



**Machine Translation Enhanced  
Computer Assisted Translation**

---

## **D4.2 - Second version of MateCat Tool**

---

**Author:** Alessandro Cattelan

**Dissemination level:** Public

**Date:** October 15<sup>th</sup>, 2013

**D4.2 - Second version of MateCat Tool**

Grant agreement no.	287688
Project acronym	MateCat
Project full title	Machine Translation Enhanced Computer Assisted Translation
Funding scheme	Collaborative project
Coordinator	Marcello Federico (FBK)
Start date, duration	November 1st 2011, 36 months
Dissemination level	Public
Contractual date of delivery	September 30th, 2013 (M23)
Actual date of delivery	October 15th, 2013
Deliverable number	D4.2
Deliverable title	Second version of MateCat Tool
Type	Report
Status and version	Final, V1.0
Number of pages	13
Contributing partners	Translated
WP leader	Translated
Task leader	Translated
Author	Alessandro Cattelan
Reviewers	Marco Trombetti, Marcello Federico
EC project officer	Kimmo Rossi
MateCat partners are:	Fondazione Bruno Kessler (FBK), Italy
	Université Le Mans (LE MANS), France
	The University of Edinburgh (UEDIN), UK
	Translated S.r.l. (TRANSLATED), Italy

For copies of reports, updates on project activities and other MateCat-related information, contact:

FBK

MateCat

Marcello Federico

Povo - Via Sommarive 18

I-38123 Trento, Italy

federico@fbk.eu

Phone: +39 0461 314 552

Fax: +39 0461 314 591

Copies of reports and other material can also be accessed via <http://www.matecat.com>

© 2013, Alessandro Cattelan

No part of this document may be reproduced or transmitted in any form, or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission from the copyright owner.

## **Executive Summary**

During the second year of the project, the MateCat tool has been further developed and has been used in production by the commercial partner. While still in early development, it currently offers all of the functionalities to perform translation or post-editing work using commercial or open source MT servers. The CAT tool functionalities have been enhanced both in terms of usability for the translators and of information provided to the MT researchers/developers.

This document presents the main features of the CAT tool at the current development stage.

# Table of Contents

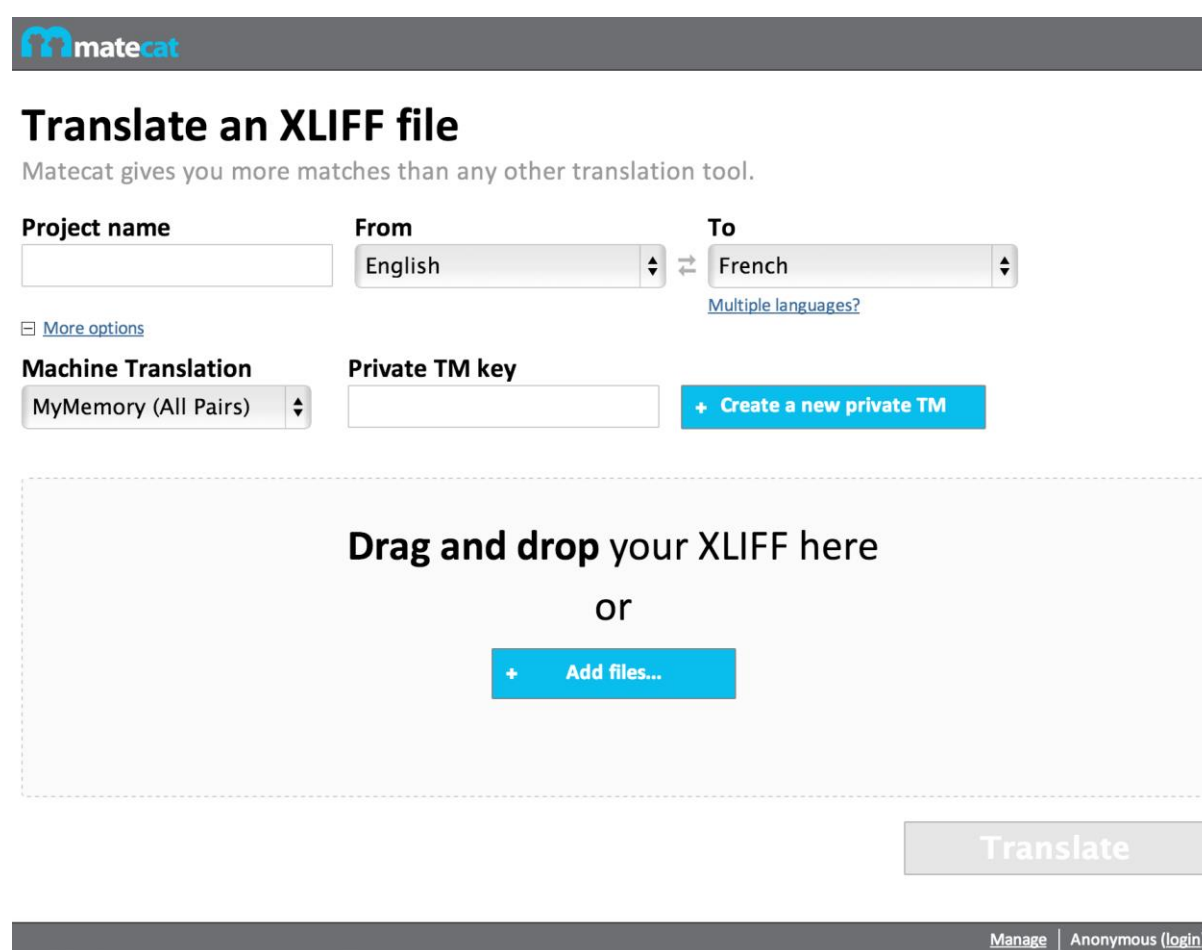
Executive Summary.....	3
1. File importer .....	5
2. Management panel.....	6
3. Editor .....	7
3.1 Understanding the URL.....	7
3.2 Understanding the UI.....	7
3.2.1 Top bar.....	8
3.2.2 Active segment .....	9
3.2.3 Bottom bar .....	10
3.2.4 Navigating through files .....	10
3.2.5 Using TM matches / MT suggestions .....	10
3.2.6 Quality Assurance Messages.....	11
4. Editing log .....	12
5. CAT Tool URL.....	13

