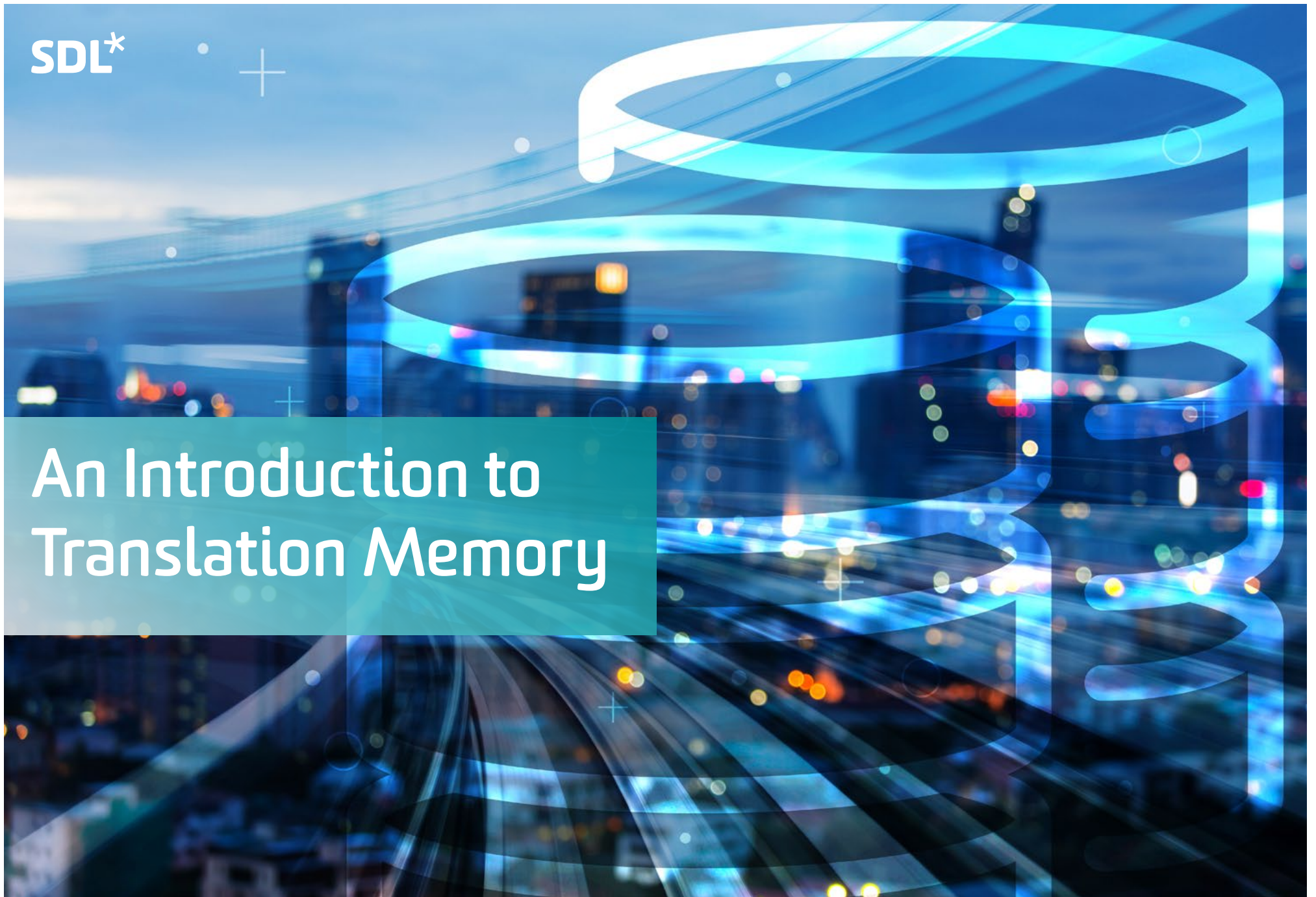SDL*

An Introduction to Translation Memory

This ebook provides an introduction to translation memory technology. Read it from cover to cover, or jump to a particular section by clicking a link below.
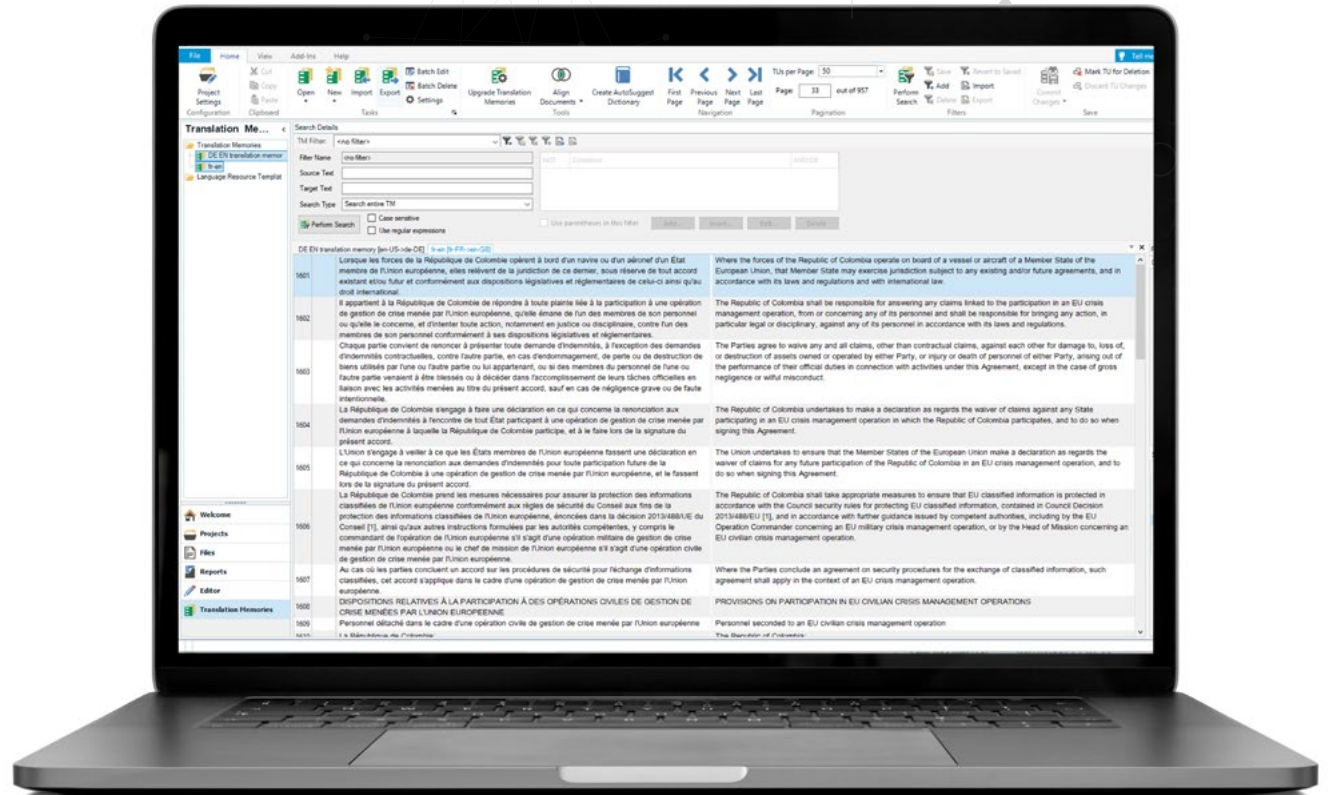
What is a translation memory?

A translation memory (TM) is a database that stores segments of source language text alongside their translated counterparts. Segments are usually whole sentences, but can also be shorter (e.g. a single word in a cell within a Microsoft Excel document) or longer (e.g. an entire paragraph).

Translation memories are built up gradually as you translate, so you start with an empty TM and it grows, sentence by sentence, over the course of months and years. A TM can also be created from existing resources.

As you work, your translation memory flags up identical or similar segments that you have translated in the past, to help you work faster and more consistently. Newly translated segments are added to the TM.

