



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

Danina Lupsa
danina.lupsa@capstan.be

ILSAs AND TRANSLATION

International assessments:

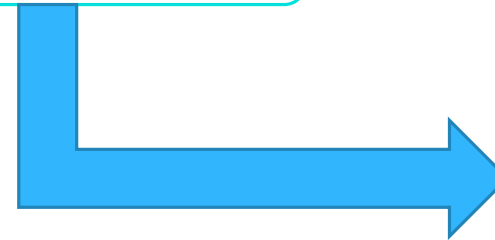
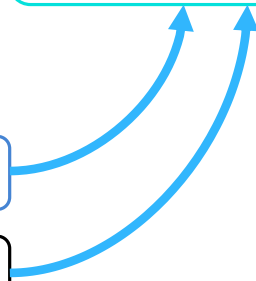
- Gather comparable data
- **Multilingual content** is needed to present the material in the language of the respondents



TRANSLATION

Computing

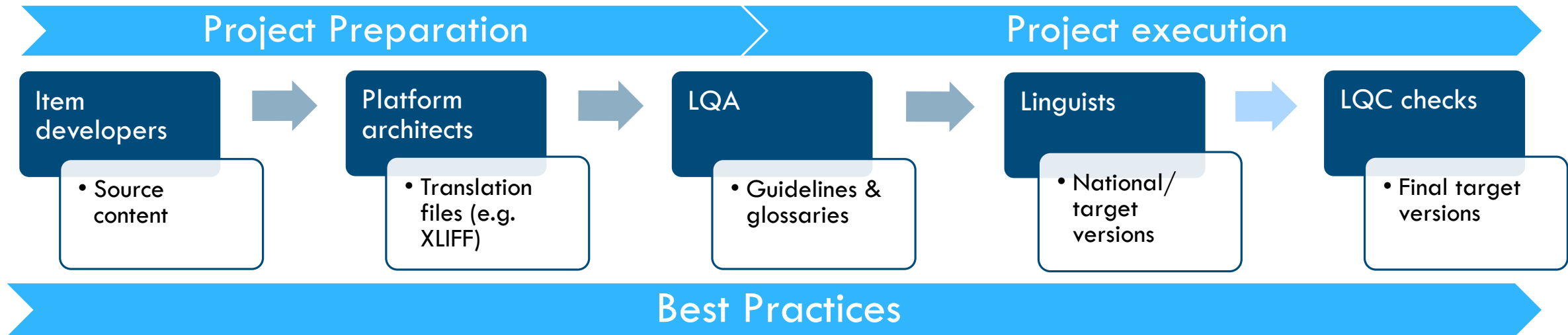
Technical evolution



BEST PRACTICES

MULTILINGUAL CONTENT: LIFECYCLE

Stages of the multilingual content life cycle



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

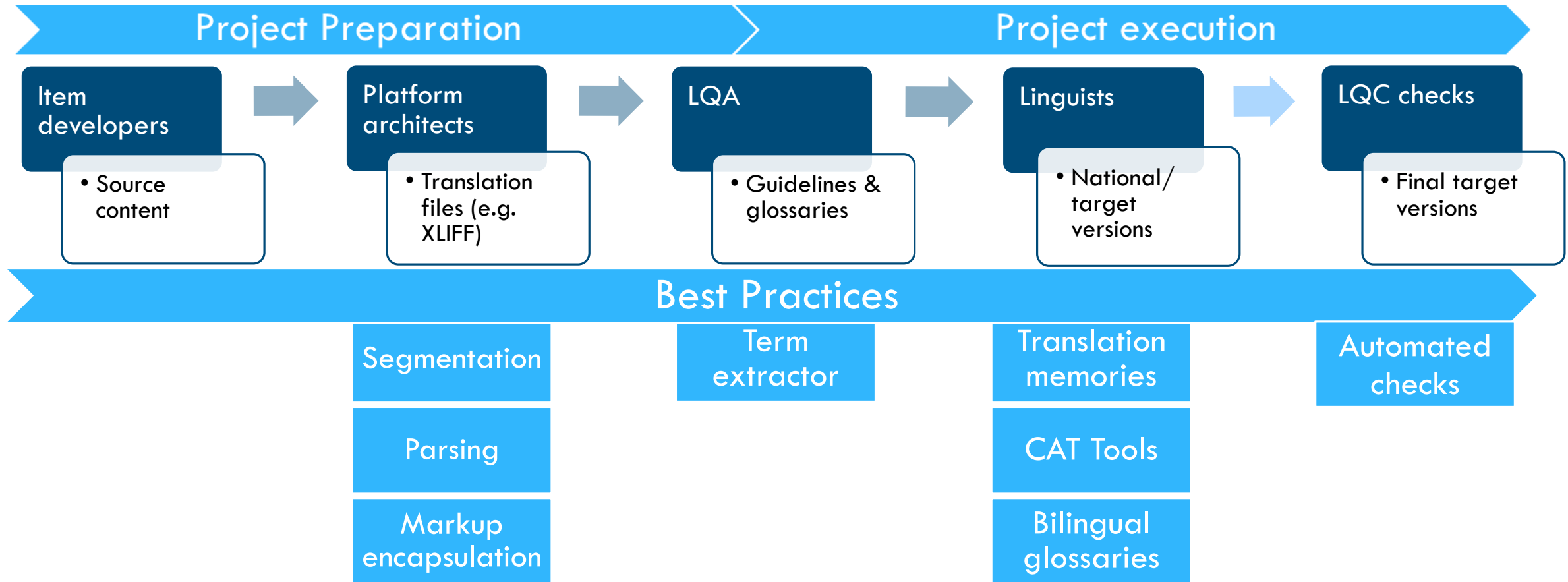
Standards:

- **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



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:

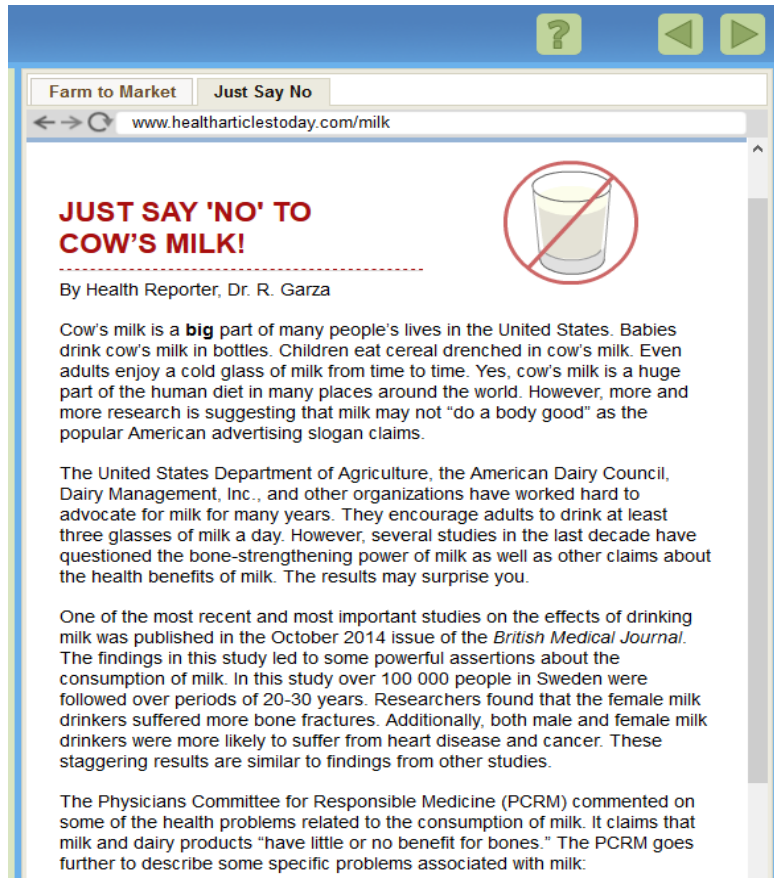
- Things are more complicated
- More time and money than necessary are spent

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

SEGMENTATION: FULL PARAGRAPHS

This is a full paragraph that has not been segmented.