## 1. File importer

The first component of the CAT tool is the file importer available at the following URL:

<http://demo.matecat.com>

Users can translate any XLIFF files in over 70 languages. Also, by clicking on the *More options* link, they can select the machine translation server to use and create or enter their own private translation memory key.



The screenshot shows the MateCat web interface. At the top is the MateCat logo. Below it is the heading "Translate an XLIFF file" with a subtext "Matecat gives you more matches than any other translation tool." The interface includes several input fields and buttons:

- Project name:** A text input field.
- From:** A dropdown menu currently set to "English".
- To:** A dropdown menu currently set to "French". A double-headed arrow is between the "From" and "To" menus. A link "Multiple languages?" is below the "To" menu.
- More options:** A link with a small icon to its left.
- Machine Translation:** A dropdown menu currently set to "MyMemory (All Pairs)".
- Private TM key:** A text input field.
- Create a new private TM:** A blue button with a plus icon.
- Drag and drop your XLIFF here or:** A large dashed box containing this text and a blue button with a plus icon and the text "Add files..."
- Translate:** A large grey button.

At the bottom right, there are links for "Manage" and "Anonymous (login)".

## 2. Management panel

When creating a project, users are given the option to log in to the MateCat tool to add their projects to the management panel as displayed below.

The screenshot shows the MateCat web interface. At the top, the logo 'matecat' is on the left, and 'Showing active projects' is in the center. On the right, there is a dropdown menu with a downward arrow. Below the header, the title 'Projects List' is displayed. Underneath, the project name 'Connection\_en-GB\_fr-FR.xliff' is shown with four small icons (trash, folder, resume, unarchive) to its left. Below the project name, it says 'Machine Translation: MyMemory (All Pairs)'. A table follows with the following structure:

Create Date	Job	Progress	Actions
Today, 12:37	English > French <a href="http://demo.matecat.com/translate/.../7497-2rva9trz">http://demo.matecat.com/translate/.../7497-2rva9trz</a>	<div style="width: 100%;"></div>	

MATECAT BBOL-201310110245

*Management panel*

From the management panel, users can access the translator editor by clicking on the URL of the translation job, check the progress of their translations and perform a set of actions, namely:

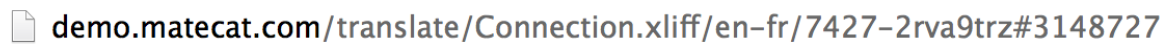
	<b>Change job password</b> - This will modify the password of the job.
	<b>Cancel job/project</b> - This will delete the job/project and disable the URL.
	<b>Archive Job/project</b> - This will prevent any changes to be applied to the job/project. However, the URL will still work and is accessible in read-only mode.
	<b>Resume project</b>
	<b>Unarchive project</b>

## 3. Editor

The editor is the main component of the MateCat tool where translators perform their work.

