Beyond traditional computer-assisted translation tools

A localisation perspective

Lucía Morado Vázquez and Silvia Rodríguez Vázquez
http://diarium.usal.es/codex

Equivalences 2014 « Traduction et technique »
24 October 2014 – Kursaal, Bern

Contents

• Introduction
  – Cod.eX
  – Computer-assisted translation tools
  – Localisation

• A localisation perspective
  – Other tools
    • Localisation specific tools
    • Domain specific tools
  – The role and importance of standards of localisation
  – Translation technologies training

• Summary and conclusion
Who are we?

Research group within TIM (Département de Traitement Informatique Multilingue)

and

Universidad de Salamanca

Cod.eX

Specialised in:

- Web accessibility and localisation
- Multilingual web management
- Standards of localisation
- Translation technologies training
- Localisation training
- Post-editing (MT)
Computer-assisted translation (CAT) tools

CAT tools automate certain facets of human translation processing in order to enhance overall translator productivity (adapted from Folaron 2010)

CAT tools are used to support the translator, by eliminating repetitive work, automating terminology lookup activities, and recycling previously translated texts. (Esselink 2000, p.359)

Related* concepts:
– TMS: Translation Memory Systems
– TENT: Translation Environment Tools

2010
HOW MANY CAT/TF TOOL DO YOU USE ROUTINELY FOR PREPARATION AND TRANSLATION AND POST-TRANSLATION CHECKING?

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5 *</th>
</tr>
</thead>
<tbody>
<tr>
<td>163 (25%)</td>
<td>115 (18%)</td>
<td>65 (10%)</td>
<td>15 (2%)</td>
<td>2 (0.3%)</td>
</tr>
</tbody>
</table>

Respondents: 372

2013
HOW MANY CAT/TF TOOL DO YOU USE ROUTINELY FOR PREPARATION AND TRANSLATION AND POST-TRANSLATION CHECKING?

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5 *</th>
</tr>
</thead>
<tbody>
<tr>
<td>163 (18%)</td>
<td>115 (13%)</td>
<td>65 (7%)</td>
<td>15 (2%)</td>
<td>2 (0.2%)</td>
</tr>
</tbody>
</table>

Respondents: 708

WHICH TRANSLATION ENVIRONMENT TOOL(S) DO YOU USE WITH SOME FREQUENCY?

<table>
<thead>
<tr>
<th>2010</th>
<th>2013</th>
</tr>
</thead>
<tbody>
<tr>
<td>OmniTI</td>
<td>SDL Trados 2007</td>
</tr>
<tr>
<td>DejaVu</td>
<td>SDL Trados Studio</td>
</tr>
<tr>
<td>SDLX</td>
<td>WordFast Classic</td>
</tr>
<tr>
<td>SDLX</td>
<td>WordFast Pro</td>
</tr>
<tr>
<td>Star Transit</td>
<td>memoQ</td>
</tr>
<tr>
<td>Across</td>
<td>SDLX</td>
</tr>
<tr>
<td>Other</td>
<td>Star Transit</td>
</tr>
<tr>
<td>Absolutely none and proud of it</td>
<td>Absolutely none and proud of it</td>
</tr>
</tbody>
</table>

Respondents: 458

2013 Translation Environment Tools Survey, Kevin Lossner
CAT Tools

Suited for:

- Technical / repetitive texts (Bowker 2005)
- New versions/ updates
- Big projects:
  - Style and terminology coherence
  - Shared translation memory (several translators)
- Similar texts of the same client
CAT Tools

Advantages

– Easier exchange of linguistic resources
– Reuse of translated texts
– Time (Brkc et al 2009, Morado 2012)
– Style coherence (Bowker 2005, p.15)
– Concordance search
– Analysis of internal text repetition
– Project statistics

Disadvantages

– Investment:
  • Money (licences)
  • Learning (de Saint Robert 2008; García 2008)
– Context reduction (segment focus)
– Risk of error propagation (de Saint Robert 2008)
– Translation memory maintenance
– File management skills
– Complex invoicing system
Localisation (L10n): “Adapting a digital product to a given language or culture, so that it seems natural to that region”

(GALA, 2011)

— If you are not an English native speaker, localisation is all over the place
— Most of the software that you are using has already been localised
**Localisation**

**International**

WELCOME TO THE UNIVERSITY OF GENEVA

The University of Geneva is an institute dedicated to research, teaching and dialogue. It is the site of academic creativity and the transmission of knowledge. Its roots in foundations in 1559 by the Swiss Confederation have developed over the centuries, and the University is the University of the 21st century and is renowned for its internationality in teaching and research. It is home to over 40,000 students and researchers from all over the world.

