Training translation students to evaluate CAT tools using Eagles: a case study

Marianne Starlander and Lucía Morado Vázquez

University of Geneva
Faculty of Translation and Interpreting
Marianne.Starlander@unige.ch
Lucia.Morado@unige.ch
Table of contents

- Introduction
- CAT Course
- CAT tool evaluation
  - Evaluation methods
  - EAGLES
- Case study
  - Introduction
  - General outcome
  - Questionnaire
- Future work
Introduction

Motivation

Alumni soon being considered as CAT tool experts in their new jobs (often internships)

Skills needed:

- Practical knowledge about CAT tools
- Evaluation skills
Evaluation in the translator’s training

- Not generally identified as one of the necessary skills or competences that students need to acquire during a typical translation technology course

- But: Pym (2012): “ability to evaluate the suitability of a tool in relation to technical needs and price”

- Convey learning and assessing techniques to students rather than training on specific industry tools
CAT Course
CAT Course - Facts

- Optional subject within the MA in Translation at the University of Geneva
- 1 semester duration – 5 ECTS
- 60-65 students
- 1 lecturer, 1 lab assistant
CAT Course - methodology

➢ 1 hour theoretical course:
   A. Introduce the students to CAT tools
   B. Focus on translation memory systems (TMS)
   C. CAT tool evaluation: applied to TM systems

➢ 1-2 hours in the lab room:
   A. Translation kits
   B. Evaluation assignment
CAT Course: focus on TMS

Basic TM system functionalities:

- Translating with a TM system
- Alignment
- Translation memory management
- Terminology management
- Project creation and management
CAT Course: hands-on TMS

Two popular commercial tools used:

SDL Trados Studio 2011

MultiTRANS Prism
CAT Course: TMS assignment

Two assignments: Real case scenario using both systems (SDL Trados 2011 and Multitrans Prism)

Translation toolkit:
- Instructions
- Reference files (to be aligned)
- Translatable file (.docx)

Deliverables:
- Invoice
- TBX file
- TMX file
- Translated file (.docx)
CAT Tool Evaluation
EAGLES – 7-step recipe

Why EAGLES?

- EAGLES = Expert Advisory Group on Language Engineering Standards
- European Project (started 93 - report 96)
- Several deliverables, including:
  - Seven-step recipe (Geneva, 1999)
  - Aim: provide an evaluation framework for natural language processing systems
Seven steps

1. Define the aim of the evaluation
2. Elaborate a task model
3. Define top level quality characteristics
4. Produce detailed requirements for the system under evaluation
5. Devise the metrics to be applied to the system for the requirements produced under 4
6. Design the execution of the evaluation
7. Execute the evaluation
Case study
Case study

CAT tool comparison task

- Two CAT tools evaluated and compared.
- EAGLES 7-step recipe
- Report
- Oral presentation
- Real case scenarios
- Group activity
How to explain the 7-step recipe

- Determine a precise context of use
  - «there is no such thing as a best system, but a best system for a particular situation» (Rico, 2001)
  - Context oriented approach

- Give example for the task model
  - Functionality: core characteristic evaluated
  - Ex: TM → provide previously translated sentences

- Elaborate a quality model

- How to chose the metrics
Define the metrics

- Different types: yes/no; scale; time, success rate, error rate
- Important: avoid subjectivity → determine metrics and scales

- Example: term retrieval

<table>
<thead>
<tr>
<th>Points</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Good</td>
</tr>
<tr>
<td>2</td>
<td>Satisfactory</td>
</tr>
<tr>
<td>1</td>
<td>Not satisfactory</td>
</tr>
<tr>
<td>0</td>
<td>Not existant</td>
</tr>
</tbody>
</table>

Shortcut  -> good
1 click   -> satisfactory
Max. 2 clicks  -> satisfactory
> 2 clicks  -> not satisfactory
Type of evaluation

- Feature inspection: checklist
  - mostly yes/no

- Benchmarking testing: the performance of individual functions, system modules or the entire system can be evaluated
  - only on one selected feature according to chosen user-context
Case study

Real case scenarios: 4 choices

1. Newly graduated freelance translator: 7/11 groups

2. Experienced freelance translator (15 years) without CAT tool experience: 3/11 groups

3. In-house translator (10 years same company) so far using SDL Trados: 1/11 group

4. Translation support manager in an International Organisation: 0/11
## TMS combination

<table>
<thead>
<tr>
<th>System 1</th>
<th>System 2</th>
<th>Number of groups</th>
<th>Functionality chosen</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDL Trados 2011</td>
<td>Wordfast Anywhere</td>
<td>3/11</td>
<td>Alignment (2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Terminology sharing</td>
</tr>
<tr>
<td>SDL Trados 2011</td>
<td>Wordfast Classic</td>
<td>2/11</td>
<td>Alignment</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Terminology</td>
</tr>
<tr>
<td>SDL Trados 2011</td>
<td>Wordfast Pro</td>
<td>1/11</td>
<td>Terminology</td>
</tr>
<tr>
<td>SDL Trados 2011</td>
<td>Across</td>
<td>1/11</td>
<td>Translation</td>
</tr>
<tr>
<td>SDL Trados 2011</td>
<td>Déjà Vu</td>
<td>1/11</td>
<td>Translation</td>
</tr>
<tr>
<td>Multitrans Prism</td>
<td>Wordfast Anywhere</td>
<td>1/11</td>
<td>Exchanging TMs</td>
</tr>
<tr>
<td>Multitrans Prism</td>
<td>Wordfast Pro</td>
<td>1/11</td>
<td>Alignment</td>
</tr>
<tr>
<td>SDL Trados 2011</td>
<td>OmegaT</td>
<td>1/11</td>
<td>Alignment</td>
</tr>
</tbody>
</table>
General outcome

