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


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.
$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.
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)
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)
When the export is ready, an email will be sent to your email address with a link to download the translation memory.

Quando l'esportazione è pronta, verrà inviata una e-mail al tuo indirizzo e-mail con un link per scaricare la memoria di traduzione.
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?

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...
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