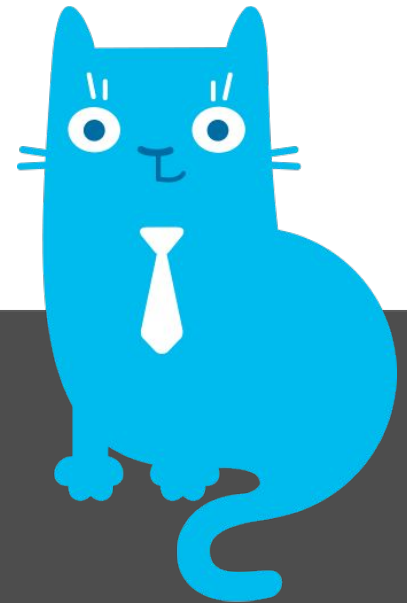


Introduction to MateCat

Luca De Franceschi - Associate Product Manager

Le Mans, 12 June 2018



A little bit of history

1940s and **'50s**: Machine Translation is all the rage

1960s: Automatic Language Processing Advisory Committee (ALPAC) report resulted in a halt to public financing for MT

1970s: Focus on helping the human translators

1980s: TSS (Translation Support System) by Alpnet, TextTools by INK, Trados is born (1984)

1990s and **2000s**: Multiplication of translation memory tools: Trados, Wordfast, DéjàVu, etc.

2000s: Return of MT: Statistical Machine Translation, Moses, Google Translate, etc.

2014 - 16: Neural MT: Google, DeepL, Microsoft, etc.

Translation Market

\$40 billion in translation services

Large industry and growing (6.2% CAGR '14).

98.7% of professional services and 1.3% of translation software.

40,000 translation companies (LSPs)

Fragmented: only 100 companies above \$100M.

Median LSP \$500K/y

3-5 employees, 20 freelancers for only few languages.

Language Service Providers

Using In-House Translators

The oldest and/or industry-specialized ones that mostly use in-house translators. They **outsource peaks and non-supported languages to other LSPs (10%)** and freelancers (10%).

Using Freelance Translators

The large majority of LSPs today. They outsource a big chunk of their business to freelancers (80%) and **a smaller amount (20%) to other LSPs**.

Outsourcing to other LSPs

Most of the large LSP, historically called MLV, are sales-driven companies. They are often the result of merging many LSPs under one brand. They **primarily outsource to other LSPs**.

Common Pains

Can't afford to say no to customers

The competition is fierce – **you say no once, you risk losing your customer.**
Today when an LSP has a request that it cannot manage, it picks up the phone and calls freelancers and LSPs until one is available.

Outsourcing = No margins

Requests come fast and delivery terms are often less than 48 hours. When in rush, **LSPs forget about their margins to find an outsourcer quickly.**

Peaks vs. Quality

Median LSPs have few translators per language. Finding a freelancer or LSP experienced in a specific field and/or with a specific customer is difficult.
Controlling quality is an issue.

What is a CAT tool?

Computer Assisted Translation tools:

- Spell checkers
- Grammar checkers
- Terminology managers
- Electronic dictionaries
- Concordancers
- Alignment tools
- Project management tools
- Translation memory

CAT tool lingo

- Segments
- Alignment
- Concordancer (TM Search)
- 100% match - Exact match
- XX% match - Fuzzy match
- Bilingual file (.xliff)
- TM (public and private)
- Post - editing
- Revision
- Repetitions
- Internal repetitions
- Tag
- TMX (Translation Memory eXchange format)

Translation Memory

Overview

- “Database” of translated **segments** in a given language pair
- If any segment was translated in the past, it will be returned as an exact or fuzzy match (re-use of content)
- A CAT automatically suggests the best **match** for a given segment
- Online TM (MyMemory) and offline TM (TMX)