All participants managed to finished the task
11 groups of 4-5 students
11 reports – around 10 pages
Oral presentations were clear and followed
the guidelines (5 minutes time constraints)
### Example

**Graph:**
- **Joshua Gryphon**
- **Moutier**
- **Geneva**
- **SDL Trados**
- **Time well spent on the train**
- **Wordfast Anywhere**
- **Costs of software**
- **commuting by train**

**Table:**

<table>
<thead>
<tr>
<th>Alignment functionality</th>
<th>Time spent for alignment</th>
<th>fast Internet</th>
<th>slow Internet</th>
<th>no Internet</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1-5</td>
<td>1, 3, 5</td>
<td>1, 3, 5</td>
<td>1, 3, 5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No functionality: 1</td>
<td>No functionality: 1</td>
<td>No functionality: 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Partial functionality: 3</td>
<td>Partial functionality: 3</td>
<td>Partial functionality: 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Full functionality: 5</td>
<td>Full functionality: 5</td>
<td>Full functionality: 5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>20+ min: 1</th>
<th>16 – 20 min: 2</th>
<th>11 -15 min: 3</th>
<th>5 – 10 min: 4</th>
<th>&lt;5 min: 5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Example: result tables

<table>
<thead>
<tr>
<th>Functionality</th>
<th>Wordfast Anywhere</th>
<th>Multitrans Prism</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exporting in TMX</td>
<td>yes – 1 point</td>
<td>yes – 1 point</td>
</tr>
<tr>
<td>Number of clicks needed to export a TM into .tmx</td>
<td>14 clicks – 0 point</td>
<td>5 clicks – 1 point</td>
</tr>
<tr>
<td>Time needed to export into .tmx</td>
<td>5 minutes 19 seconds – 0 point</td>
<td>30 seconds – 1 point</td>
</tr>
<tr>
<td>Sharing function exists</td>
<td>yes – 1 point</td>
<td>yes – 1 point</td>
</tr>
<tr>
<td>Number of clicks needed to share TM</td>
<td>9 to 10 clicks – 0 point</td>
<td>5 clicks – 1 point</td>
</tr>
<tr>
<td>Time needed to share a TM</td>
<td>56 seconds – 0 point</td>
<td>30 seconds – 1 point</td>
</tr>
<tr>
<td>FINAL SCORE</td>
<td>2 points</td>
<td>6 points</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Module</th>
<th>Trados</th>
<th>Déjà Vu</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anayse</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Création d’une projet</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Traduction</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Compatibilité de format</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Langues</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Assistance technique</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Prix</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Total de points</td>
<td>26</td>
<td>21</td>
</tr>
</tbody>
</table>
Main difficulties

- EAGLES/ISO Terminology and concepts
- Some ambitious evaluation plans → superficial evaluations
- Subjectivity when choosing the final scale and interpretation
  - Clarification of certain aspects of EAGLES (notably the classification into characteristics)
  - More concrete examples needed
Results from the questionnaire
The EAGLES seven steps recipe was easy to understand and implement

The EAGLES methodology helped me to establish my own evaluation criteria
Attitude towards the task

I enjoyed the scenario driven evaluation task we had to implement during the CAT course.

This activity represented an excessive workload compared to the other assignments fulfilled during the CAT course.
Future use

I think that the evaluation methodology learnt during the CAT course will be useful in my future career as a translator.

In my professional life, I will use the EAGLES method to design an objective CAT tool evaluation before buying a CAT system.
EAGLES as part of the MA CAT course

I would recommend the lecturers to continue including this evaluation method in the forthcoming CAT course.

I think that software evaluation skills should be part of the translator’s training curricula.
Discussion and future work

- Clarification of certain aspects of EAGLES (notably the classification into characteristics)

- More concrete examples needed

- Predefined set of pre-defined scenarios?

- Time-constraint: reflects real life
Better introduction of quality model

- Explain the ISO quality characteristics and sub-characteristics with an every-day life example
Pre-purchase evaluation for a pair of trousers

Context:
- Man
- Size: 1m90
- Usage: wedding in summer
- Specific requirements:
  - be able to reuse in future for other occasion
  - budget: maximum 150 CHF
Quality model for pair of trousers

Quality of trousers

Size

Design

Reusability

Price
Quality model – more details

Quality

- size
- design
- reusability
- price

- colour
- cut
- fabric
Conclusion

- EAGLES can be used in a translation training program: Successful completion rate
- Students understood the utility of the evaluation task
- Some improvements have already been implemented

Thank you!
References

- EAGLES 7 steps (1999):

- Cf. references in the paper