME, CESAR, METANET4U, META-NORD).
  - New chair of Exec. Board: Jan Hajic.

- **META**
  - Multilingual Europe Technology Alliance, 800+ members in 70 countries
  - [http://www.meta-net.eu/members](http://www.meta-net.eu/members)
Pan-European infrastructure, bringing together providers and consumers of language data, tools and services.

LRs are documented, uploaded, stored, catalogued, downloaded, shared – to improve visibility, documentation, identification, availability, interoperability.

META-SHARE includes repository software, a metadata model, licensing kit, statistics.

29 distributed repositories maintained by 37 organisations in 25 countries.

2,500+ resources (corpora: 49%, lexical: 38%, tools/services: 12%), covering ca. 100 languages.

7,000+ downloads in total; ca. 70% of all LRs have been downloaded.

http://www.meta-net.eu
Welcome to META-SHARE!

META-SHARE is developed within the META-NET Network of Excellence

About the project

META-NET is designing and implementing META-SHARE, a sustainable network of repositories of language data, tools and related web services documented with high-quality metadata, supported in centres. It will provide a uniform search and access service to shared resources. Data and tools can be both 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.

META-SHARE will start by integrating nodes and centres represented by the partners of the META-NET consortium. Initial partners are:

- CNR - National Research Council of Italy
- DFKI – Deutsches Forschungszentrum für Künstliche Intelligenz
- ELIA – Evaluations and Language Resources Distribution Agency
- EUN – European University Association
- FCL – Fédération des Centres Linguistiques
- INRIA – Institute National de Recherche en Informatique et en Automatique
- KUL - Katholieke Universiteit Leuven
- LISA – Linguistic and Informational Service for American Heritage
- MPI – Max Planck Institute for Psychological and Social Research
- NAU – National Academy of Sciences Ukraine
- SMNL – Swedish National Language Lab
- UCL – University College London
- VUB – Vrije Universiteit Brussel
- VU – Vrije Universiteit Amsterdam
- WLU – Wilfrid Laurier University

META-SHARE Managing Nodes

Select network node

Please select one of the following META-SHARE network nodes to proceed. For an explanation of the differences between META-SHARE Managing Nodes and other META-SHARE nodes, see this page.

META-NET

About the partners

META-SHARE will start by integrating nodes and centres represented by the partners of the META-NET consortium. It will gradually be extended to encompass additional nodes/centres and provide more functionality with the goal of turning it into an as large distributed infrastructure as possible.

META-SHARE Managing Nodes

Other META-SHARE Nodes

The META-SHARE Infrastructure

META-SHARE aims to provide a multi-layer infrastructure that:

- makes available quality LRs and related metadata over the META-SHARE network;
- ensures that such LRs and metadata are properly preserved and maintained;
- makes a set of services available to all META-SHARE members and users;
- promotes the use of widely acceptable standards for language resource building which ensures the maximum possible interoperability of LRs;
- allows associated third parties to easily use their LRs over the META-SHARE network;
- allows any potential user of the LRs to easily and legally acquire the requested high-quality LRs.

Connect and Share! – Our vision

META-SHARE Members set up and maintain their own Language Resource Repositories, the META-SHARE Network Nodes, acting as Repository Service Providers. Thus, we can do so by using the open-source META-SHARE software which is freely available at GIPSY.

META-SHARE Members who do not wish to set up their own language resources repository can deposit their language resources in the META-SHARE repositories maintained by other META-SHARE Network Members who undertake to host and/or be responsible for those LRs (relying repositories).

META-SHARE Managing Nodes are META-SHARE Network Members that provide core services to all META-SHARE Network Members, acting as Core Service and/or Support Service providers.

Third Parties are able to use the search services and easily acquire LRs from the META-SHARE network in accordance with the relevant licensing terms and conditions.

By submitting your PDWP Framework Programme and/or EC funded projects of the European Commission through the Contracts Frameworks (FP5/6/7) and/or the Communities’ Support Frameworks (ESPA I/II, ESPA III/IV, ESPA V), you are agreeing to the “Terms of Service” which are available at META-SHARE.
Summary: “At Least 21 European Languages in Danger of Digital Extinction!”
CRACKER (CSA, 2015–2017)

CRACKER – Cracking the Language Barrier: Coordination, Evaluation and Resources for European MT Research

<table>
<thead>
<tr>
<th>No.</th>
<th>Institution</th>
<th>Country</th>
<th>Coordinator</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>DFKI</td>
<td>Germany</td>
<td>Georg Rehm (Coordinator)</td>
</tr>
<tr>
<td>2</td>
<td>CUNI</td>
<td>Czech Republic</td>
<td>Jan Hajic</td>
</tr>
<tr>
<td>3</td>
<td>ELDA</td>
<td>France</td>
<td>Khalid Choukri</td>
</tr>
<tr>
<td>4</td>
<td>FBK</td>
<td>Italy</td>
<td>Marcello Federico</td>
</tr>
<tr>
<td>5</td>
<td>ATHENA RC</td>
<td>Greece</td>
<td>Stelios Piperidis</td>
</tr>
<tr>
<td>6</td>
<td>UEDIN</td>
<td>UK</td>
<td>Philipp Koehn</td>
</tr>
<tr>
<td>7</td>
<td>USFD</td>
<td>UK</td>
<td>Lucia Specia</td>
</tr>
</tbody>
</table>
Communities

