

TM & CAT tools *for* Tibetan

*How to improve translation quality while
saving time & money*

TM - Translation Memory

*A Translation Memory (or **TM**) is a tool that saves translation segments so that you can **reuse** them. The benefits of translation memory include:*



- **Consistency** in translations on large projects (Cloud-based, team TM)
- Increases in **Accuracy & Quality**
- Streamlined Translating
- **Reduced costs** (20%-80%)
- Efficient editing & revision process

CAT - Computer Assisted Translation



A **CAT** tool is a **computer program** that assists translators by:

1. **Splitting** source texts in segments, making translating *easier*—by assisting with up to **80%** of the text—which makes the process up to **40% faster**
2. **Displaying** the source text & the translation side-by-side, for quick and easy reference, while referencing **TM** databases & suggesting translations for words & segments
3. **Saving** newly translated units in the **TM** database, so that they can be reused later for the same, or another, text; plus, through special "fuzzy search," **CAT** tools don't only find exact matches—they also work on similar segments

CAT Tools

Features of a typical CAT tool, which can be personalized & shared, include....

- **Multilingual Terminology**
- **Translation Memory**
- **Machine Translation**
- **Dictionaries/ Glossaries/ Thesauruses**
- **Quality Control & Editing**
- **Comments between Translators & Editors**

Terminology

SmartCAT can identify all terms in a text and indicate the correct translation (based on your glossary) regardless of a word's grammatical form.

Translation Memory

Translators can see matches from your centralized TM. All they have to do is insert matches and start editing. Final translation versions are automatically saved, so your TM grows without any additional work.

Machine translation

Choose one of the built-in engines or connect your own machine translation engine.

Dictionaries

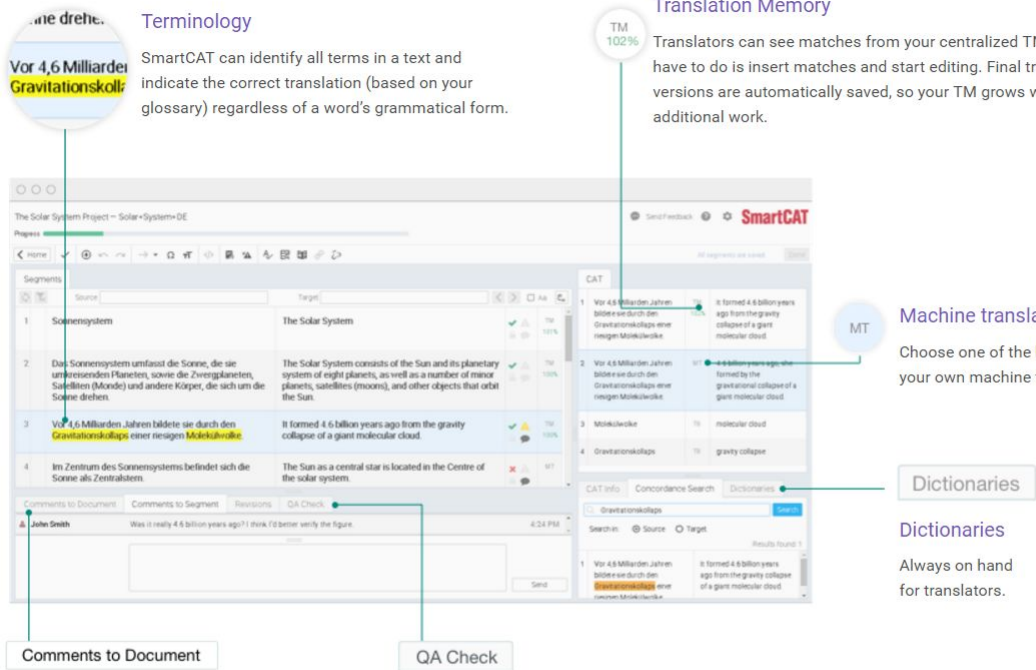
Dictionaries
Always on hand
for translators.

Comments

Translators and editors can leave comments about an entire document or individual segments.

