Translation Technology in ILSAs: striving to combine apps, discernment and capacity-building
ILSAs AND TRANSLATION

International assessments:
- Gather comparable data
- **Multilingual content** is needed to present the material in the language of the respondents

TRANSFORMATION

Computing

Technical evolution

BEST PRACTICES
MULTILINGUAL CONTENT: LIFECYCLE

Stages of the multilingual content life cycle

**Best Practices**

**Translation industry best practices:**
- Optimal preparation of projects for translation
- Optimal working environment and language assets
- Linguistic quality assurance and quality control
TRANSLATION BEST PRACTICES

Standards:
- Procedures: language assets (glossaries, translation memories, dictionaries, parallel corpora, etc.)
- Formats: enabling exchange (translation memories, terminology, segmentation rules)
- Basic functionality: Computer Assisted Translation (CAT) tools

Benefits:
- Interoperability
- Better working conditions for linguists (ergonomics and usability)
- Higher quality and higher productivity
BEST PRACTICES: SOME EXAMPLES

**Project Preparation**
- Item developers
  - Source content
- Platform architects
  - Translation files (e.g. XLIFF)

**LQA**
- Guidelines & glossaries

**Linguists**
- National/target versions

**LQC checks**
- Final target versions

**Best Practices**
- Segmentation
- Parsing
- Markup encapsulation
- Term extractor
- Translation memories
- CAT Tools
- Bilingual glossaries
- Automated checks
TRANSLATION BEST PRACTICES: OFTEN NEGLECTED

Instead of complying with standards or using existing available technology, architects sometimes reinvent the wheel:

- Creating new translation interfaces:
  - Lack ergonomics
  - Lack the necessary functionalities we already have in existing CAT tools

As a result:

- Things are more complicated
- More time and money than necessary are spent

**Impact on quality:** suboptimal process → suboptimal results
PROJECT PREPARATION
PROJECT PREPARATION: SENTENCE-BASED SEGMENTATION

Long documents need to be split so that they can be handled more efficiently:

- Shorter chunks entail a smaller cognitive effort
- Leveraging translation materials easier

General practice: split full texts in sentences.

- As a rule of thumb, the sentence is the basic unit of processing

Texts/documents need to be split at the end of each sentence.

- After a full stop (.), a question mark (?), an ellipsis (…), etc.
- Exceptions: abbreviations, e.g. “Mr. Robinson”

Segmentation rules and exceptions are written in regular expressions, e.g.

- Before: \[.\?\!]+
- After: \s

International standard: SRX
This is a full paragraph that has not been segmented.

It is more difficult to translate than single sentences.
SEGMENTATION: BY SENTENCE

This is the same paragraph segmented by sentence.

This is easier to translate or revise, as the linguist can focus on one sentence at a time.
PROJECT PREPARATION: ENCAPSULATING MARKUP

Besides text, documents may contain some codes (markup):  
• Document structure (e.g. lists, tables)  
• Formatting information (e.g. bold, italics)

Markup is essential for the environment but may be intricate for linguists.

XLIFF specifications:  
▪ parsing structural markup  
▪ encapsulating inline markup
Neither segmentation nor markup encapsulation are applied in this file:
MARKUP: ENCAPSULATED

Markup encoding and segmentation can make a huge difference for the linguists.
This segment is so long that there is virtually no space left in the translation editor to enter the translation. Real example from PIAAC, linguists’ nightmare.
PROJECT EXECUTION
**Prehistoric approach**

- Source files (in Word, Excel)
- Key Terms definitions (in Word)
- Bilingual glossary (in Excel or a Word table)
- Guidelines (in Word, Excel)

**Reference files:**
- Materials from the previous cycle (in Word, Excel)
- Related research

**Computer assisted approach**

- Project download link / Project package
- Source: `.xliff`
- Termbase: `.tbx`
- Guidelines
- Translation memory from previous cycle: `.tmx`

**Reference files:**
- Related research
**LINGUISTIC QUALITY ASSURANCE: GLOSSARIES**

Glossaries: often underused and underrated.

ILSAs may not contain large amount of technical terminology, but there are many words and expressions that should be translated consistently (e.g. recurring instructions).

```
1 consuming farm
2 consumption
2 consumption of milk
3 contains
2 council
30 cow's
30 cow's milk
24 dairy
2 dairy foods
2 dairy foods association
5 dairy milk
5 dairy milk products
```
ISSUES: TYPO

The source version talks about airplanes and a space suit, to which the respondent will need to refer to obtain full credit.
AVOIDABLE ISSUES: TYPO

The term “space suit” was translated as “special suit”.

That mistranslation makes it impossible to reply correctly to the question, because the hint to “space” is missing.
Bilingual glossaries and key term definitions can be imported in CAT tools.

Displayed in a separate pane ("Glossary" pane in OmegaT).

Predictive typing
TRANSLATION: CAT TOOLS

“A specialized program that helps a translator to increase productivity and obtain a higher degree of consistency and quality, by leveraging on special translation-boosting functionalities and language assets such as translation memories and term bases.”