A brief history

In the beginning...

The 1960s saw the birth of multilingual glossaries, initially with punched-card input and later in magnetic tape format. In the 1970s, the original term-by-term concept evolved with the addition of KWIC (keyword in context) searches, developed through the European Commission's Eurodicautom and the Canadian Government's Termium bank. A decade later, early translation memory systems introduced paired sentence-storing capability in MS-DOS, so that exact matches could be retrieved from previously translated segments.

However, it wasn't until the 1990s, with the advent of the Windows operating system on the one hand, and Trados Workbench on the other, that translators could start to glean the benefits of computer-aided translation systems in general, and translation memories in particular.

1960s Multilingual glossaries, initially with punched-card input and later in magnetic tape format

1970s The original term-by-term concept evolved with the addition of KWIC searches

1980s Early translation memory systems introduced paired sentence-storing capability in MS-DOS

1990s Windows operating system and Trados Workbench enable computer-aided translation

Latest developments in translation memory technology

Click here to discover the latest developments in translation memory technology!

References
Hutchins, J. The origins of the translator's workstation. Machine Translation, vol.13, no.4 (1998), p. 287-307 García, I. Long term memories: Trados and TM turn 20. The Journal of Specialised Translation, Issue 04 — July 2005

Why use a translation memory?

More to it than "Never translate the same sentence twice"

There are many benefits of using a TM:

Work faster

Imagine you translate a set of instructions for a washing machine. A year later, the same company sends you the user manual for a new model. Some of the text will be identical, other parts will have minor changes and the rest will be brand new. Your translation memory will retrieve exact (100%) matches, flag up similar segments ("fuzzy matches") and leave the rest for you to translate from scratch.

Ensure consistency

Once you've decided how to translate a tricky phrase, your translation memory will remind you to use it in the future. It might be your own solution, a matter of client preference or a stock phrase. Whatever the reason, your TM will help you use it consistently.

Get suggestions for similar content

Sub-segment matches in translation memories are offered through AutoSuggest in SDL Trados Studio. If you translate software, for example, *Hold down the Ctrl key* will be a common expression, but the rest of the sentence may vary too much to get a good fuzzy match. AutoSuggest 2.0 in SDL Trados Studio offers these relevant snippets as you type.

Sub-segment matches are also automatically provided in the Translation Memory Results window through the upLIFT Fragment Recall functionality in more recent versions of SDL Trados Studio.

Learn more about upLIFT technology on the 'latest developments in translation memory technology' page

See how you translated certain words or expressions in the past

Even if you don't get fuzzy matches from your translation memory or prompts from AutoSuggest, you can still look up one or more words at a time from your source text to see how you translated them in the past. In SDL Trados Studio, this is known as concordance. It's similar to a glossary feature but has the added benefits of seeing the words in context and identifying when and where they were translated. Check how you translated challenging words and phrases in the past by running an automated concordance check.

Comply with document version requirements

When you translate a new version of a formal document (e.g. a draft contract or agreement), unchanged text must be identical to previous versions. Pre-translate the source file with your translation memory to comply with version management.

All these benefits apply when you have a mature translation memory

Is an empty translation memory any use at all?

Yes! A brand new TM is a useful TM.

The following features work even if your TM is empty:

Repetitive text

When you translate a long, repetitive document, your translation memory offers fuzzy matches from segments that you save, one by one, as you work.

Auto-propagation

Identical segments are common in tables, forms and questionnaires. As soon as you have confirmed the first instance of a repetition, it is auto-propagated throughout the file. If you merge several files, identical segments are automatically translated in all files. Number-only segments are localized on the fly to match target language conventions and they are also auto-propagated.

Auto-substitution

Dates, times, numbers and measurements are "recognized tokens" in source segments and localized automatically, so you can insert them easily in the corresponding target segment.

Fuzzy match: A match that is less than 100%

Translation memory myths

Myth # 1: TMs are only useful for repetitive texts

Concordance and auto-substitution (previous page) are useful for all texts, repetitive or not. For creative texts, translators have even been known to use concordance for the opposite effect: to avoid repetition.