It is more difficult to translate than single sentences.


 OmegaT 4.1.2_2 :: 3_Verification

Project Edit Go To View Tools Options Help

Editor - PISA_fra-ZZZ_REA_R557-CowsMilk_eng.xlf

serait pas « tout le bien qu'on se fait », comme l'affirme un célèbre slogan publicitaire américain.

The United States Department of Agriculture, the American Dairy Council, Dairy Management, Inc., and other organizations have worked hard to advocate for milk for many years. They encourage adults to drink at least three glasses of milk a day. However, several studies in the last decade have questioned the bone-strengthening power of milk as well as other claims about the health benefits of milk. The results may surprise you.

Depuis de nombreuses années, le ministère américain de l'agriculture, le Conseil américain des produits laitiers, l'entreprise Dairy Management, Inc. et d'autres organisations multiplient leurs efforts pour faire la promotion du lait. Ils encouragent les adultes à boire au moins trois verres de lait par jour. Pourtant, plusieurs études menées ces dix dernières années contestent que le lait renforce les os, ainsi que d'autres supposés effets bénéfiques pour la santé. Les résultats pourraient vous surprendre.

Last modified by unknown

One of the most recent and most important studies on the effects of drinking milk was published in the October 2014 issue of the <i>British Medical Journal</i>. The findings in this study led to some powerful assertions about the consumption of milk. In this study over 100 000 people in Sweden were followed over periods of 20-30 years. Researchers found that the female milk drinkers suffered more bone fractures. Additionally, both male and female milk drinkers were more likely to suffer from heart disease and cancer. These staggering results are similar to findings from other studies.

Une des études les plus récentes et les plus importantes sur les effets de la consommation de lait a été publiée dans le numéro d'octobre 2014 du <i>British Medical Journal</i>. Les résultats de cette étude ont donné lieu à des affirmations tranchées au sujet de la consommation de lait. Dans cette étude menée en Suède, plus de 100 000 personnes ont été suivies sur des périodes allant de 20 à 30 ans. Les chercheurs ont constaté que les femmes qui buvaient du lait subissaient davantage de fractures osseuses. Par ailleurs, autant les hommes que les femmes buvant du lait étaient plus susceptibles de souffrir de maladies cardiaques et de cancer. Ces résultats stupéfiants sont similaires aux conclusions d'autres études. <segment 0021>

The Physicians Committee for Responsible Medicine (PCRM) commented on some of the health problems related to the consumption of milk. They claim that milk and dairy products "have little or no benefit for bones." The PCRM goes further to describe some specific problems associated with milk:

Le Comité des médecins pour une médecine responsable, le CMMR (Physicians Committee for Responsible Medicine, PCRM) s'est exprimé sur certains problèmes de santé liés à la consommation de lait. Il affirme que le lait et les produits laitiers « ne présentent que peu ou pas de bienfaits pour les os ». Le CMMR va plus loin en décrivant certains problèmes précis associés au lait :

Glossary

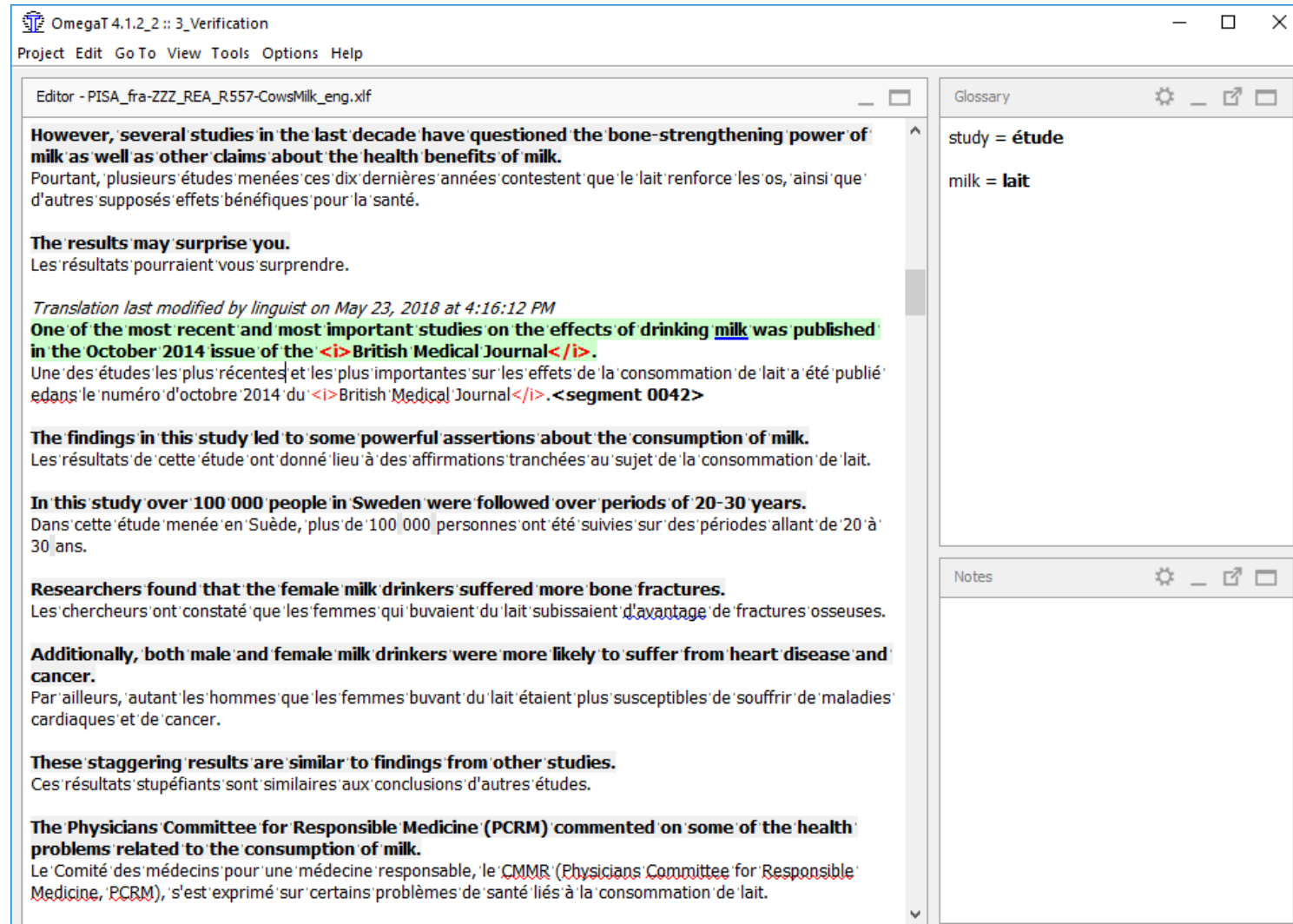
study = étude
 milk = lait

Notes

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.



OmegaT 4.1.2.2 :: 3_Verification

Project Edit Go To View Tools Options Help

Editor - PISA_fra-ZZZ_REA_R557-CowsMilk_eng.xlf

However, several studies in the last decade have questioned the bone-strengthening power of milk as well as other claims about the health benefits of milk.
 Pourtant, plusieurs études menées ces dix dernières années contestent que le lait renforce les os, ainsi que d'autres supposés effets bénéfiques pour la santé.

The results may surprise you.
 Les résultats pourraient vous surprendre.

Translation last modified by linguist on May 23, 2018 at 4:16:12 PM

One of the most recent and most important studies on the effects of drinking milk was published in the October 2014 issue of the <i>British Medical Journal</i>.
 Une des études les plus récentes et les plus importantes sur les effets de la consommation de lait a été publiée dans le numéro d'octobre 2014 du <i>British Medical Journal</i>. <segment 0042>

The findings in this study led to some powerful assertions about the consumption of milk.
 Les résultats de cette étude ont donné lieu à des affirmations tranchées au sujet de la consommation de lait.

