CAFETRAN ESPRESSO - Translation Memories options

Updated for CafeTran Espresso 2019

This document offers an overview of CafeTran's Translation Memory options.

Some memory-related settings can be found in [Edit > Preferences (Options) > Memory](#).

You can find additional information on [Working with Translation Memories](#) In the official CafeTran Knowledge Base (Solutions).

Curated by [Jean Dimitriadis](#) (EN-FR/EL-FR translator).

If you want to check out the latest CafeTran developments, review the [official update annoucements](#) or the aggregated [ChangeLog](#).

TM context menu options
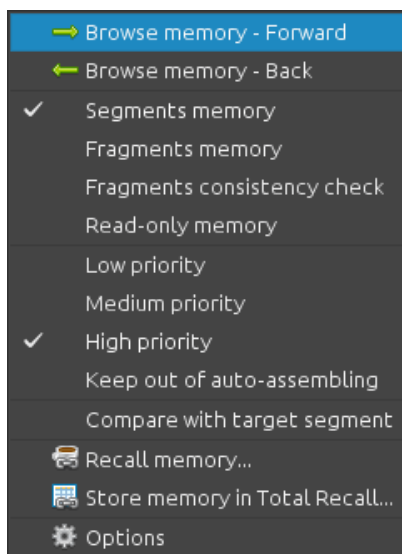
The TM context menu can be accessed by right-clicking inside the pane of a specific Translation Memory.

Browse memory

- **Forward**
- **Back**

This allows you to browse the selected memory Back or Forward. TMX units are displayed in pages, based on the number of segments defined in Preferences (Options) > General > Project page size (units) [default value is 50].

Note: You can achieve the same with the Back and Forward buttons in the Quick search bar.

Memory type

Quickly check or change the memory type settings for the TM, which can also be set via the TM Options (see below).

Segments memory = If enabled, the translation memory stores the segments of the project you work on as they are added to memory (see more detailed explanations in TM options below).

Fragments memory = If enabled, you can have the TM store terms and fragments (see more detailed explanations in TM options below).

Fragments consistency check = If enabled, the option allows you to check for segment pairs that do not have terms consistent with entries in this translation memory, when running the QA > Fragments consistency check (Memory) check.

Read-only memory = This sets the memory as read-only. No segments are stored when adding segments to memory.
Note: If a memory is set as Read-only, you may not access its Options, unless you close and reopen it.

Priority

Quickly check or change the priority level for the TM, which can also be set via the TM Options (see below). This affects fuzzy matching and auto-assembling results.

- **Low priority**
- **Medium priority**
- **High Priority**

Keep out of auto-assembling = If enabled, this lets you remove the specific resource from auto-assembling calculations. Keeping only the relevant resources for auto-assembling (and adjusting their priority, see above) helps get better results.

Compare with target segment = If enabled, this option shows the red "Different translation" warning in the TM tab for Exact matches. It can also be set for individual TMs after the right-click.

Recall memory = Recall the contents of the selected database table to this translation memory in the lexical context of the current project. Also available through Memory > Recall memory. This is the same as selecting a Total Recall table to recall from the Dashboard.

Store memory in Total Recall = Save the currently selected translation memory in a Total Recall database table. Also available through Memory > Store memory in Total Recall.

Options = This opens the all-important TM options dialog (see below).

TM options

The TM options dialog offers many important settings for your Translation Memories.

Memory type

In CafeTran, TMX translation memories can be used for storing segments, fragments/terms, or both.

Segments memory = Checkbox. If enabled, the translation memory stores the segments (Translation Units or TUs) of the project you work on as they are added to memory (see related Actions menu items and Target segment editor buttons in "CafeTran Menu and Interface").

With this setting ON, the TMX acts as a standard Translation Memory.

Processing tags = Checkbox. If enabled, CafeTran also stores information on the source and segment tags.

Suggestion: Tags count toward fuzzy match calculation, so this is a nice option to keep enabled for ProjectTMs. Maybe less so if you use Translation Memories to store segments across multiple projects. However, saved tags can easily be removed if needed, by selecting this TM filter options > Remove internal tags (see below) when opening a TM.

