Why Do We Need Language Technology?

Lori.Thicke@Lexcelera.com
About Lexcelera

- Founded in Paris in 1986
- First LSP in France ISO 9001 quality certified
- Offices in Europe, North & South America
- Known for Quality MT
- Leading expert in RBMT and SMT
- Major MT deployments since 2007
- MT engine agnostic
- Awarded EU grant for SMT (Project ACCEPT)
• Founded in Paris in **1993**
• French **charity**
• As of 2010 also a **US 501 (c)(3) charity**
• Over **1200** translators around the world
• Automated Translation **Workspace** powered by ProZ.com
• **3,978,450** words donated since May 2011
• Nearly **$3 million** donated since 1993
• **70+** non-profits supported
• **73** language pairs
Why do we need language technology?

Shortage of Translators

Social Web

Fragmented market

Growing Volumes

Growing Global Reach

Lack of Visibility

New Needs

Shrinking Deadlines
Why do we need language technology?
Translation is important because information is important

- 97.14% believed that greater access to translation would help individuals in Africa understand their legal rights.
- 95.85% believed that greater access to translation would help protect human rights in Africa.
- 94.92% believed that greater access to translation would have a positive impact on the collective health of people living in Africa.
- 91.96% believed that greater access to translation would help people in Africa contribute to the political process.
- 88.78% believed that greater access to translation would help prevent international, civil, and ethnic conflict in Africa.

63.07% of African translators said greater access to translated information could have prevented the loss of life of people in their family or circle of friends.

Download the full report, “The Need for Translation in Africa”
In Europe, information access depends on what language you speak.

- English
- French
- Italian
- Spanish
- Swedish
- Czech
- Romanian
- Lithuanian
- Slovak
- Estonian
- Maltese
- Estonian
- Latvian
- Lithuanian
- Slovak
- Romanian
- Czech
- Swedish
- Spanish
- Italian
- French
- English
## Language discrimination around the globe

<table>
<thead>
<tr>
<th>Language</th>
<th>Number of Native Speakers</th>
<th>Number of Wikipedia articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>330,000,000</td>
<td>3,888,898</td>
</tr>
<tr>
<td>Bengali</td>
<td>180,000,000</td>
<td>23,087</td>
</tr>
<tr>
<td>Hausa</td>
<td>25,000,000</td>
<td>270</td>
</tr>
<tr>
<td>Somali</td>
<td>16,000,000</td>
<td>2,324</td>
</tr>
<tr>
<td>Nepali</td>
<td>13,900,000</td>
<td>19,215</td>
</tr>
<tr>
<td>Shona</td>
<td>10,800,000</td>
<td>873</td>
</tr>
<tr>
<td>Zulu</td>
<td>10,300,000</td>
<td>305</td>
</tr>
<tr>
<td>Bulgarian</td>
<td>9,100,000</td>
<td><strong>128,045</strong></td>
</tr>
<tr>
<td>Kikuyu</td>
<td>7,200,000</td>
<td>122</td>
</tr>
<tr>
<td>Lithuanian</td>
<td>3,200,000</td>
<td><strong>148,682</strong></td>
</tr>
<tr>
<td>Latvian</td>
<td>1,500,000</td>
<td>40,237</td>
</tr>
<tr>
<td>Swahili</td>
<td>800,000</td>
<td>23,263</td>
</tr>
<tr>
<td>Maltese</td>
<td>400,000</td>
<td>2,980</td>
</tr>
</tbody>
</table>
Lexcelera's Machine Translation Services

- Quality-driven
- Automated
- Engine agnostic
- Process-based

Pilots
Customization
Machine Translation
Post-editing
“Cost: Savings of 30%.

“Turnaround: Time to market reduced by one third.

“Quality: Contrary to all expectations, using MT has improved the translation quality.

Online Help (French): 8.25/10 for MT, while the traditional TM translation was 8/10.

Courseware (German): Score 8/10, with the reviewer commenting ‘It was nearly 9, … it was the best translation of courseware I ever read.’”
MT is not a Tool. MT is a Process

**PRE-PRODUCTION**
1. Authoring
2. Customization
3. Piloting

**PRODUCTION**
4. Processing
5. Post-Editing

**POST-PRODUCTION**
6. QA/Metrics
7. Maintenance
Quality at the Source for the Social Web

Goals

- Develop a user-friendly strategy for (automatically) pre-editing User- Generated Content for Statistical Machine Translation (WP2)

- Address the post-editing bottleneck by creating engaging online tools and leveraging monolingual subject-matter experts as post-editors (WP5)

- Improve learning and develop feedback loops to improve SMT results for community data (WP4)

- Improve reliability of forum content translation using Text Analytics (WP3)
ACCEPT

✔ Join our Special Interest Group
✔ Check your user-generated content
✔ Use our SDK to build your own plug-ins
✔ Integrate with your systems
✔ Leverage our evaluation framework
New Language Technology: Made in Europe

- Post-Editing Environment
- Terminology Control for SMT
- TM Integration
Thank you!

Contact
Lori.Thicke@Lexcelera.com