In this study over 100 000 people in Sweden were followed over periods of 20-30 years.
 Dans cette étude menée en Suède, plus de 100 000 personnes ont été suivies sur des périodes allant de 20 à 30 ans.

Researchers found that the female milk drinkers suffered more bone fractures.
 Les chercheurs ont constaté que les femmes qui buvaient du lait subissaient davantage de fractures osseuses.

Additionally, both male and female milk drinkers were more likely to suffer from heart disease and cancer.
 Par ailleurs, autant les hommes que les femmes buvant du lait étaient plus susceptibles de souffrir de maladies cardiaques et de cancer.

These staggering results are similar to findings from other studies.
 Ces résultats stupéfiants sont similaires aux conclusions d'autres études.

The Physicians Committee for Responsible Medicine (PCRM) commented on some of the health problems related to the consumption of milk.
 Le Comité des médecins pour une médecine responsable, le CMMR (Physicians Committee for Responsible Medicine, PCRM), s'est exprimé sur certains problèmes de santé liés à la consommation de lait.

Glossary

study = étude

milk = lait

Notes

PROJECT PREPARATION: ENCAPSULATING MARKUP

```
1  <!DOCTYPE html>
2  <html>
3  <head>
4  <meta charset="utf-8" />
5  </head>
6  <body>
7  <p>My father has <span style="color:darkolivegreen;font-weight:bold;
   display:block;word-wrap:break-word;width:50px;white-space:
   normal;border:5px solid red;background-color:black;font-family:
   :courier;padding:20px;">dark green</span> eyes.</p>
8  </body>
9  </html>
```

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

The screenshot shows the OmegaT 4.0 application window. The title bar reads "OmegaT 4.0". Below it is a menu bar with "Project", "Edit", "Go To", "View", "Tools", "Options", and "Help". The main editor area has a tab titled "Editor - origi/markup.xlf". The text in the editor is as follows:

```
Translation last modified by Manuel on Mar 22, 2017 at 11:22:43 AM
<opens some-attribute="some-long-value" id="
1938472938473.129374329874" rid="UNIT.2.0" fontsize="11"/> You will
now learn about the next exercises in the survey.<closes id="
1938472938473.129374329874" rid="UNIT.2.0"/><opens id="
2393743293498734923.23974392" rid="UNIT.4.0" fontsize="11"/> Please
click <b>Next</b> to continue.<closes id="
2393743293498734923.23974392" rid="UNIT.4.0"/>
[<opens some-attribute="some-long-value" id="1938472938473.129374329874"
rid="UNIT.2.0" fontsize="11"/> Ὑοὺς οἰκτῶν τὰ ἀρτὰ ἀβούτ' ἐλὲ νῆεξ' ἐκεῖς ἐζιγίεζ' ἵη' ἐλὲ
αὐτοῦ ἐξ. τοῖς ἐμ' ἱπεύμ' δότ' ὁπ' ]
<closes id="1938472938473.129374329874" rid="UNIT.2.0"/><opens id="
2393743293498734923.23974392" rid="UNIT.4.0" fontsize="11"/> πλῆσθ' ἐξέλεξ'
<b>Νέεξ</b> τοῖς ἐμ' ἱπεύμ' τοῖς ἐμ' ἱπεύ.<closes id="
2393743293498734923.23974392" rid="UNIT.4.0"/>]<segment 0041>
```

The screenshot shows the OmegaT 4.0.0 application window. The title bar reads "OmegaT 4.0.". Below it is a menu bar with "Project", "Edit", "Go To", "View", "Tools", "Options", and "Help". The main area is titled "Editor - optim/markup.xlf.xlf". It contains two translation units from an XLIFF file.

Segment 1:

- Translation last modified by Manuel on Apr 18, 2017 at 1:24:41 PM
- Source: <x0/>You will now learn about the next exercises in the survey.
- Target: [

Segment 2:

- Translation last modified by Manuel on Mar 22, 2017 at 11:23:37 AM
- Source: <x0/><x1/>Please click<x2/>Next<x3/> to continue.<x4/>
- Target: [<x0/><x1/>þləázè'çřic'k<x2/>Nèχf<x3/>'fô'çõñfiñúè'lôřèm'ĩpəúm.<x4/>]