Read-only = Checkbox. This sets the memory as read-only. No segments are stored when adding segments to memory.

Note: If a memory is set as Read-only, you may not access its Options, unless you close and reopen it.

Fragments memory = Checkbox. Fragments or subsegments are parts of segments and phrases. They can consist of one or more terms (which in turn can consist of one or more words). If you select this option, you can have the Translation Memory store terms and fragments.

With this setting ON, the TMX can act as as a flexible glossary/termbase/fragment base for all intents and purposes, with all the advantages the TMX format can offer.

You can either use it in conjunction with Segments memory option, for storing entire segments and fragments in the same TM, or separately, for example for storing fragments that can be reused across projects.

If this option is ON, the fragment TM will be offered as an option to store fragments as you add them (see related Actions menu items and Target segment editor buttons in "CafeTran Menu and Interface").

Fragments consistency check = If enabled, the option allows you to check for segment pairs that do not have terms consistent with entries in this translation memory, when running the QA > Fragments consistency check (Memory) check.

Preliminary memory matching only = Use the TM only for Preliminary memory matching.

To overcome this issue of huge memories, you can perform the preliminary memory matching of the whole source document so that the matches will be available instantly, without any delay.

Also see Preferences > Memory > Switch to preliminary matching threshold in "CafeTran Preferences".

Related links:
Preliminary Memory Matching

Language pair

- **Source language** = Select the source language/language variant.
- **Target language** = Select the target language/language variant.

Note: Clicking anywhere in the Language pair section inverts the chosen language pair.

Options

Encoding = Change the character encoding for the TM. UTF-8 (default) is generally sufficient.

Priority = Drop-down menu. Choices: Low priority, Medium priority, High priority

Workflow integration = Drop-down menu. Choices: Automatic, Preliminary memory matching, Manual. Default: Automatic.

- **Automatic** = After you load a translation memory either from TMX files or from Total Recall, CafeTran performs the automatic matching. The currently-translated segment is matched automatically against this translation memory to produce the results of matching.
- **Preliminary memory matching** = As your translation memory is getting bigger and bigger over time and reaches thousands of segments, the automatic matching may get slower as the current source segment is matched against that large amount of translation memory segments. To overcome this issue of huge memories, you can perform the preliminary memory matching of the whole source document so that the matches will be available instantly, without any delay. The preliminary memory matching may take a while but you do not have to wait until it completes because matches are available right after the segment is processed. It means that you can translate at your own human pace while the preliminary matching is taking place in the background.
- **Manual** = The memory is not being matched against your current-translated segment, but you can conduct manual concordance searches in source or target memory entries.

Related links:
Preliminary Memory Matching

Matching type = Drop-down menu. Choices: Fuzzy & hits, Fuzzy, Fuzzy without word separator. Default: Fuzzy & hits.

- **Fuzzy & hits** = With this option, CafeTran analyzes source segments on a word basis (fuzzy matching), and performs statistical analysis of subsegments to determine their translation (fragment matches aka Hits). See an explanation of Fragment Hits.
- **Fuzzy** = With this option, CafeTran only performs fuzzy matching analysis. Disabling subsegment analysis speeds up the matching process.
- **Fuzzy without word separator** = With this option, CafeTran analyzes source segments on a character basis, which is suitable for languages without a word boundary (e.g. Chinese or Japanese).

Minimal prefix length = This option lets you set the minimal allowed length of prefixes, if Prefix matching box is checked.

Prefix matching (%) = Checkbox and drop-down menu. Choices: Fixed length, 10, 20, 30, 40, 50, 60, 70, 80, 90. Default: Fixed length.

When this option is selected, CafeTran will analyze the beginnings of words (here called prefixes) and discard any endings responsible for inflection of words. It is an option which increases significantly the number of

hits for highly inflected languages. The length of prefixes is set by a percentage number. The bigger the percent number the longer the prefix of words which the program will analyze. The length can also be fixed, when the "fixed" option selected, instead of a set percentage length. It means that all the words will have the minimal prefix length, no matter their actual length.