### 3.1 Understanding the URL

The MateCat URL contains information about the job to translate. An explanation of the different elements of the URL can be found below.

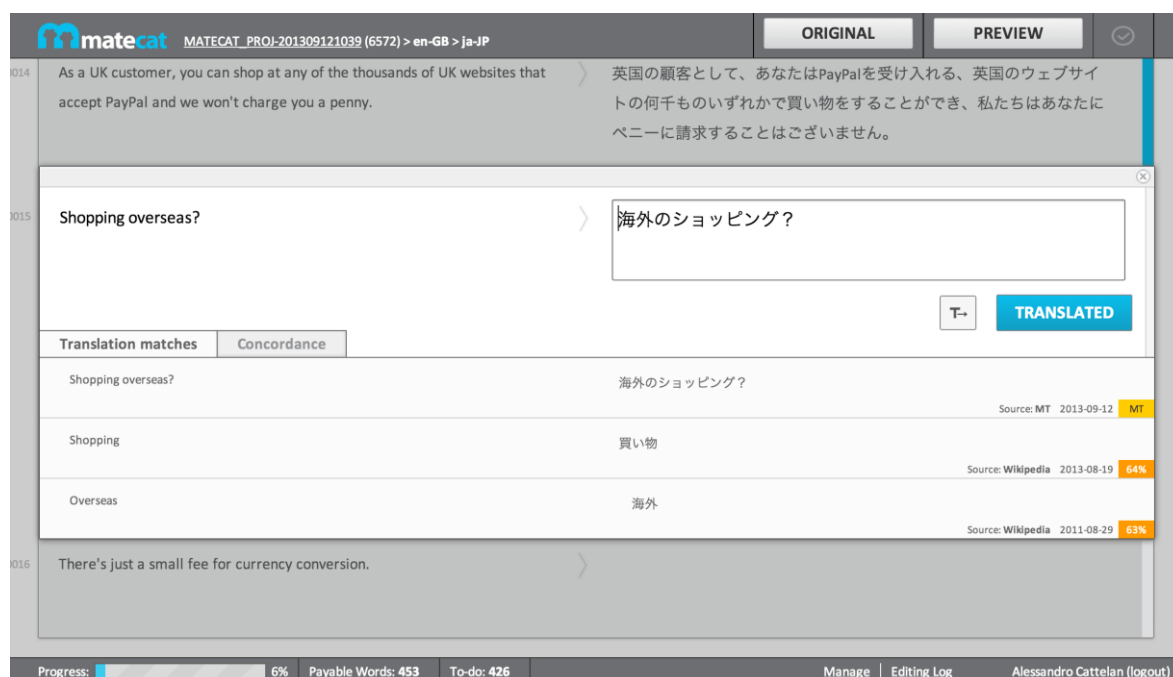





demo.matecat.com/translate/Connection.xliff/en-fr/7427-2rva9trz#3148727

- **http://demo.matecat.com** - Base URL for the MateCat tool.
- **/translate** - This indicates the translation editor page.
- **/Connection.xliff** - The name of the project to translate.
- **/en-fr** - The language pair of the translation.
- **/7427-2rva9trz** - The job ID (7427) followed by the job password (unique for any jobs).
- **#3148727** - The unique identifier of the active segment. This number changes every time a new segment is opened.

### 3.2 Understanding the UI

The following image presents the translation editor of the MateCat tool. In the following sections, a description of the main component of the editor are presented.

**D4.2 - Second version of MateCat Tool***Translation editor***3.2.1 Top bar**

	Clicking on the logo takes the users to the home page of the MateCat tool where they can create a new project.
<b>MATECAT_PROJ-201309121039 (6572) &gt; en-GB &gt; ja-JP</b>	The name of the job and language pair combination.
<b>ORIGINAL</b>	Download the original file(s).
<b>PREVIEW</b>	Download a preview of the translated file(s). The preview contains the translation and machine translation for the segments that have not been edited.
<b>DOWNLOAD TRANSLATION</b>	The <i>Preview</i> button is replaced by the <i>Download translation</i> button once the translation is complete (100% in the progress bar).
	QA module - No errors reported.
	QA module - 1 error message reported. Click on the icon to go to the segment with the error to fix it.