OmegaT 4.1.2.2 :: 1_Translation

Project Edit Go To View Tools Options Help

Editor - PISA_eng_ZZZ_REA_R548-ChickenForum_eng_FT2018.xlf

<bx clone="yes" id="oBLOCK_FRAGMENT.0.T4" rid="BLOCK_FRAGMENT.0.T4" cba:blockname="T4"/><bx clone="yes" id="oBLOCK_FRAGMENT.0.T11" rid="BLOCK_FRAGMENT.0.T11" cba:blockname="T11"/><bx clone="no" id="oPARAGRAPH.1.0" rid="PARAGRAPH.1.0"/><bx clone="yes" id="oPHRASE.0.0" rid="PHRASE.0.0" cba:bold="true" cba:fontname="Times New Roman" cba:italic="true"/>What is markup??<ex id="cPHRASE.0.0" rid="PHRASE.0.0"/><ex id="cPARAGRAPH.0.0" rid="PARAGRAPH.0.0"/><bx clone="no" id="oPARAGRAPH.1.0" rid="PARAGRAPH.1.0"/><bx clone="yes" id="oPHRASE.1.0" rid="PHRASE.1.0" cba:bold="true" cba:fontname="Times New Roman" cba:italic="true"/><ex id="cPHRASE.1.0" rid="PHRASE.1.0"/><ex id="cPARAGRAPH.1.0" rid="PARAGRAPH.1.0"/><bx clone="no" id="oPARAGRAPH.2.0" rid="PARAGRAPH.2.0"/><bx clone="yes" id="oPHRASE.2.0" rid="PHRASE.2.0" cba:fontname="Times New Roman" cba:fontsize="9"/>The source content might include codes, e.g. any HTML tag used to apply a certain behavior or property to part of the text.<ex id="cPHRASE.2.0" rid="PHRASE.2.0"/><ex id="cPARAGRAPH.2.0" rid="PARAGRAPH.2.0"/><bx clone="no" id="oPARAGRAPH.3.0" rid="PARAGRAPH.3.0"/><bx clone="yes" id="oPHRASE.3.0" rid="PHRASE.3.0" cba:fontname="Times New Roman" cba:fontsize="9"/><ex id="cPHRASE.3.0" rid="PHRASE.3.0"/><ex id="cPARAGRAPH.3.0" rid="PARAGRAPH.3.0"/><bx clone="no" id="oPARAGRAPH.4.0" rid="PARAGRAPH.4.0"/><bx clone="yes" id="oPHRASE.4.0" rid="PHRASE.4.0" cba:fontname="Times New Roman" cba:fontsize="9"/>Often source content is HTML or some sort of similar markup language, where therefor tags are used to define layout, formatting and/or structure. The preparation of the source files as XLIFF entails dealing with those codes as appropriate.<ex id="cPHRASE.4.0" rid="PHRASE.4.0"/><ex id="cPARAGRAPH.4.0" rid="PARAGRAPH.4.0"/><bx clone="no" id="oPARAGRAPH.5.0" rid="PARAGRAPH.5.0"/><bx clone="yes" id="oPHRASE.5.0" rid="PHRASE.5.0" cba:fontname="Times New Roman" cba:fontsize="9"/><ex id="cPHRASE.5.0" rid="PHRASE.5.0"/><ex id="cPARAGRAPH.5.0" rid="PARAGRAPH.5.0"/><bx clone="no" id="oPARAGRAPH.6.0" rid="PARAGRAPH.6.0"/><bx clone="yes" id="oPHRASE.6.0" rid="PHRASE.6.0" cba:fontname="Times New Roman" cba:fontsize="9"/><ex id="cPHRASE.6.0" rid="PHRASE.6.0" cba:bold="true" cba:fontname="Times New Roman" cba:italic="true"/>What does CIEM do?<ex id="cPHRASE.7.0" rid="PHRASE.7.0"/><ex id="cPARAGRAPH.7.0" rid="PARAGRAPH.7.0"/><bx clone="no" id="oPARAGRAPH.8.0" rid="PARAGRAPH.8.0"/><bx clone="yes" id="oPHRASE.8.0" rid="PHRASE.8.0" cba:bold="true" cba:fontname="Times New Roman" cba:italic="true"/><ex id="cPHRASE.8.0" rid="PHRASE.8.0"/><ex id="cPARAGRAPH.8.0" rid="PARAGRAPH.8.0"/><bx clone="no" id="oPARAGRAPH.9.0" rid="PARAGRAPH.9.0"/><bx clone="yes" id="oPHRASE.9.0" rid="PHRASE.9.0" cba:fontname="Times New Roman" cba:fontsize="9"/>Codes can be of two kinds:<ex id="cPHRASE.9.0" rid="PHRASE.9.0"/><ex id="cPARAGRAPH.9.0" rid="PARAGRAPH.9.0"/><bx clone="no" id="oPARAGRAPH.10.0" rid="PARAGRAPH.10.0"/><bx clone="yes" id="oPHRASE.10.0" rid="PHRASE.10.0" cba:fontname="Times New Roman" cba:fontstyle="4"/><ex id="cPHRASE.10.0" rid="PHRASE.10.0"/><ex id="cPARAGRAPH.10.0" rid="PARAGRAPH.10.0"/><bx clone="no" id="oPARAGRAPH.11.0" rid="PARAGRAPH.11.0"/><bx clone="yes" id="oPHRASE.11.0" rid="PHRASE.11.0" cba:fontname="Wingdings" cba:fontstyle="8"/>><ex id="cPHRASE.11.0" rid="PHRASE.11.0"/><bx clone="yes" id="oPHRASE.12.0" rid="PHRASE.12.0" cba:bold="true" cba:fontname="Times New Roman" cba:italic="true"/>Job Data Bank<ex id="cPHRASE.12.0" rid="PHRASE.12.0"/><ex id="cPARAGRAPH.11.0" rid="PARAGRAPH.11.0"/><bx clone="no" id="oPARAGRAPH.12.0" rid="PARAGRAPH.12.0"/><bx clone="yes" id="oPHRASE.13.0" rid="PHRASE.13.0" cba:fontname="Times New Roman" cba:fontstyle="9"/>I.e. codes that embed a sentence, or stand outside of a sentence, or between sentences, or operate at a higher level than the sentence should not be included in segments.<ex id="cPHRASE.13.0" rid="PHRASE.13.0"/><ex id="cPARAGRAPH.12.0" rid="PARAGRAPH.12.0"/><bx clone="no" id="oPARAGRAPH.13.0" rid="PARAGRAPH.13.0"/><bx clone="yes" id="oPHRASE.14.0" rid="PHRASE.14.0" cba:fontstyle="4"/><ex id="cPHRASE.14.0" rid="PHRASE.14.0"/><ex id="cPARAGRAPH.13.0" rid="PARAGRAPH.13.0"/><bx clone="no" id="oPARAGRAPH.14.0" rid="PARAGRAPH.14.0"/><bx clone="yes" id="oPHRASE.15.0" rid="PHRASE.15.0" cba:fontname="Wingdings" cba:fontstyle="8"/>><ex id="cPHRASE.15.0" rid="PHRASE.15.0"/><bx clone="yes" id="oPHRASE.16.0" rid="PHRASE.16.0" cba:bold="true" cba:fontname="Times New Roman" cba:italic="true"/>Intrasentential codes<ex id="cPHRASE.16.0" rid="PHRASE.16.0"/><ex id="cPARAGRAPH.14.0" rid="PARAGRAPH.14.0"/><bx clone="no" id="oPARAGRAPH.15.0" rid="PARAGRAPH.15.0"/><bx clone="yes" id="oPHRASE.17.0" rid="PHRASE.17.0" cba:fontname="Times New Roman" cba:fontstyle="9"/>I.e. codes that stand inside a sentence must be represented as inline elements also called "content markup" according to the guidelines of the XLIFF specification<ex id="cPHRASE.17.0" rid="PHRASE.17.0"/><ex id="cPARAGRAPH.15.0" rid="PARAGRAPH.15.0"/><bx clone="no" id="oPARAGRAPH.16.0" rid="PARAGRAPH.16.0"/><bx clone="yes" id="oPHRASE.18.0" rid="PHRASE.18.0" cba:fontstyle="4"/><ex id="cPHRASE.18.0" rid="PHRASE.18.0"/><ex id="cPARAGRAPH.16.0" rid="PARAGRAPH.16.0"/><bx clone="no" id="oPARAGRAPH.17.0" rid="PARAGRAPH.17.0"/><bx clone="yes" id="oPHRASE.19.0" rid="PHRASE.19.0" cba:fontname="Wingdings" cba:fontstyle="8"/>><ex id="cPHRASE.19.0" rid="PHRASE.19.0"/><bx clone="yes" id="oPHRASE.20.0" rid="PHRASE.20.0" cba:bold="true" cba:fontname="Times New Roman" cba:italic="true"/>Courses<ex id="cPHRASE.20.0" rid="PHRASE.20.0"/><ex id="cPARAGRAPH.17.0" rid="PARAGRAPH.17.0"/><bx clone="no" id="oPARAGRAPH.18.0" rid="PARAGRAPH.18.0"/><bx clone="yes" id="oPHRASE.21.0" rid="PHRASE.21.0" cba:fontname="Times New Roman" cba:fontstyle="9"/>A leading opening tag appearing before the beginning of a sentence and its corresponding closing tag appearing after the end of the sentence are an example of suprasentential codes:<ex id="cPHRASE.21.0" rid="PHRASE.21.0"/><ex id="cPARAGRAPH.18.0" rid="PARAGRAPH.18.0"/><bx clone="no" id="oPARAGRAPH.19.0" rid="PARAGRAPH.19.0"/><bx clone="yes" id="oPHRASE.22.0" rid="PHRASE.22.0" cba:fontstyle="4"/><ex id="cPHRASE.22.0" rid="PHRASE.22.0"/><ex id="cPARAGRAPH.19.0" rid="PARAGRAPH.19.0"/><bx clone="no" id="oPARAGRAPH.20.0" rid="PARAGRAPH.20.0"/><bx clone="yes" id="oPHRASE.23.0" rid="PHRASE.23.0" cba:fontname="Wingdings" cba:fontstyle="8"/>>><ex id="cPHRASE.23.0" rid="PHRASE.23.0"/><bx clone="yes" id="oPHRASE.24.0" rid="PHRASE.24.0" cba:bold="true" cba:fontname="Times New Roman" cba:italic="true"/>They don't need to appear in the translation editor and the translator does not need to see them:<ex id="cPHRASE.24.0" rid="PHRASE.24.0"/><ex id="cPARAGRAPH.20.0" rid="PARAGRAPH.20.0"/><bx clone="no" id="oPARAGRAPH.21.0" rid="PARAGRAPH.21.0"/><bx clone="yes" id="oPHRASE.25.0" rid="PHRASE.25.0" cba:fontname="Times New Roman" cba:fontstyle="9"/>On the other hand, the file markup_span.html from the sample project includes the following text:<ex id="cBLOCK_FRAGMENT.0.T11" rid="BLOCK_FRAGMENT.0.T11"/><ex id="cPHRASE.25.0" rid="PHRASE.25.0"/><ex id="cPARAGRAPH.21.0" rid="PARAGRAPH.21.0"/><bx clone="no" id="oPARAGRAPH.22.0" rid="PARAGRAPH.22.0"/><bx clone="yes" id="oPHRASE.26.0" rid="PHRASE.26.0" cba:fontstyle="4"/><ex id="cPHRASE.26.0" rid="PHRASE.26.0"/><ex id="cPARAGRAPH.22.0" rid="PARAGRAPH.22.0"/><bx clone="no" id="oPARAGRAPH.23.0" rid="PARAGRAPH.23.0"/><bx clone="yes" id="oPHRASE.27.0" rid="PHRASE.27.0" cba:fontname="Wingdings" cba:fontstyle="8"/>><ex id="cPHRASE.27.0" rid="PHRASE.27.0"/><bx clone="yes" id="oBLOCK_FRAGMENT.0.T8" rid="BLOCK_FRAGMENT.0.T8" cba:blockname="T8"/><bx clone="yes" id="oPHRASE.28.0" rid="PHRASE.28.0" cba:bold="true" cba:fontname="Times New Roman" cba:italic="true"/>Mediation<ex id="cBLOCK_FRAGMENT.0.T8" rid="BLOCK_FRAGMENT.0.T8"/&