Translation Memory

```
<tu tuid = "515933822" srclang="en-US" creationdate="20160908T123213" changedate="
20160908T123213" datatype = "plaintext">
  <prop type="x-project_id">427324</prop>
  <prop type="x-project_name">Exporting_Private_Translation_Memories.docx</prop>
  <prop type="x-job_id">467988</prop>
    <tuv xml:lang = "en-US">
      <seg>When the export is ready, an email will be sent to your email
      address with a link to download the translation memory.</seg>
    </tuv>
    <tuv xml:lang = "it-IT">
      <seg>Quando l'esportazione è pronta, verrà inviata una e-mail al tuo
      indirizzo e-mail con un link per scaricare la memoria di traduzione.</seg>
    </tuv>
</tu>
```

Glossary

TM

- Segment pairs
- Format: .tmx
- Similar or identical content is automatically inserted
- On average, segments are long (phrases)

Glossari

- List of entries (similar to a dictionary)
- Can contain context information (comments)
- Format: .xlsx
- A CAT tool flags if there are available entries
- Usually, they need to be inserted manually

How to keep formatting?

13

`<g id="1">lexiQA </g> <bx id="2"/> (<g id="3">www.lexiQa.net </g>) is <g id="4">now </g> in` **>** `<g id="1">lexiQA </g> <bx id="2"/> (<g id="3">www.lexiQa.net </g>) <g id="4">funziona </g>`
`full operation with <g id="5">MateCat </g> .` `<g id="5">test </g> .`

TRANSLATED

⚠ More/fewer whitespaces found next to the tags. (1)

Translation Matches (2) TM Search Glossary

14 [Add private resources](#)

Tags represent formatting information and metadata that are compressed and easily movable across the segment

The CAT tool market today



And many, many, many, many more...

SDL Studio

SDL Trados Studio - ANNUAL PLAN DETAILS revised.docx_en-US_ni-NL

File Home Review Advanced View Add-Ins Help

Project Settings Configuration Batch Tasks File Actions Copy Paste Clear Formatting QuickInsert Concordance Search Translation Memory Select Previous Match Select Next Match Apply Translation Show Translations Add New Term Quick Add New Term Terminology Confirm Copy Source to Target Copy All Source to Target Clear Target Segment Merge Segments Split Segments Change Segment Status Navigation Find Replace Select All Editing

Editor

ANNUAL PLAN DETAILS
PGI.docx_en-US_ni-NL.s

Generiek (algemeen) - Translation Results

Project Settings

Term Recognition

1 <hyperlink value="http://www.pgi.com/imeet/pricing/#annual"> ANNUAL PLAN DETAILS </hyperlink> - iMeet Plus

1 <hyperlink value="http://www.pgi.com/imeet/pricing/#annual"> ANNUAL PLAN DETAILS </hyperlink> - iMeet Plus CM <hyperlink value="http://www.pgi.com/imeet/pricing/#annual"> JAARABONNEMENT </hyperlink> - iMeet Plus

Generiek (algemeen)

6-6-2015 11:47:13 pc-pieter\pieter Beens

Generiek (algemeen) - Translation Results

Concordance Search Comments Messages (0)

ANNUAL PLAN DETAILS revised.docx [Translation en-US-ni-NL] PGI.docx_en-US_ni-NL.sdlxiff [Translation]