- META-NET incl. META-SHARE and META
- MT evaluation initiatives – WMT, IWSLT, MT Marathons
- MT and other LT industry
- Language resources – META-SHARE, ELRA
- HT/MT evaluation tool – translate5
- Translation industry and translation profession
- MT user communities
Selected Activities

- **Production of resources** (e.g., for WMT 2016 and 2017, IWSLT 2015-2017)
- **Tools** (quality control, evaluations; towards the idea of a smart workbench for translators)
- **Strategies and roadmaps** (SRIA, Roadmap for European MT Research)
- **Exchange and sharing** facility for resources (META-SHARE)

**META-SHARE**

**Maintenance of Operations and Outreach**

- Provide services, adapt them to evolving user requirements and licensing landscape
  - adapt, streamline and extend the metadata schema;
  - adapt licensing toolkit to new international licensing setups;
  - streamline and simplify operations for repository providers and data depositors.
- Technical support and bug fixing

http://www.cracker-project.eu • http://www.meta-net.eu
Funded actions in the call H2020-ICT-17 “Cracking the Language Barrier”:

- One Research and Innovation Action
  - QT21 (Quality Translation 21) – DFKI

- Three Innovation Actions
  - HimL (Health in my Language) – UEDIN
  - MMT (Modern Machine Translation) – Translated
  - TraMOOC (Transl. of MOOCs) – HUB

- Two Coordination and Support Actions
  - CRACKER – DFKI
  - LT_Observatory – Zabala
META-FORUM 2010-2015

- META-FORUM 2010 – Nov. 17/18, Brussels, Belgium
  Challenges for Multilingual Europe
- META-FORUM 2011 – June 27/28, Budapest, Hungary
  Solutions for Multilingual Europe
- META-FORUM 2012 – June 20/21, Brussels, Belgium
  A Strategy for Multilingual Europe
- META-FORUM 2013 – Sept. 19/20, Berlin, Germany
  Connecting Europe for New Horizons
- META-FORUM 2015 – April 27, Riga, Latvia
  Technologies for the Multilingual Digital Single Market

http://www.meta-net.eu
META-FORUM 2015

- META-FORUM 2015
- April 27 in Riga, Latvia
- Riga Summit 2015 on the Multi-lingual Digital Single Market
- Highly compressed programme (one day instead of two)
- Two important components:
  - Emerging SRIA
  - Further community fusing
- http://www.meta-forum.eu
Communities & Stakeholders

META FORUM 2015
Technologies for the Multilingual Digital Single Market

... and many more research centres, companies, EU projects etc.
Digital Single Market

- The DSM is one of the most important priorities of the EU.
- Expected to add 340b€ to European GDP, 100,000s of new jobs.
  1. Better access for consumers and businesses to digital goods and services (harmonised consumer rules, geo-blocking, copyright etc.)
  2. Shaping the environment for digital networks and services to flourish
  3. Creating a European digital economy and society with long-term growth potential (smart industrial system, standards, cloud computing etc.)
- Language topic not included in early drafts.
- EC presented the DSM strategy on May 05, 2015.
- Unfortunately, language was a complete and utter no show.
DSM must be multilingual!

- Europe’s DSM must be multilingual to be integrated and unified.
- Otherwise it will consist of 20+ isolated markets that don’t have any bridges between them – language barriers are market barriers!
- Multilingualism is at the very heart of the European idea.
- 24 official and many more regional, minority, immigrant languages as well as those of important trade partners.
- DSM cannot be hampered by language barriers!

http://www.meta-net.eu
THREE PRIORITY AREAS FOR ACHIEVING THE MULTILINGUAL DIGITAL SINGLE MARKET

Multilingual access to all digital goods and services across Europe

Geo-blocking: due to nationality, location, or residence

Language-blocking: languages they do not speak

Geo-blocking and language-blocking are barriers to access

Both geo-blocking and language-blocking are daily problems for tens of millions of EU citizens.

Customers are six times more likely to buy from sites in their native language.

English is not the answer

52% of EU customers do not purchase from English-language sites.

Adding even a few languages to an SME’s website beyond English can have a major impact on revenue. Large organizations today often localize products and websites into fifty or more languages to increase market share.

Most EU languages address less than 3% of the market, fundamentally limiting SMEs operating in countries where those languages are spoken.

http://www.meta-net.eu
Facts and Figures

Geo-blocking and language-blocking are barriers to access

**Geo-blocking:**
- keeps customers from accessing content due to nationality, location, or residence
- can be worked around by tech-savvy customers
- prevents some cross-border commerce

**Language-blocking:**
- keeps customers from accessing content in languages they do not speak
- customers never even know what they cannot find
- is unavoidable: no-one speaks all languages; however, current online translation is insufficient
- prevents customers from even trying to conduct cross-border commerce
- disproportionately impacts speakers of less common languages

Both geo-blocking and language-blocking are daily problems for tens of millions of EU citizens.
Europe today is not a single market: it is a separated into 28+ small language markets.