PROJECT EXECUTION

LANGUAGE ASSETS

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  Standard: .xliff

Termbase  Standard: .tbx

Guidelines

Translation memory from previous cycle
 Standard: .tmx

Reference files:

- Related research

LINGUISTIC QUALITY ASSURANCE: GLOSSARIES

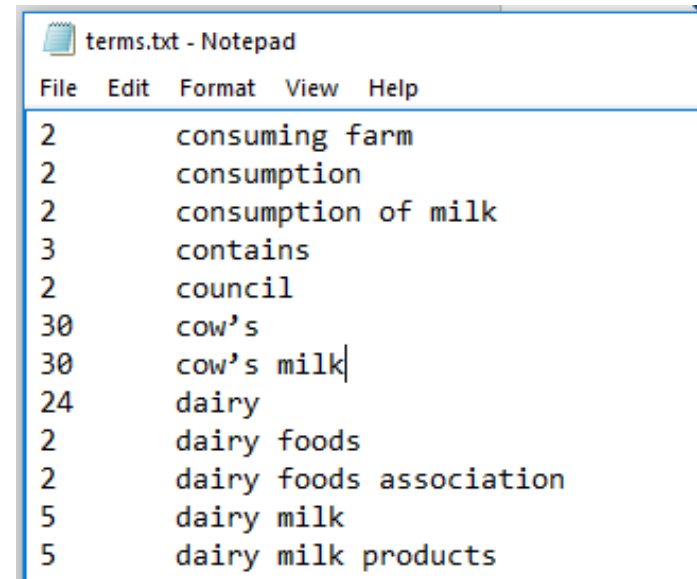
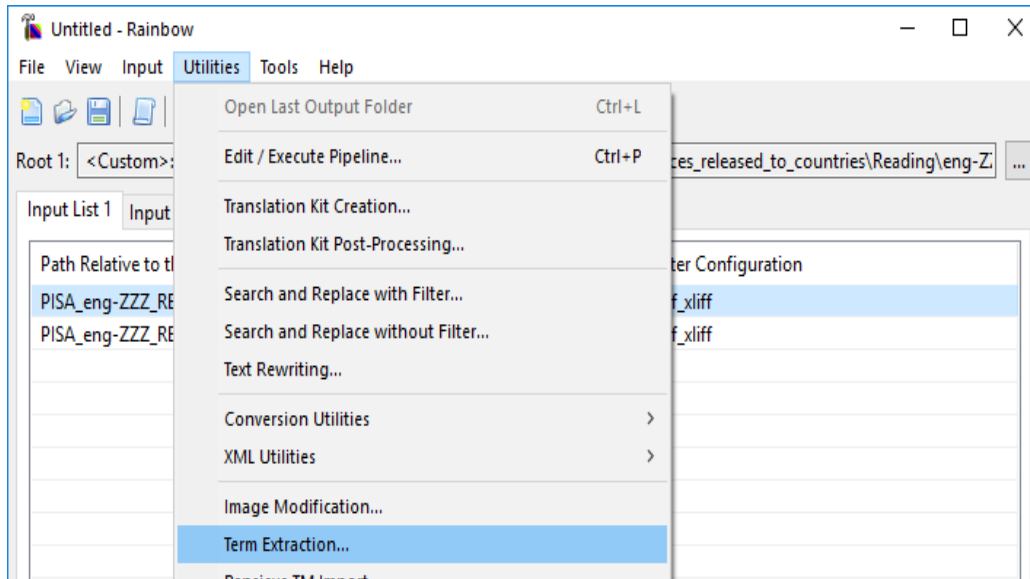
Source Text

Automated term extraction

Term validation

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).





-- select -- ▼	Metadata	SEARCH
English ▼	click on X	
fra-ZZZ ▼	Target expression	

Container	Content type	Domain
TEST ▼	All ▼	All ▼

Advanced search options i

Source:

+w	[abc]	Aa	≠	[...]	.+	?*
----	-------	----	---	-------	----	----

Target:

+w	[abc]	Aa	≠	[...]	.+	?*
----	-------	----	---	-------	----	----

► Search log

#	file	eng	fra-ZZZ
1	TEST_Pisa_Cognitive_Recurring_Glossary	Click on a choice and then type an explanation to answer the question.	Pour répondre à la question, cliquez sur l'un des choix de réponse puis tapez votre explication.
2	TEST_Pisa_Cognitive_Recurring_Glossary	Click on the NEXT arrow to continue.	Pour continuer, cliquez sur la flèche « SUIVANT ».
3	TEST_Pisa_Cognitive_Recurring_Glossary	Click on the choices in the table to answer the question.	Pour répondre à la question, cliquez sur vos choix de réponse dans le tableau.

ISSUES: TYPO

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

Height record for hot air balloons

The Indian pilot Vijaypat Singhania beat the height record for hot air balloons on November 26, 2005. He was the first person to fly a balloon 21,000 metres above sea level.

Record height: 21,000 m

Earlier record: 19,800 m

Oxygen: only 4% of what is available at ground level

Side slits can be opened to let out hot air for descent.

Size of conventional hot air balloon

Height: 49 m

Fabric: Nylon

Inflation: 2.5 hours

Size: 453,000 m³ (normal hot air balloon 481 m³)

Weight: 1800 kg

Gondola:
Height: 2.7 m Width: 1.3 m

Enclosed pressure cabin with insulated windows

Aluminium construction, like airplanes

Vijaypat Singhania wore a space suit during the trip.

Question 34: BALLOON

Vijaypat Singhania used technologies found in two other types of transport. Which types of transport?

1.

2.

Full Credit

Code 2: Refers to **BOTH airplanes AND spacecraft** (in either order). *[can include both answers on one line]*