Automatic quality assurance

Detect more than 20 types of mistakes in real time.



DO YOU
CAT?



YES I
DO!

FOR...

Cost



Why pay to translate
twice, thrice, ...

Time



Make translation
more efficient

Consistency



Ensure consistency
across teams

INPUT: UNTRANSLATED



Sentence / Phrase / Word

Translation Memory

OUTPUT:

Quem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur.

1. Untranslated - The sentence must be translated from scratch

Sentence has never been translated

2. 50-99% Fuzzy Match - Many recurring words & phrases within the sentence can be **reused**

The translator translates the sentence

A similar sentence has been translated

TM sends the translation & shows the differences

The sentence has been translated

TM gives the translation

3. 100% Match - Most of the document is translated; just need to double-check translation/order/syntax

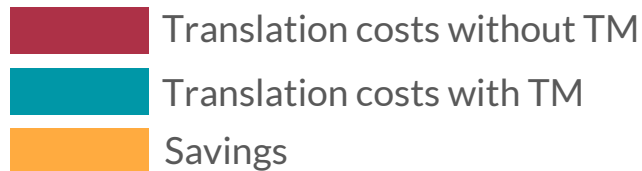
Translator modifies the sentence

The translator checks its accuracy

Translation is added to the **TM** for the future

Comparison of Costs

with & without TM



Save up to
40% in
both time
and money



Year 1
10,000 words
30% repetition




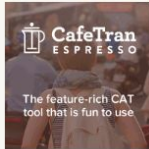




























Year 2
300,000 words
40% repetition



Year 3
600,000 words
50% repetition

Some Existing CAT Tools...

 <p>SDL Trados Studio 2015</p> <p>★★★★☆ (19 reviews)</p>	 <p>memoQ 2013</p> <p>★★★★★ (25 reviews)</p>	 <p>Wordfast Classic</p> <p>★★★★★ (8 reviews)</p>	 <p>CafeTran Espresso</p> <p>★★★★★ (5 reviews)</p>	 <p>MateCat</p> <p>★★★★★ (23 reviews)</p>	 <p>Déjà Vu X3</p> <p>★★★★★ (2 reviews)</p>	 <p>SDL Trados Studio 2011 - Starter Edition</p> <p>★★★★★ (0 reviews)</p>	 <p>Text United Translation Software</p> <p>★★★★★ (0 reviews)</p>	 <p>Atril Déjà Vu X2 Professional</p> <p>★★★★★ (11 reviews)</p>	 <p>Wordfast Pro</p> <p>★★★★★ (12 reviews)</p>
 <p>Fluency Translation Suite 2013</p> <p>★★★★★ (29 reviews)</p>	 <p>Transit NXT Professional+</p> <p>★★★★★ (4 reviews)</p>	 <p>Across Personal Edition</p> <p>★★★★★ (30 reviews)</p>	 <p>MetaTaxis for Word</p> <p>★★★★★ (5 reviews)</p>	 <p>Wordfast Anywhere</p> <p>★★★★★ (6 reviews)</p>	 <p>LogiTerm Pro</p> <p>★★★★★ (2 reviews)</p>	 <p>Smartling Translation Management System</p> <p>★★★★★ (37 reviews)</p>	 <p>SDL TRADOS Studio 2011 - Freelance</p> <p>★★★★★ (21 reviews)</p>	 <p>Lilt</p> <p>★★★★★ (1 review)</p>	 <p>SmartCAT</p> <p>★★★★★ (76 reviews)</p>
 <p>Swordfish Translation Editor</p> <p>★★★★★ (7 reviews)</p>	 <p>Memsources Cloud</p> <p>★★★★★ (8 reviews)</p>	 <p>OmegaT</p> <p>★★★★★ (8 reviews)</p>	 <p>XTM Cloud</p> <p>★★★★★ (11 reviews)</p>	 <p>MultiTrans Prism Freelance</p> <p>★★★★★ (1 review)</p>	 <p>Heartsome Translation Studio</p> <p>★★★★★ (3 reviews)</p>	 <p>Verifika - Quality Assurance Tool for Translators</p> <p>★★★★★ (0 reviews)</p>	 <p>MadCap Lingo</p> <p>★★★★★ (2 reviews)</p>	 <p>GeoWorkZ - Translation Workspace</p> <p>★★★★★ (3 reviews)</p>	 <p>OLanguage</p> <p>★★★★★ (0 reviews)</p>