Current eCommerce growth within Europe is about (565 million) opportunity languages present.

Online Population (510 million)

Chinese

Europe today is not a single market: for (165 million) of European SMEs currently sell cross-language.

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

Japanese

A

AG

J
A Strategy for the MDSM!

- Strategic R&I Agenda for the Multilingual Digital Single Market
- Core: Technology Solutions
- Data economy is an inherent component – LT for effective multilingual data value chains.

http://www.meta-net.eu
Strategic Agenda for MDSM

- Presented at META-FORUM 2015 and Riga Summit for the first time.
- Version 0.5 – work in progress
- Builds upon many strategy papers and roadmaps prepared by several European projects, incl. the META-NET SRA (2013).
- Input and feedback collected at the Riga Summit 2015 to be used for upcoming versions.

http://www.meta-net.eu
Joint EFNIL and NPLD Panel

- Joint EFNIL and NPLD panel at META-FORUM 2015.
- A joint position paper was presented.

"Languages are not only a means of communication. They also have embedded in them people’s values, aspirations and hopes." (European Roadmap for Linguistic Diversity 2015, NPLD)

"Many European languages run the risk of becoming victims of the digital age as they are under-represented and under-resourced online. Huge regional market opportunities remain untapped because of language barriers." (Multilingual Europe: A Challenge for Language tech. MultiLingual April/May 2011, page 51/52)


http://www.meta-net.eu
On Friday, March 20, 2015, we published an open letter to the EC on http://multilingualeurope.eu.

On Monday, March 23, 2015, we informed President Juncker and all Commissioners about the campaign and the 1300+ signatures.

By now more than 3500 signatures!

Who signed?

- 5 Members of the European Parliament
- 150+ high-level representatives from industry (CxO level)
- 1200+ professors
- 400+ project or research managers
- 20+ entrepreneurs and founders
- hundreds of language and language technology professionals, officials, researchers, administrators and representatives from related stakeholder groups
A federation of European projects and organisations working on technologies for a multilingual Europe.

Multi-lateral Memorandum of Understanding; 8 organisations and 19 projects on board already (including FP7 and H2020-ICT15).

Selected areas of collaboration: data management and repositories (incl. Data Management Plan), tools, shared tasks and evaluations.

Goal: provide one platform and umbrella organisation for the whole community.
Website: information about all projects and organisations.

Downloadable documents.

List of events.

Upcoming: news, blog etc.

Several new members will join the initiative soon.

http://www.cracking-the-language-barrier.eu
• Connecting Europe Facility (CEF Digital) is a set of pan-European Digital Service Infrastructures (e.g., cross-border e-justice, e-procurement, e-health, Europeana, Open Data Portal)

• CEF AT to provide multilinguality for these Digital Service Infrastructures via the MT@EC system

• CEF AT goals:
  – To make public services usable by all EU users, irrespective of their working language and language skills
  – To facilitate cross-border information exchange in public administration

• Free access to machine translation tool (MT@EC):
  – Apply for ECAS account: https://webgate.ec.europa.eu/cas/eim/external/register.cgi
  – Send an email to DGT-MT@ec.europa.eu asking for activation of MT@EC
• Identify language resources relevant to CEF AT in national public service administrations across Europe
• Provide language resources relevant to CEF AT to the EC
• Provide best services back by adapting CEF AT according to the needs of national public service administrations
• Optimize Europe’s Digital Service Infrastructures according to the needs of national public service administrations
• Contribute to the success of the Multilingual Digital Single Market (MDSM)
• ELRC is organising workshops in all countries participating in CEF.

• Target audience: representatives of public administrations and public services (data holders).

• All delegates of the EFNIL 2015 conference who want to know more about ELRC and CEF AT are invited to contact my colleague Andrea Lösch <andrea.loesch@dfki.de>.

• More information: http://www.lr-coordination.eu
Summary and Conclusions

- Europe is only starting to recognise the potential/importance of LT.
- Language Technology will be a key ingredient of our future IT.
- EU has a unique opportunity for strategic investment into our future.
- EU needs to decide if it wants to contribute – or if it can afford not to.
- Multilingual public services play an increasingly important role (CEF).
- Strong support of open letter a big surprise – maybe an opportunity?
- We must increase our collaboration to be able to make a difference.

Next steps:
  (a) updated (and simplified) version of the MDSM SRIA;
  (b) extension of the Cracking the Language Barrier initiative;
  (c) maybe try to get the support of the European citizen?
Thank you!

Available at the registration desk

http://www.meta-net.eu
http://www.meta-forum.eu
http://multilingualeurope.eu
http://rigasummit2015.eu
http://www.cracker-project.eu
http://www.cracking-the-language-barrier.eu
http://www.lr-coordination.eu