- 1. Aircraft
- 2. Spacecraft

New Delhi

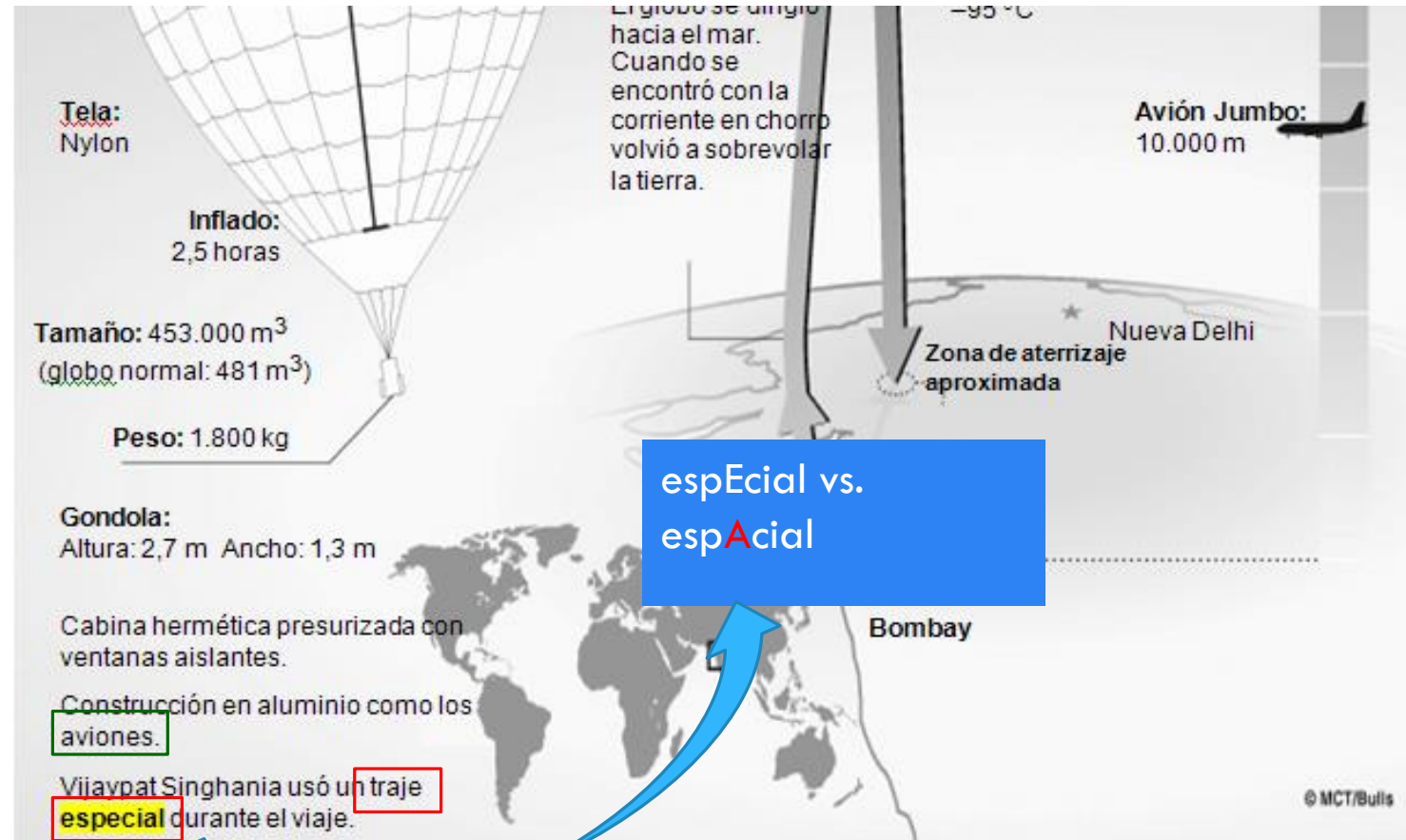
Mumbai

© MCT/Bullis

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.

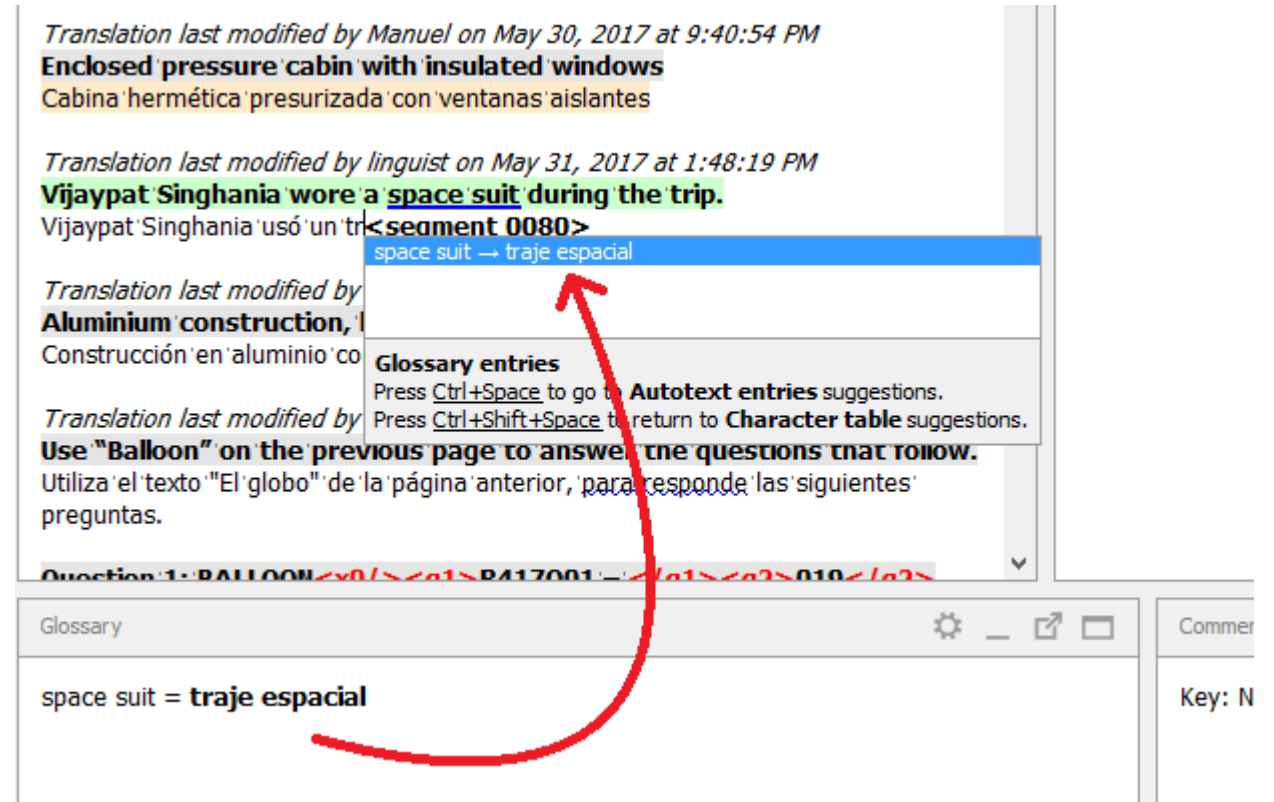


TRANSLATION: COMPUTER ASSISTED TRANSLATION TOOLS

Bilingual glossaries and key term definitions can be imported in CAT tools.

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

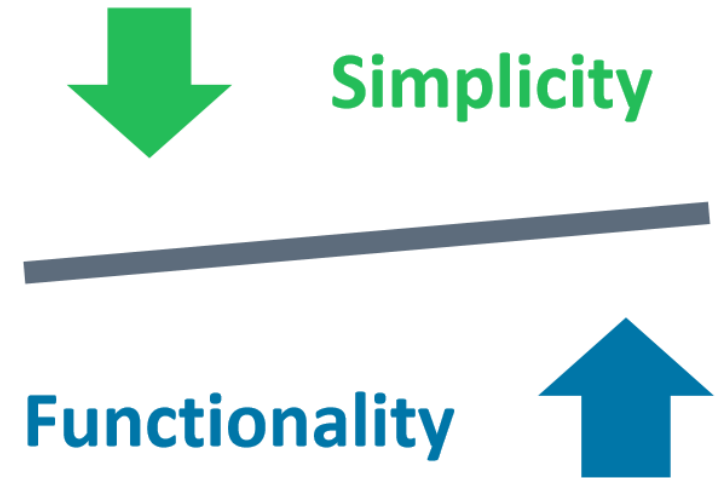
Predictive typing



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.



