Repurposing language resources for multilingual websites

Fernando Serván
Food and Agriculture Organization of the United Nations (FAO),
Rome, Italy
Context

- International organization with >190 member countries
- Global mandate on agriculture, food standards, animal diseases, genetic resources, forestry, fisheries.
- 5 regional offices, 13 sub-regional offices, 67 country offices.
- Languages: English, French, Spanish, Arabic, Chinese, Russian.
## Our users

<table>
<thead>
<tr>
<th>ranking</th>
<th>English name of Language</th>
<th>language code</th>
<th>percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>English</td>
<td>en</td>
<td>53.2%</td>
</tr>
<tr>
<td>2</td>
<td>Spanish</td>
<td>es</td>
<td>15.2%</td>
</tr>
<tr>
<td>3</td>
<td>French</td>
<td>fr</td>
<td>13.0%</td>
</tr>
<tr>
<td>4</td>
<td>German</td>
<td>de</td>
<td>3.8%</td>
</tr>
<tr>
<td>5</td>
<td>Italian</td>
<td>it</td>
<td>2.6%</td>
</tr>
<tr>
<td>6</td>
<td>Russian</td>
<td>ru</td>
<td>1.6%</td>
</tr>
<tr>
<td>7</td>
<td>Chinese</td>
<td>zh</td>
<td>1.4%</td>
</tr>
<tr>
<td>8</td>
<td>Dutch</td>
<td>nl</td>
<td>1.2%</td>
</tr>
<tr>
<td>9</td>
<td>Japanese</td>
<td>ja</td>
<td>1.2%</td>
</tr>
<tr>
<td>10</td>
<td>Portuguese</td>
<td>pt</td>
<td>1.2%</td>
</tr>
<tr>
<td>11</td>
<td>Turkish</td>
<td>tr</td>
<td>0.8%</td>
</tr>
<tr>
<td>12</td>
<td>Arabic</td>
<td>ar</td>
<td>0.5%</td>
</tr>
</tbody>
</table>
Our content

- Governing bodies’ documents
- Normative documentation (standards)
- News and campaigns
- Technical information
- Internal communication

Need to make it available in all languages
Our budget -- <$
Our content

- Governing bodies’ documents → HT
- Normative documentation (standards) → HT
- News and campaigns → HT.... MT?
- Technical information → HT.... MT?
- Internal communication → HT ....MT?
Testing MT – Our project

Re-use of *legacy* translations for non-translated content

Integration **TM – MT** (doc – web ?)

Improve accessibility to the website in other languages

A tool for HT processes?
Testing MT – Our project

SMT engine
EN/ES
EN/AR

Search engine

query in AR/ES
result in AR/ES

query in EN
result in EN

LANGUAGE MODEL
and PHRASE TABLE

TBX

Terminological db

TMX

Translation memories

XML

Text bases (mono)
Testing MT – Our project result

The Multilingual Web – The Way Ahead
Luxembourg, 15 March 2012
Testing MT – Our project result

The Multilingual Web – The Way Ahead

Luxembourg, 15 March 2012
**Needs/opportunities (1)**

**Documentation**: technical documentation is not summarized and it is complex (except *moses-for-mere-mortals*)

**Best practices**: a network of practitioners exist in the UN system, however, no documented best practices are available.
**Needs/opportunities (2)**

**Standards:** each part of the process and language resource has its own standard (TBX, TMX, XLIFF) but...

**Interoperability:** it is difficult to make them work together

translate=no ------ attributes

**MultilinguaWeb-LT initiative provides a good opportunity**
¡Gracias!

Fernando Serván, Food and Agriculture Organization of the United Nations (FAO), Rome, Italy

Special thanks to:

Giuseppe Forte and Nick Waltham, Food and Agriculture Organization of the United Nations (FAO), Rome, Italy