INTERNATIONAL

UNIVERSITÉ DE GENEVE

Programme des études doctorales

BENNEVUE À L'UNIVERSITÉ DE GENEVE

L'Université de Genève condamne l'instabilité de la recherche, l'enseignement et la culture. Elle a pour but d'offrir un cadre de travail propice à la créativité scientifique et à la transmission de connaissances.

Depuis sa fondation en 1559, l'UNIGE a favorisé un renforcement des compétences et de la qualité des formations universitaires. Ce qu'elle contribue à la place pour les établir sur le marché international.

L'UNIGE est représentée par le réseau international de Genève, pôle inter- et multiculturel, avec une large gamme d'activités sur le monde. La volonté de développer toujours plus sa vocation de centre de la recherche en sciences humaines et sociales, avec la contribution du centre de recherche interdisciplinaire, la vocation de devenir un "centre mondial", un lieu de convergence des savoirs, un lieu de rencontre des sciences humaines et sociales.

INTERNATIONAL

UNIVERSITÉ DE GENEVE

Programme des études doctorales

Inscription et informations pratiques.

Programme: Acquis pour les études doctorales, inscription au Département de l'Université.

Inscription: Les inscriptions sont ouvertes à partir du 1er septembre.

modalités: Les informations pratiques et les modalités de inscription sont disponibles sur le site Internet de l'Université.

Nous nous réjouissons de vous voir sur le site internet de l'Université. Le programme des études doctorales est réceptif aux appels de candidature et fait partie de l'inscription des étudiants à l'Université.
A localisation perspective: Localisation and CAT tools
Localisation and CAT tools

— Localisation-specific tools (SDL Passolo, Alchemy Catalyst, Lingobit localiser...)

— Many different products and file formats to deal with:

<table>
<thead>
<tr>
<th>File Formats</th>
<th>Tools</th>
</tr>
</thead>
<tbody>
<tr>
<td>png, xcf, svg, gif, html, css, exe, sb, json, sln, xlsx, po, jpg, docx, xlf, blk, jsn, scm, apk, zip, xml, jpg, png, gif, jpg, ....</td>
<td>CAT tools, Image editors, Advanced text editors, OCR (Optical Character Recognition) tools</td>
</tr>
</tbody>
</table>

Translators/localisers should also work with domain-specific tools: image editors, programming environments...
### Localisation and CAT tools

#### Image localisation

<table>
<thead>
<tr>
<th>File Formats</th>
<th>Tools</th>
</tr>
</thead>
</table>
| png gif xcf jpg svg | CAT tools
| Image editors Advanced text editors OCR (Optical Character Recognition) tools |

---

![Image of tools and software interfaces for localization and CAT tools.](image-url)
### Localisation and CAT tools

#### Web localisation

<table>
<thead>
<tr>
<th>File Formats</th>
<th>Tools</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mainly html (static webpages) and css.</td>
<td>CAT tools</td>
</tr>
<tr>
<td></td>
<td>Advanced text editors</td>
</tr>
<tr>
<td></td>
<td>Web developing tools</td>
</tr>
</tbody>
</table>

---

**Diagram:**

- **Localisation and CAT tools**
- **Web localisation**
- **File Formats**
  - Mainly html (static webpages) and css.
- **Tools**
  - CAT tools
  - Advanced text editors
  - Web developing tools
Localisation and CAT tools

Software localisation

File Formats

<table>
<thead>
<tr>
<th>exe</th>
<th>po</th>
<th>cs</th>
<th>xml</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Tools

- Localisation tools
- Advanced text editors
- IDEs (Integrated Development Environments)
Localisation and CAT tools

Software localisation

<table>
<thead>
<tr>
<th>File Formats</th>
<th>Tools</th>
</tr>
</thead>
<tbody>
<tr>
<td>exe</td>
<td>Localisation tools</td>
</tr>
<tr>
<td>po</td>
<td>Advanced text editors</td>
</tr>
<tr>
<td>cs</td>
<td>IDE (Integrated Development Environment)</td>
</tr>
<tr>
<td>xml</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>
Localisation and CAT tools

**Game localisation**

File Formats

<table>
<thead>
<tr>
<th>exe</th>
<th>po</th>
<th>cs</th>
<th>xml</th>
</tr>
</thead>
</table>

Tools

- Localisation tools
- Advanced text editors
- Spreadsheet programs
- IDE (Integrated Development Environment)
Standards of localisation
Standards

**TMX**
- Translation Memory eXchange
- Standarised format to store and interchange translation memories
- Current version: TMX 1.4b
- Well supported by the CAT tool ecosystem
TMX – Translation Unit

Standards

SRX

– Segmentation Rule Exchange
– Segmentation rules can be defined for a specific text and language
– It increases the possibilities of reusing translated material (leveraging)
ITS