TRANSLATION: OMEGAT

Color coding for navigation
between segments

Glossary pane

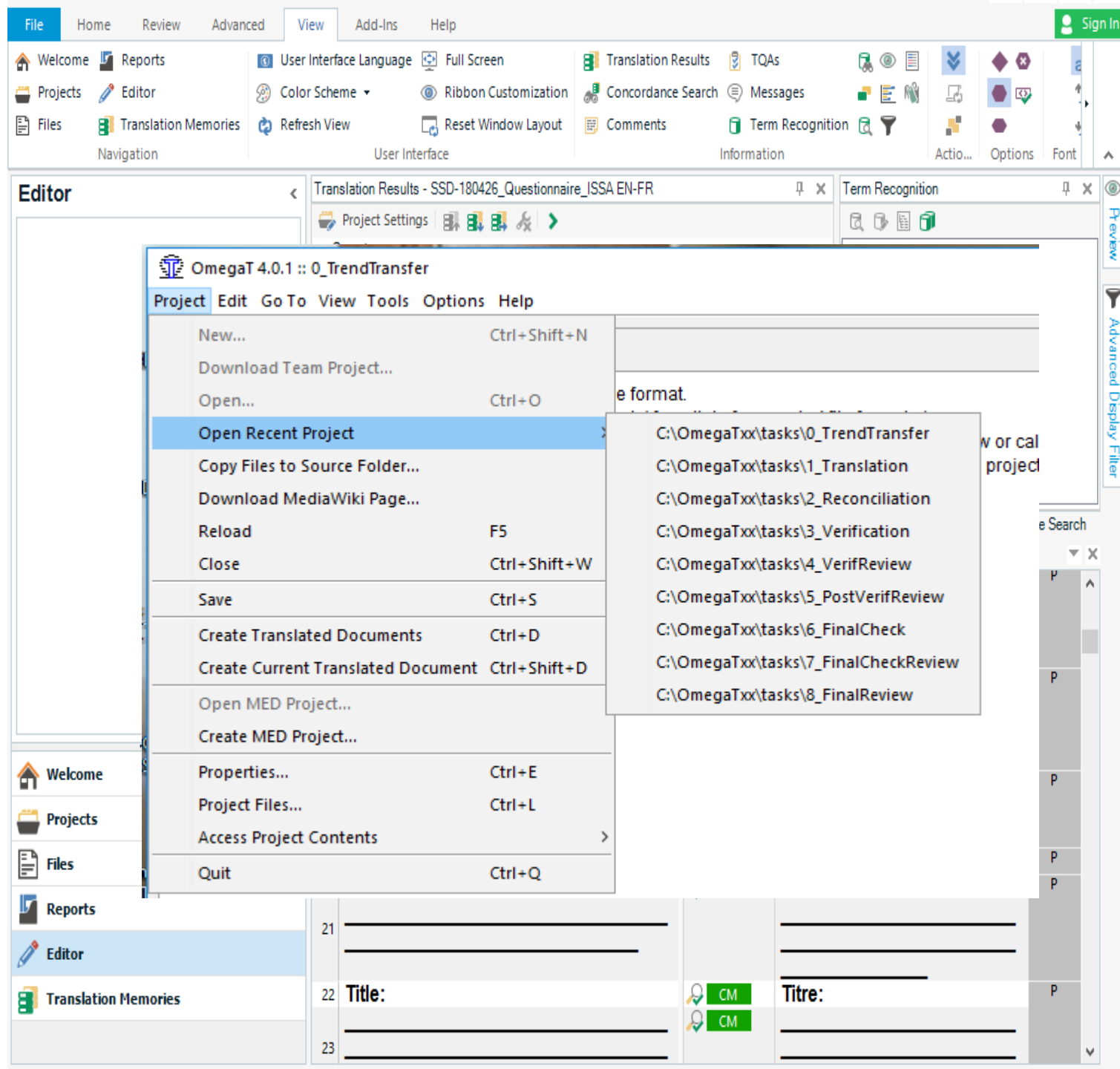
Underlined terms

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

neither satisfied nor dissatisfied

Dissatisfied

Insatisfecho

Very dissatisfied

Would you like to participate to a similar event?

Yes

No

Don't know

If yes, how often?

Very often

Often

Rarely

Never

How satisfied are you with your accommodation in Barcelona?

Very satisfied

Satisfied

Dissatisfied

Insatisfecho

CREATING TRANSLATION MEMORIES

How?

- Create translation memories through alignment



Satisfaction_quest
t_Madrid_EN



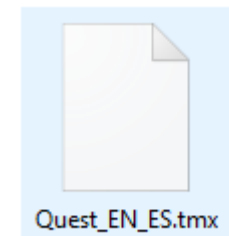
Satisfaction_quest
t_Madrid_ES



LF Alignment Editor 1.5 - aligned_Questionnaire_example_for_alignment_EN-Questionnaire_example_f...

1	How satisfied are you with your trip to Madrid?	¿Cuán satisfecho está con su viaje a Madrid?
2	Very satisfied	Muy satisfecho
3	Satisfied	Satisfecho
4	Neither satisfied nor dissatisfied	Ni satisfecho ni insatisfecho
5	Dissatisfied	Insatisfecho
6	Very dissatisfied	Muy insatisfecho
7	Would you like to participate to a similar conference?	¿Le gustaría participar en una conferencia parecida?
8	Yes	Sí
9	No	No
10	Don't know	No sé
11	How often?	¿Con qué frecuencia?
12	Very often	Muy a menudo
13	Often	A menudo

Merge (F1) Split (F2) Shift up (F3) Shift down (F4)



Quest_EN_ES.tmx

TRANSLATION MEMORIES

Hierarchy

More reliable versions: how can a distinction be made?

- Applying penalties to less reliable translation memories

Very dissatisfied
<segment 0006>

Fuzzy Matches

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

Trend management

How can trend content be visible, recognizable and protected?

- Distinction: color coding
- Protection: locked segments

<i> Refer to "Farm to Market Dairy" on the right.

Click on a choice to answer the question. </i>

Pour répondre à la question, cliquez sur l'un des choix de réponse

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

<segment 0222>

Consuming milk and milk products leads to obesity.

Milk is a good source for essential vitamins and minerals.



LINGUISTIC QUALITY CONTROL

LQA spells out the requirements.

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

Issues (5)

Options

All (5)

Spelling (3)

Tags (2)

Segment	Type	Description
166	Spelling	www.lafermelaitiere.com
171	Spelling	d'innonmbrables
217	Tags	Missing x2
221	Tags	Missing x1
223	Spelling	consomation

Cow's Milk

Le lait de vache

Le lait de vache

Apply Fix

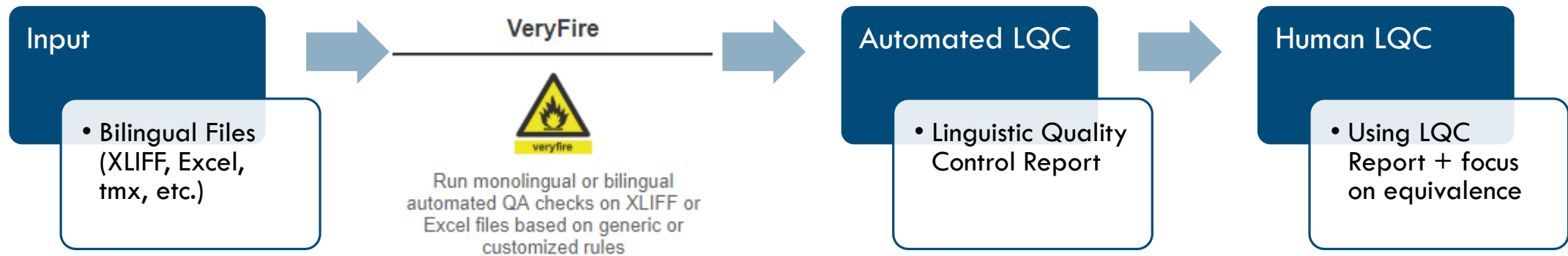
Jump To Segment

Refresh

Close

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.

Rulesets	
Scope:	
<input checked="" type="checkbox"/> Project	<input type="checkbox"/>
<input checked="" type="checkbox"/> Generic	<input type="checkbox"/>
<input checked="" type="checkbox"/> Language	<input type="checkbox"/>
<input checked="" type="checkbox"/> xxx-XXX	<input type="checkbox"/>

Select rules applying to all target languages.