**D4.2 - Second version of MateCat Tool****3.2.2 Active segment***The active segment*

The active segment is the segment the user is editing. Users enter their translation or edit the pre-translated segment in the input field to the right.

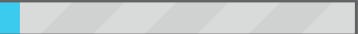
Once users are done editing the segment, they click on the *Translated* button to confirm their translation and move to the next segment. A number of keyboard shortcuts are available which are displayed by hovering the cursor over the buttons.

The *Translation matches* and *Concordance* tabs provide suggestions from machine translation or translation memory. To the right, the tool presents information about the origin of the match, the date it was created and the match percentage.

The bar to the right of each segment indicates its status using colors. Clicking on it shows the following options to set/change the status of a segment:

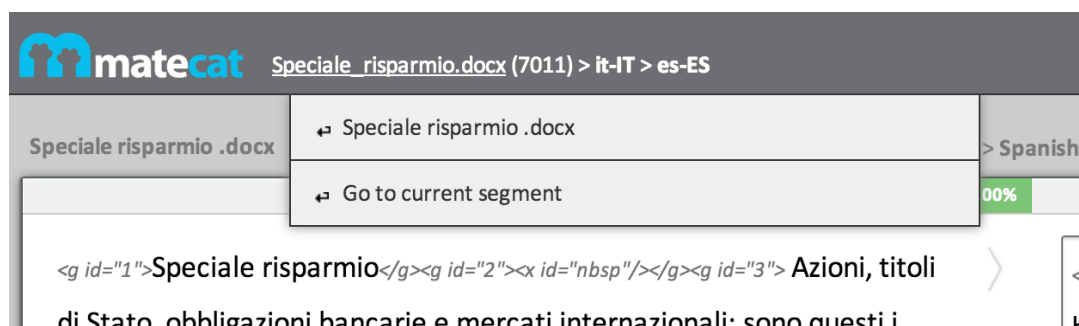
*Status dialog*

**D4.2 - Second version of MateCat Tool****3.2.3 Bottom bar**

Progress:  6%	The progress bar indicates the real time status of the translation.
Payable Words: 453	Number of payable (equivalent) words for the entire job.
To-do: 426	Number of words left to translate for the entire job.

**3.2.4 Navigating through files**

Clicking on the job name at the top of the page will display a list of the files of the job. From there, users can click on the file name to go to the beginning of that file or to move to the active segment by clicking on *Go to current segment*.

*Navigation menu***3.2.5 Using TM matches / MT suggestions***MT suggestion*

Hovering the cursor over the matches displays the shortcut to select the match and insert it into the translation input field. On Mac OS X, the shortcut is *ALT+CMD+[n]* as shown above.

**D4.2 - Second version of MateCat Tool**

Users can also remove incorrect matches from the TM by clicking on the dustbin icon



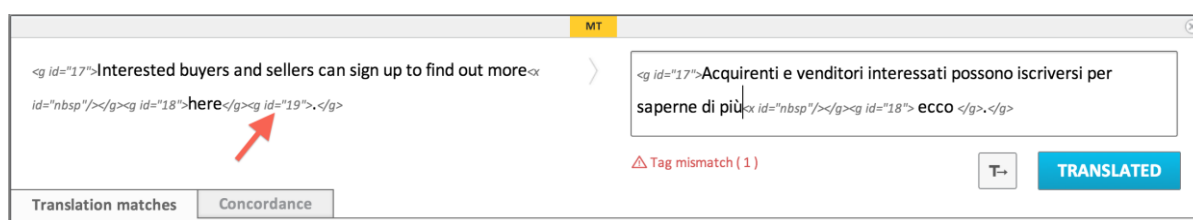
that is displayed to the right of each segment when hovering the cursor over it.

**3.2.6 Quality Assurance Messages**

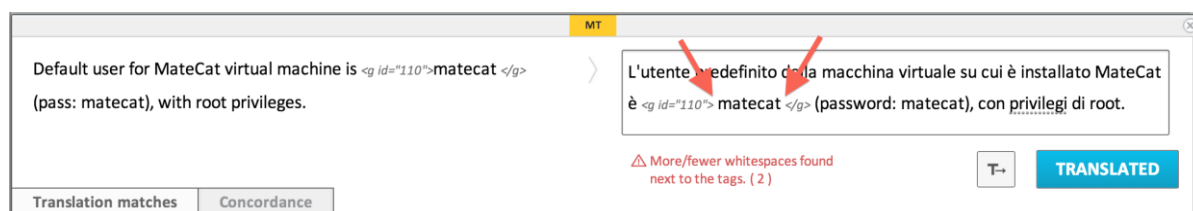
At the top right of the MateCat window, you will find an icon which warns of any potential QA issues in the translation. The warning sign contains a red box indicating how many issues are found in the project. The following example indicates that the project contains one QA issue:



Clicking on the icon takes the user to the segment with the issue to fix it.