... for Tibetan:



SmartCAT
★★★★★
(76 reviews)



OmegaT
★★★★★
(8 reviews)



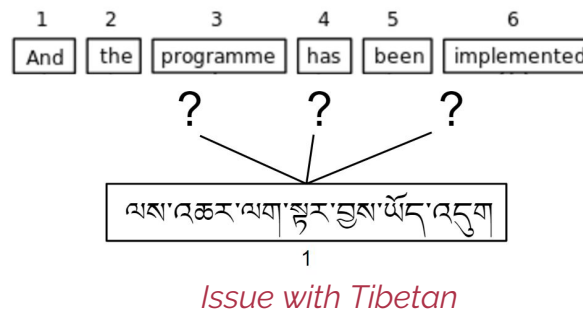
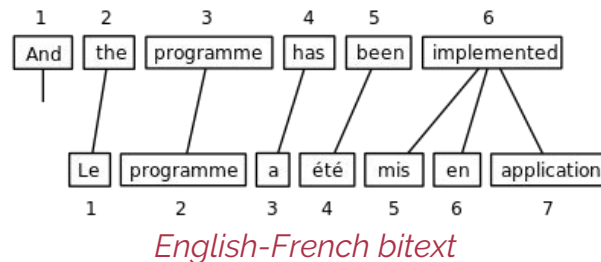
Why CAT isn't used for Tibetan

Texts in a **TM** are called **bitext** because they are synchronized with their translations at the sentence or at the word level.

The process of creating bitext is called **alignment** and is done using specialized programs called **word aligners**.

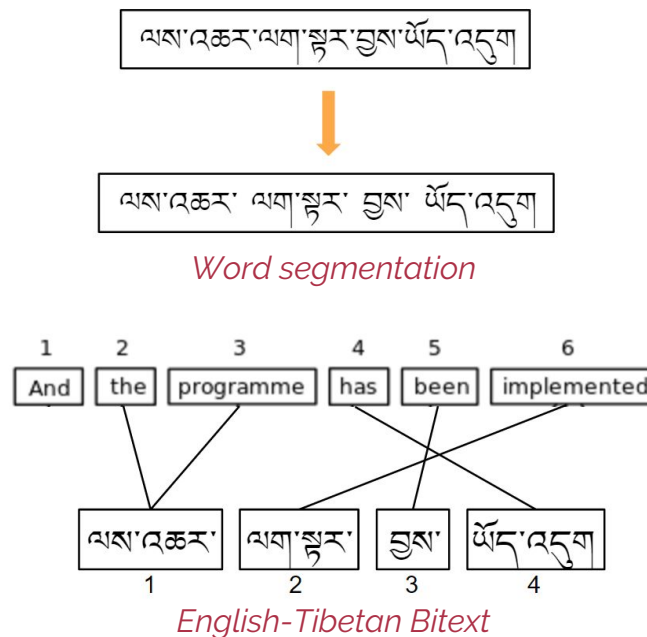
Currently, word aligners can't be used for Tibetan, because in Tibetan sentences, words are not separated from each other.

Without alignment, TM can't be generated; thus, it becomes very difficult to use CAT.



What Needs to Be Done?

1. Create a **segmenter** to cut sentences into words
2. **Align sentences manually** for existing translations
3. Create a **word aligner** for Tibetan
4. Extract **TMs** and glossaries of words and phrases
5. Set up a public **Tibetan TM exchange platform** on GitHub or Google Drive
6. Prepare **tutorials** to show translators how to use CAT tools like **SmartCAT**



More Information

[One of the best online CAT tools: SmartCAT](#)

[Best offline CAT tool: OmegaT](#)

[Comparison of CAT tools on proz.com](#)

[Word segmentation](#)