European Language Resource Coordination

ELRC Vilnius Workshop | 24.02.2016

Andrejs Vasiljevs | ELRC/Tilde
Twitter map on the Language borders in Europe
Multilingual Europe

1.1B EUR

EU institutions spend per annum on translation services
The Second Gutenberg Effect

Languages that will not be used in the modern technologies are going to face digital extinction.
Will smaller languages survive in the 21st century? Only if the speed of language technology development is faster than the rate of their diminishment.

— Alvin Toffler —
Language Whitepaper Series:
30 volumes published by Springer

http://www.meta-net.eu/whitepapers/overview
## Machine Translation support for European languages

<table>
<thead>
<tr>
<th>Excellent support</th>
<th>Good support</th>
<th>Moderate support</th>
<th>Fragmentary support</th>
<th>Weak/no support</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>French</td>
<td>Catalan</td>
<td>Basque</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spanish</td>
<td>Dutch</td>
<td>Bulgarian</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>German</td>
<td>Croatian</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hungarian</td>
<td>Czech</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Italian</td>
<td>Danish</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Polish</td>
<td>Estonian</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Romanian</td>
<td>Finnish</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Galician</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Greek</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Icelandic</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Irish</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Latvian</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Lithuanian</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Maltese</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Norwegian (Bokmål, Nynorsk)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Portuguese</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Serbian</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Slovak</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Slovene</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Swedish</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Welsh</td>
<td></td>
</tr>
</tbody>
</table>

Workshop Context

- Automated Translation at EC for many years ...
- Machine Translation at EC: MT@EC

- New **CEF Automated Translation** is special
  - Focus on public services, administrations, ministries, Pan-European digital services ...
  - CEF.AT is free for public administrations

- Your data is needed to make CEF Automated Translation better for you and for everybody
European Language Resource Coordination (ELRC) Objective

To establish a permanent **Language Resource Coordination** mechanism that will feed the CEF Automated Translation DSI with relevant **Language Resources** in **all official languages** of the EU and CEF Associated countries in order to **improve the quality**, coverage and performance of automated translation systems and solutions in the context of current and future **CEF digital services**.
European Language Resource Coordination (ELRC) is a tender operating under the EU’s Connecting Europe Facility SMART 2014/1074 programme.
Tilde @ ELRC Consortium

• A leading European developer of language technologies and provider of localization services
• 120 employees in Riga, Tallinn and Vilnius offices
• Strong R&D team (70+ scientific publications)
• Cooperation with leading European universities
• Key player in the major European industry and research associations META-NET, ELRA, LT-Innovate, BDVA
• Coordinator and participant in large scale EU H2020 and FP projects
ELRC Main Activities

- LR Coordination **Workshops** in all CEF countries
- Annual LR Coordination **conferences**
- Setting up a network of National Anchor Points and Language Resource Board
- Promote **best practice** cases:
  - e.g. language technology application for public services in the Baltic States
- Technical **Helpdesk**
- Advisory and **consultancy** service in the area of LRs for MT
- **Resource** identification and collection
The CEF conference "Towards a Connected Multilingual Europe" will be held on Day Three of the Riga Summit. Pan-European digital service infrastructures (DSIs) foreseen in the CEF programme will deliver online public services for EU Member State administrations, citizens and economic operators. Automated Translation (AT) is one of the DSIs, and will provide automated translation services to all other CEF DSIs that necessitate multilingual communication or exchange of documents and other linguistic content between administrations, citizens and economic operators.

http://rigasummit2015.eu
ELRC Portal: www.lr-coordination.eu
Workshop Objectives

• Raise awareness
  – on the value and importance of data held by public services in overcoming language barriers

• Engage you and your institutions
  – in the identification and sharing of data for CEF.AT

• Help with legal and technical issues
  – associated with the collection and/or provision of data by public service administrations

• Share data for the best results for your needs and your language
  – with European Commission and FIN-CLARIN
  – Open data to strengthen Finnish language in the Digital age

• Promote CEF.AT services that can be useful for you
Languages — the heart of Multilingual Europe

Supporting Lithuanian language is supporting Europe,
Supporting Europe is supporting Lithuanian language