**Examples of QA issues****Tag Mismatch**

The tag `<g id="19">` in the source segment, indicated by the red arrow, was not added to the target segment. This generates a tag mismatch. To fix the issue, users can copy and paste the tag from the source segment to the target and place it in the same position as the source. Alternatively, they can click on the tag in the source segment and drag and drop it to the target segment.


**Whitespaces mismatch**

MateCat warns the user when more or fewer spaces are found in the target segment next to the tags. In the image above, there are two extra whitespaces indicated by the red arrows. To fix the issue, users delete the extra spaces between the tags and the word.

## 4. Editing log

The editing log displays statistics that can be used to check the performance of the translators and to assess the quality of the MT server in terms of time-to edit, post-editing effort and HTER.

Statistics are logged in real time and are available both in a web UI and in a CSV file which can downloaded from the same page.

 Dry_Run_-_Field_test_2 (7450) > en-GB > fr-FR <div> <a href="#">&lt; Back to Translation</a> <a href="#">Export All Data in CSV</a> </div>								
Secs/Word	Job ID	Segment ID	Words	Suggestion source	Match percentage	Time-to-edit	PE Effort	TER (Beta)
26.8	7450	<a href="#">3128589</a>	18.00	Machine Translation	86%	08m:01s	56%	48%
Segment	Connection Timeout (in minutes) – Enter the number of minutes after which the connection should be timed out.							
Suggestion	Connexion Délai d'attente (en minutes) - Saisissez le nombre de minutes après lequel la connexion doit être dépassé.							
Translation	Connection Timeout (in minutes) – Introduisez le nombre de minutes au bout desquelles la connexion doit être interrompue.							
Diff View	<del>Connexion Délai d'attente (en</del> Connection Timeout (in minutes) <del>–Saisissez – Introduisez</del> le nombre de minutes <del>après lequel au bout</del> desquelles la connexion doit être <del>dépassé: interrompue.</del>							

Secs/Word	Job ID	Segment ID	Words	Suggestion source	Match percentage	Time-to-edit	PE Effort	TER (Beta)
9.6	7450	<a href="#">3128597</a>	46.00	Machine Translation	86%	07m:19s	60%	58%
Segment	When this option is set to “Yes” for a connection, a check box for Deploy to TC is displayed on “Save definition” dialog and another check box for TC Mobility is displayed on both Schedule Now popup and Delivery Options of Output tab in Scheduler component.							
Suggestion	Lorsque cette option est réglé sur "Oui" pour une connexion, une case pour Déployer à TC est affichée sur "Enregistrer" de définition de dialogue et une autre case pour TC Mobilité est affichée sur les deux Planifier Maintenant popup et Delivery Options d'onglet Sortie dans composant Scheduler.							
Translation	Lorsque cette option est réglée sur « Yes » pour une connexion, une case Deploy to TC s'affiche sur la boîte de dialogue « Save definition » et une autre case TC Mobility s'affiche sur la fenêtre contextuelle Schedule Now et sur Delivery Options, dans l'onglet Output de la composante Scheduler.							
Diff View	Lorsque cette option est <del>réglé</del> <del>réglée</del> sur <del>"Oui"</del> « Yes » pour une connexion, une case <del>pour Déployer à</del> <del>Deploy to</del> TC est <del>affichée</del> s'affiche sur <del>"Enregistrer" de définition</del> la boîte de dialogue <del>« Save definition »</del> et une autre case <del>pour TC Mobilité est affichée</del> Mobility s'affiche sur <del>les deux Planifier Maintenant popup</del> la fenêtre contextuelle Schedule Now et sur Delivery <del>Options d'onglet Sortie</del> Options, dans							

Progress:  0% Total Words: 1,207 To-do: 0

[Manage](#) | [Alessandro Cattelan \(logout\)](#)

*The editing log page*

## **5. CAT Tool URL**

The second release of the CAT tool can be found at the following URL:

<http://demo.matecat.com>

The repository where the source code is stored can be found at the following URL:

[https://www.assembla.com/code/matecat\\_source/git/nodes](https://www.assembla.com/code/matecat_source/git/nodes)