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

Error	Row	Ruleset:Rule ID	Function:Pattern	File name	Source segment	Target segment	Comment
1	20	Generic checks	Unaltered segment	TEST_Cows_Milk_fra-FRA_1_Translation.xls	To support the use of <i>Farm to Market Dairy</i> products.	To support the use of <i>Farm to Market Dairy</i> products.	The text in the translation is identical to the source text.
2	1	Generic checks	Target segment inconsistency	TEST_Cows_Milk_fra-FRA_1_Translation.xls	Farm to Market	La Ferme laitière	Identical source segments have been translated differently.
2	2	Generic checks	Target segment inconsistency	TEST_Cows_Milk_fra-FRA_1_Translation.xls	Farm to Market	La Ferme de lait	Identical source segments have been translated differently.
3	13	Generic checks	Source segment inconsistency	TEST_Cows_Milk_fra-FRA_1_Translation.xls	Click on Yes choice No to answer the question.	Pour répondre à la question, cliquez sur l'un des choix de réponse.	Different source segments have been translated identically.
3	15	Generic checks	Source segment inconsistency	TEST_Cows_Milk_fra-FRA_1_Translation.xls	Click on Yes or No to answer the question.	Pour répondre à la question, cliquez sur l'un des choix de réponse.	Different source segments have been translated identically.
4	11	RuleSet_TEST_testing_fra-FRA_DL:2	Glossary: 'dairy products' → 'produits laitiers'	TEST_Cows_Milk_fra-FRA_1_Translation.xls	"Milk proteins, milk sugar, fat, and saturated fat in dairy products pose health risks for children and encourage the development of obesity, diabetes, and heart disease."	«Les protéines de lait, le sucre de lait, les graisses et les graisses saturées des produits posent des risques pour la santé des enfants et favorisent le développement de l'obésité, du diabète et des maladies cardiaques».	Translation not compliant with glossary.



CONCLUSIONS

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

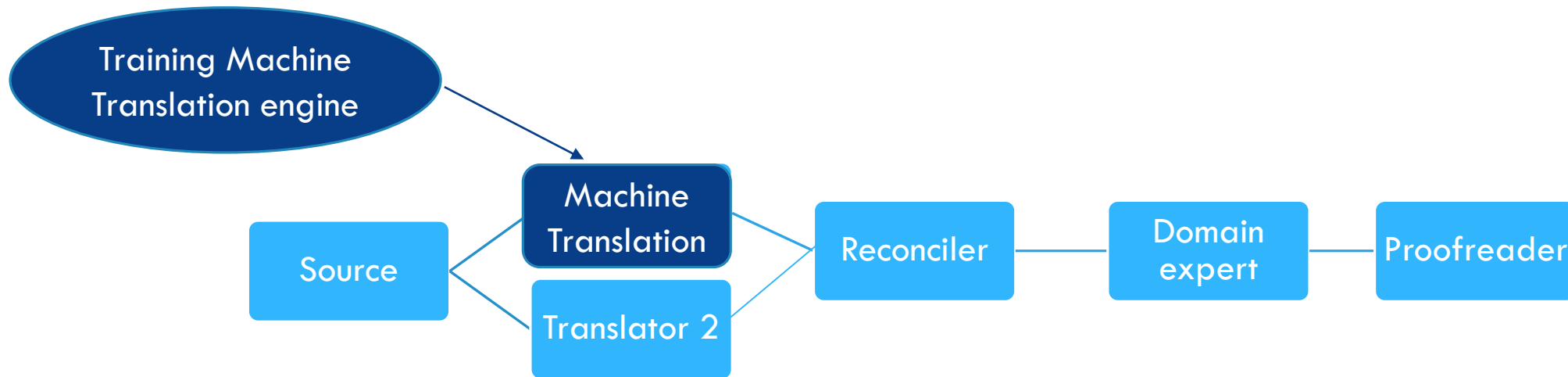
Optimal processes require planning and preparation

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

And all players will benefit from translation technology

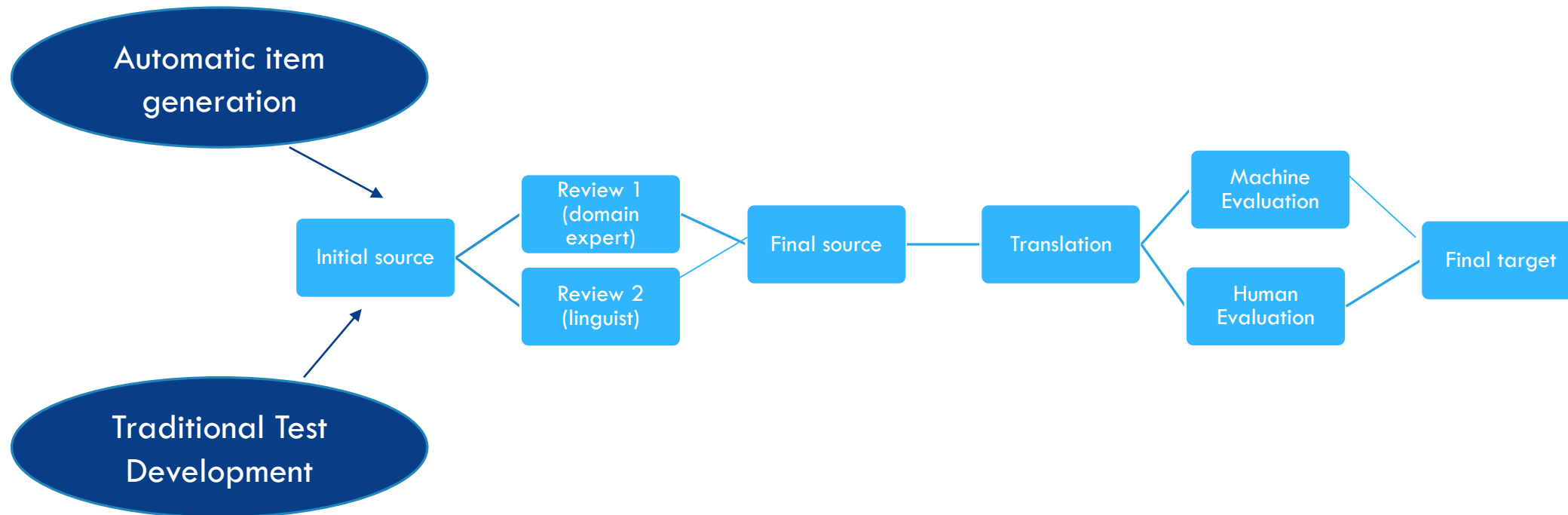
TRANSLATION TECHNOLOGY: WHAT'S NEXT?

- Incorporating Machine Translation (MT) in current workflows?



TRANSLATION TECHNOLOGY: WHAT'S NEXT?

- Creating new workflows that allow better use of new technologies?






THANK YOU FOR YOUR ATTENTION!

QUESTIONS?

TRANSLATION: SERVER BASED REPOSITORY OF TMs



-- select --
▼

English

fra-ZZZ
▼

Metadata
SEARCH

of Xandar
X

Target expression

Container
▼

PISA

Content type
▼

Translations

Domain
▼

All

Advanced search options
i

Source:

+w

[abc]

Aa

≠

[...]

.+

?*

Target:

+w

[abc]

Aa

≠

[...]

.+

?*

► Search log

#	file	eng	fra-ZZZ
1	PISA_fra-ZZZ_CPS_C100-Xandar_eng_1-Init.tmx <small>commonFragments_1_tab1Q4_519415e7c79ab1.75913225</small>	What proportion of Xandar is desert?	Quelle proportion du Xandar est occupée par le désert ?
2	PISA_fra-ZZZ_CPS_C100-Xandar_eng_1-Init.tmx <small>commonFragments_1_commonFragments_14_519415e7c7b4a3.32164229</small>	10 percent of Xandar is desert.	10 pourcents du Xandar sont occupés par le désert.
3	PISA_fra-ZZZ_CPS_C100-Xandar_eng_1-Init.tmx <small>commonFragments_1_commonFragments_18_519415e7c7c3c9.37240833</small>	The climate of Xandar is mostly tropical.	Le climat du Xandar est essentiellement tropical.