Source Text	Target Text	Quality	Tags
1 <hyperlink value="http://www.pgi.com/imeet/pricing/#annual"> ANNUAL PLAN DETAILS </hyperlink> - iMeet Plus	1 <hyperlink value="http://www.pgi.com/imeet/pricing/#annual"> JAARABONNEMENT </hyperlink> - iMeet Plus	85%	P
2 http://www.pgi.com/imeet/pricing/#annual	2 http://www.pgi.com/imeet/pricing/#annual	100%	TAG
3 €29 / £24 per month subject to a minimum of 12 months	3 € 29 per maand bij een abonnement van minimaal 12 maanden	98%	P
4 iMeet includes;	4 iMeet biedt:	100%	P
5 Up to 35 participants	5 Tot wel 35 deelnemers	100%	P
6 Unlimited use of the iMeet video and web service offering	6 Onbeperkt gebruik van de video- en webdiensten van iMeet	100%	P
7 Unlimited use of softphone audio	7 Onbeperkt gebruik van "softphone"-audio	100%	P
8 5 GB of file storage	8 5 GB opslagruimte	100%	P
9 Applications available for Microsoft Outlook, Microsoft and Mac desktop, iOS and Android devices	9 Toepassingen voor Microsoft Outlook, Microsoft- en Mac-computers en iOS- en Android-apparaten	100%	P
10 Training and 24 hour support included.	10 Training en dag en nacht ondersteuning	100%	P
11 Not included in the price;	11 Niet bij de prijs inbegrepen:	100%	P
12 Any audio over the telephone.	12 Audio via de telefoon	100%	P
13 This is available and priced on a per minute basis.	13 Deze functie is verkrijgbaar op basis van een bedrag per minuut	100%	P
14 Interested in this package;	14 Geïnteresseerd?	100%	P
15 Enter your details here and one of our meeting experts will contact you within 24 hours.	15 Laat uw gegevens hier achter. Onze medewerkers nemen dan binnen 24 uur contact met u op!	100%	P
16 <hyperlink value="http://www.pgi.com/imeet/pricing/#annual"> ANNUAL PLAN DETAILS </hyperlink> - iMeet Plus	16 <hyperlink value="http://www.pgi.com/imeet/pricing/#annual"> JAARABONNEMENT </hyperlink> - iMeet Plus	85%	P
17 http://www.pgi.com/imeet/pricing/#annual	17 http://www.pgi.com/imeet/pricing/#annual	100%	TAG
18 €55 / £46 per month subject to a minimum of 12 months	18 € 55 per maand bij een abonnement van minimaal 12 maanden	98%	P
19 iMeet Plus includes;	19 JAARABONNEMENT - iMeet Plus	100%	P
20 Up to 35 participants	20 Tot wel 35 deelnemers	100%	P
21 Unlimited use of the iMeet video and web service offering	21 Onbeperkt gebruik van de video- en webdiensten van iMeet	100%	P
22 Unlimited use of softphone audio	22 Onbeperkt gebruik van "softphone"-audio	100%	P
23 Unlimited local dial in and dial out to landline locations listed below*	23 20 GB opslagruimte	100%	P
24 20 GB of file storage	24 Training en dag en nacht ondersteuning	100%	P
25 Applications available for Microsoft Outlook, Microsoft and Mac desktop, iOS and Android devices.	25 Niet bij de prijs inbegrepen:	100%	P
26 Training and 24 hour support included.			
27 * Austria, Australia, Belgium, Canada, China (Shanghai, Beijing), Denmark, Finland, France, Germany, Ireland, Italy, Netherlands, Norway, Singapore, South Africa, Spain, Sweden, Switzerland, UK, USA (fair usage applies)			
28 Not included in the price;			

All segments | INS | 44.00% | 29.07% | 26.93% | Chars: 27 | 11:47 | 6-6-2015

What is MateCat?

A Free Translation Tool for LSPs and freelancers

A completely free and web-based translation software (CAT tool) that handles more than 100 languages and over 60 file formats used in the industry. Disrupting the current status quo: **Offering for free what used to be charged thousands of dollars.**

Reuse made easy

Making state-of-the-art web crawling and neural machine translation available to every translator. No one in the world should translate something that has already been translated by somebody else. **Faster turnarounds and higher margins.**

Outsourcing in 1-click

1-click outsource to the best translator within the tool. We leverage Translated's T-Rank and 285,000 translators recruited so far to short-cut the industry, making **a direct link between the LSP need and the translators' skills.**

Why MateCat?

Limited use of CAT tools

- Only 60% of projects translated with a CAT tool

Security concerns

- Security issues when sending files to vendors

Limited collaboration

- Inefficient collaboration on rush and large projects

Scarce content reuse

- Limited content reuse due to small, local translation memories and poor MT integration



Join the Evolution