– Internationalization Tag Set
– It is a set of data categories which make localisation of XML content easier (adapted from Kosek 2011)

– ITS 1.0 was published in 2007
– ITS 2.0 was published in 2013

Example 3: ITS markup on elements in an XML document (local approach)

```xml
<article xmlns="http://docbook.org/ns/docbook"
          xmlns:its="http://www.w3.org/2005/11/its"
          its:version="2.0" version="5.0" xml:lang="en">
  <info>
    <title>An example article</title>
    <author its:translate="no">
      <firstname>John</firstname>
      <surname>Doc</surname>
    </author>
    <affiliation>
      <address><email>foo@example.com</email></address>
    </affiliation>
  </info>
  <para>This is a short article.</para>
</article>
```

[Source file: examples/xml/EX-basic-concepts-1.xml]
XLIFF

– It is a data container that carries localisable data from one localisation process to the other

– Its main feature is that it allows the interoperability between different tools (it is an open standard!)

– It is currently developed by OASIS

– Around ten people currently work on its development

http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=xliff

The idea is to extract the localisation-related data from its original format and merge it back when the process is finished
XLIFF

- XLIFF 1.0 approved as an OASIS Committee Specification in 2002
- XLIFF 1.2 approved in 2008
- XLIFF 2.0 approved in August 2014
XLIFF 2.0

- Core & Modules strategy.
  - The core contains only essential information
  - Eight specific modules already defined
  - They can store information about specific localisation processes:

<table>
<thead>
<tr>
<th>Translation Candidates</th>
<th>Glossary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Format Style</td>
<td>Metadata</td>
</tr>
<tr>
<td>Resource Data</td>
<td>Change Tracking</td>
</tr>
<tr>
<td>Size Restriction</td>
<td>Validation</td>
</tr>
</tbody>
</table>
### Standards

**XLIFF support in CAT tools:**

<table>
<thead>
<tr>
<th>Tool Name</th>
<th>XLIFF 1.2 Support</th>
<th>Own custom extension</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alchemy Catalyst 10</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Araya XLIFF Editor</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Microsoft LBA 5.3/MLP 5.3</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>MemoQ 5.0.64</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>MultiTrans Prism</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Okapi Framework M16</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>OmegaT 2.3.0</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>OpenTM2 V.0.9.5</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>SDL Trados Studio 2011 SP1</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Solas v1</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Swordfish Translation Editor</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Translation Factory 4.5.14</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Translation Workspace 1.10</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>XTM 6.2</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Filip & Morado (2013)

---

### Benefits of XLIFF:

- It is tool neutral
- Tool vendors do not have to work on developing support for many file formats
- There is less dependency on specific tools
- There is a better control over the flow of information
- Other benefits that XML have: customisation, wide range of rendering tools, etc.
Translation technologies training

Our objective:

– Empower the translator:
  • He/she needs to be in control of (and understand) the files he/she is manipulating

– Prepare our students to the localisation/translation market, but also:
  • Critical thinking: tools are not going to translate for you
  • Awareness of the advantages and disadvantages
Translation technologies training

Courses offered at UNIGE (directly related to CAT tools):

– Outils d’aide à la traduction (BA), fr
– Traduction assistée par ordinateur (MA), fr
– Localisation et gestion de projets (MA), en

Formation continue:

UN REGARD SUR LES MEMOIRES DE TRADUCTION
(Wordfast Classic, Wordfast Pro, Wordfast Anyware, Trados Studio 2014, Multitrans Prism, MemoQ, Wordbee)
du 7 novembre 2014 au 16 janvier 2015
Contact: Veronique.Sauron@unige.ch

Challenges:

– Constant evolution of the field and tools (Folaron 2010):
  • Our labs need to be updated every semester: new tools, new version, new areas (e.g. mobile localisation)
– Budget restriction
– Different student backgrounds (and technical competences)
• **Our approach:**
  – Constant learning process
  – Focus on tool functionalities (not on the tools themselves)
  – Participation in standards organisations (as well as conferences, tool presentations, seminars, etc.)
  – Close contact with translators and localisers, former students (companies and organisations as well)
  – Flexibility in the expected level (depending on the technical experience of students)

---

**Summary and conclusion**
— CAT tools can be your ally, but they should be used from a critical point of view:

Their usefulness depends on the specific context of use

— Translators and localiser might also need to use other domain-specific tools: not only CAT tools

— Standards of localisation are important to exchange translatable data but they also allow translators to use the tool of their choice

— CAT tools as well as computer science are in constant evolution, adapting to them represents a time-consuming task that should not be underestimated

Thank you!

lucia.morado@unige.ch
28/10/2014