Note: As the "Minimal prefix length" option for Glossaries (Preferences > Glossary > Minimal prefix matching) is shared with TM fragments, for glossaries, you can adjust the prefix length in the options for the memory.

A note on Custom prefixes:
If the inflection of a word is too high for automatic prefix matching you can enter your terms to the memory determining the prefix of a word manually. This is done by inserting the pipe character | at the end of a prefix in a word. For example, the Polish phrase "piękny dzień" (a beautiful day) has a highly inflected word "dzień" occuring in a number of various cases (dnia, dni, dniom). If you insert the pipe characters at the following positions - "pięk|ny d|zień", CafeTran will also recognize other forms of the phrase (pięknego dnia, pięknych dni etc.). Note that inserting the pipe character at the first word in the phrase - "pięk|ny" is op-tional since its inflection is quite regular and CafeTran should recognize its prefix automatically .

Match case = Checkbox. If enabled, CafeTran takes into account the text case when looking for matches, differentiating identical Translation Units that only differ in case type (uppercase or lowercase).

Greedy exact matches = Checkbox. With this option enabled, CafeTran does not stop the search at the found exact match(es) for the current segment, but it continues looking for fuzzy matches and fragments. Switching on this option is not recommended for slower computers and huge translation memories.

Duplicates option = Drop-down menu. Choices: Keep all duplicates, Keep newer duplicates, Keep old duplicates. Default: Keep newer duplicates.

- **Keep all duplicates** = Select this if you want CafeTran to keep both newer and older identical source segments in the memory.
- **Keep newer duplicates** = Select this if you want CafeTran to overwrite old Translation Units (segments) when you add a new translation for the same TU.
- **Keep old duplicates** = Select this if you want CafeTran to ignore new identical source Translation Units which are already present in the memory.

Filter button = Opens the TM filter options dialog (see below).

TM filter options

Memory filter options are only available in the dialog before opening the Memory. Once opened, the button is grayed out.

Properties filter

Default fields are:

- **Project**
- **Client**
- **Property**

The properties filter allows you to filter only the TMX units that include the value for the specified property.

If you have defined a different property, you can click the button of the property to rename it.

If you have set a TXT file for Client and/or Subject in Preferences > Definitions, here you get a Client and Subject field as an editable drop-down menu, allowing you to choose already defined Client and Subject fields.

Segment filters

- **Filter languages** = Checkbox. If enabled, this removes all languages from the TM except for the source and target language in your translation project.
- **Remove empty segments** = Checkbox. If checked, it removes empty segments (those with no translation).
- **Clean and replace foreign codes** = Checkbox. Some TMX files from third-party tools have unusual codes in the segments such as codes inside the curly brackets or emdash, endash, tab code. CafeTran clears or replaces them with equivalent unicode characters.
- **Remove source=target** = Checkbox. If checked, it removes TMX units where the target is identical to the source.
- **Remove internal tags** = Checkbox. If checked, it removes tags inside source and target segments.

- **Segment note** = Text field. Here you can insert a segment note to load only the translation units that have a particular note.
- **Regular expression** = Text field. Here you can insert a segment note to load only the translation units that match with a particular (Java) regular expression.

Total Recall options

Database - [name of database]

Recall in context (hits per word) = Checkbox and number field. This value sets the linguistic filter which analyses the current project, bringing back only the segments that contain words of the source document present in the project. The value determines the maximum number of the segments which are to be recalled for one word of the source document being translated. Checking out the "Recall in context" box lifts the filter completely and recalls all segments from the Total Recall table to the working memory.

Default value: 1000.

Related links:
[Recalling Segments in the Project Context](#)

All other TM and filter options are the same as those described above.

Priority default option is set to Low instead of High, though, and the resulting memory is set as Read only.

Note: Total Recall does not process tags.

List of changes

- 20171117 File creation
- 20180217 Migrated from Classeur.io to GitHub.com
- 20180621 Updated for CafeTran Espresso 2018 - Akua Update 10

Feedback

You can send me feedback via a ProZ message. Please start your subject with the name of the document: "CafeTran - TM options".