- Ensure consistency in a specific project and across cycles of the same project (trend management).
- Commercial and Open Source CAT tools provide the same functionalities.
OMEGAT

Color coding for navigation between segments
Glossary pane
Underlined terms
Translation and Adaptation notes
Matches from TM in track changes
Professional CAT tool accessible to all
... with a bit of training
Translation Memories

Reuse of previous material from the past.

Not necessarily 100% identical:
- Fuzzy matches (75 – 99%) are also helpful

1 translation memory for each language combination of a project.
ENSURING CONSISTENCY INSIDE A PROJECT

CAT Tools and working translation memory
CREATING TRANSLATION MEMORIES

How?

- Create translation memories through alignment
TRANSLATION MEMORIES

Hierarchy

More reliable versions: how can a distinction be made?

- Applying penalties to less reliable translation memories

Trend management

How can trend content be visible, recognizable and protected?

- Distinction: color coding
- Protection: locked segments

1. Very dissatisfied
   >>>>
   Muy insatisfecho
   <100/100/100%> Quest_EN_ES.tmx (+1 more)

2. Very dissatisfied
   >>>>
   Very insatisfecho
   <70/70/70%> penalty-030\QQ_Previous_version_penalty.tmx

According to the IDFA, with which statement do leading health professionals and organizations agree?

Consuming milk and milk products leads to obesity.
Milk is a good source for essential vitamins and minerals.
Linguistic Quality Control (LQC): 
- Checks compliance 
- Reports deviations 
- Proposes corrective action 

Expert verifiers should not spend time fixing inconsistencies in recurring terms, correcting systematic punctuation errors and typos, adding missing formatting tags, etc.

Automated QC routines enhance quality and allow verifiers/reviewers to focus on equivalence issues.
LINGUISTIC QUALITY CONTROL

Some basic checks are already available directly from CAT tools:

- completeness
- spelling
- tags
LINGUISTIC QUALITY CONTROL: VERYFIRE

Set of web-based application prototypes developed at cApStAn using regular expressions.

During the analysis, specific “rulesets” are called up based on the file name.

3 types of rulesets: generic, project-specific and language-specific.
AUTOMATED LQC

Segment level consistency checks:
- Identical source segments have been translated differently
- Identical target segments have a different source text

Glossary adherence checks: based on the glossaries created at LQA

Compliance to style guide

Completeness checks: empty target segments

Unaltered segments: translation is identical to source

Language-specific checks: language-specific punctuation

Etc…
# AUTOMATED LQC: REPORT

<table>
<thead>
<tr>
<th>Error</th>
<th>Row</th>
<th>Ruleset:Rule ID</th>
<th>Function:Pattern</th>
<th>Error type</th>
<th>File name</th>
<th>Source segment</th>
<th>Target segment</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20</td>
<td>Generic checks</td>
<td>Unaltered segment</td>
<td></td>
<td>TEST_Cows_Milk_fra-FRA_1_Translation.xls</td>
<td>To support the use of Farm to Market Dairy products.</td>
<td>To support the use of Farm to Market Dairy products.</td>
<td>The text in the translation is identical to the source text.</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>Generic checks</td>
<td>Target segment inconsistency</td>
<td></td>
<td>TEST_Cows_Milk_fra-FRA_1_Translation.xls</td>
<td>Farm to Market</td>
<td>La Ferme laitière</td>
<td>Identical source segments have been translated differently.</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>Generic checks</td>
<td>Target segment inconsistency</td>
<td></td>
<td>TEST_Cows_Milk_fra-FRA_1_Translation.xls</td>
<td>Farm to Market</td>
<td>La Ferme de lait</td>
<td>Identical source segments have been translated differently.</td>
</tr>
<tr>
<td>3</td>
<td>13</td>
<td>Generic checks</td>
<td>Source segment inconsistency</td>
<td></td>
<td>TEST_Cows_Milk_fra-FRA_1_Translation.xls</td>
<td>Click on Yes or No to answer the question.</td>
<td>Pour répondre à la question, cliquez sur l’un des choix de réponse.</td>
<td>Different source segments have been translated identically.</td>
</tr>
<tr>
<td>3</td>
<td>15</td>
<td>Generic checks</td>
<td>Source segment inconsistency</td>
<td></td>
<td>TEST_Cows_Milk_fra-FRA_1_Translation.xls</td>
<td>Click on Yes or No to answer the question.</td>
<td>Pour répondre à la question, cliquez sur l’un des choix de réponse.</td>
<td>Different source segments have been translated identically.</td>
</tr>
</tbody>
</table>
OPTIMAL PROCESSES FOR OPTIMAL RESULTS

In order to ensure equivalence of translated assessments, translation technology should intervene from project preparation to project execution

- In ILSAs, open source solutions to be preferred
- Accessible to all players: platform architects, linguists, domain experts and national centres

Optimal processes require planning and preparation

⇒ Spend more time and effort **before** the translation begins
⇒ Choose your tools with a specialist

And all players will benefit from translation technology
Incorporating Machine Translation (MT) in current workflows?
Creating new workflows that allow better use of new technologies?
THANK YOU FOR YOUR ATTENTION!  QUESTIONS?
TRANSLATION: SERVER BASED REPOSITORY OF TMS