If you work in regulatory fields, third party TMs such as the European Commission DGT translation memory are important for referring to the official translations of EU directives and other legal texts.

Even if you don't benefit directly from a TM, there are many other reasons to work with a CAT (computer-assisted translation tool) such as SDL Trados Studio:

- Translate unusual file formats even if you don't have the native program
- Never skip a sentence by mistake
- Perform quality assurance as you translate – spelling, grammar, number check, extra spaces etc.
- Retain original formatting in the target file
- Back up your work automatically in the TM sentence by sentence
- See how your translation is progressing as a percentage or as number of translated words

Myth # 2: Fuzzy matches only benefit language service providers (LSPs)

If you're a freelance translator, you may find that some LSPs ask you for a discount on fuzzy matches from translation memories they provide. If the TM is good quality, high fuzzy matches will be faster to translate than new content, so it's probably reasonable to apply a discount for them. When you work for direct clients or with your own TM, you alone will reap the benefit of 'fuzzies', and this can be an important factor when considering how much to quote for a project and how long you will take to do it.

Myth # 3: A TM is like machine translation

Not true: a translation memory is the result of your hard work. It's a convenient way of storing everything you translate, because a human memory doesn't have that capacity. But it's all come from a human brain (and any errors are human errors). Machine translation is translation of text by a computer, with no human involvement.

How to start using a translation memory

Ready, steady, go!

Alignment
You can leverage past translations by aligning source and target files. Only align past work if you're sure it's relevant to your current project.

External resources
As mentioned earlier, the European Commission has official translations available in the DGT translation memory.

You can also investigate **MyMemory**, a free plug-in from SDL AppStore, the SDL online app store.

Start from scratch
The above options may be useful, but the best method is to simply create an empty translation memory in your language pair and start translating.

How to recognize a Translation Memory file:
A TM can be identified by its .sdltm extension

If you've never used a translation memory before, you've got three options:

1. Align old translations

2. Use external resources

3. Start from scratch

MyMemory,
the world's largest TM, is a free plug-in from SDL AppStore, the SDL online app store.

How to create a new translation memory in SDL Trados Studio in three quick steps:

1. Call up the New Translation Memory wizard by **clicking Alt+Shift+N** from anywhere in SDL Trados Studio.

2. Give the TM a name, define the location where you want to save it, and check that the language pair is correct.

3. Click Finish.*

* If you want to define fields or language resources (see section on Translation memory best practices), click on the metro map at the top of the Wizard before clicking Finish in step 3.

Note that you can also create a new TM when you select New Project or Translate Single Document in the Welcome view.

Translation memory best practices

Saving, organizing and maintaining TMs – frequently asked questions

How many TMs do I need?

'Big Mama' approach

Some translators use a single translation memory for all their work. This is known as the 'Big Mama' approach. With today's powerful computers, segment and concordance look-ups in huge TMs shouldn't be a problem, unless you're dealing with monsters like the DGT TM, which has some two million translation units.

If you opt for a 'Big Mama' translation memory, define your TM fields carefully so that later you can perform filters, apply penalties and delete units as needed (see below). Another strategy with a monster TM is to pre-translate a file with it and then set up a project TM for look-ups and concordance while you work.

Use Studio 2019 navigation features to browse through big TMs effortlessly. In the Translation Memory view, display up to 1000 translation units per page, skip to a specific page or go straight to the last page.

Several smaller TMs

You may prefer to keep separate TMs for your main clients and general TMs for other clients and fields of work. If you translate for different target audiences (formal, informal, adults, children, etc.) or use different styles (advertising copy, technical instructions, etc.), these translations should be stored in separate TMs. Name your TMs clearly. Add language suffixes if you work in different language pairs or language variants. Appending the year of creation to the TM file name also helps to identify a TM quickly.

Where should I save my TMs?

By default, SDL Trados Studio stores new TMs in **C:\Users\Username\ Documents\Studio 2019\Translation Memories**. Use this location or create another TM folder elsewhere. You may prefer to keep client-specific TMs in a client-specific folder.

What details should I enter in custom fields?

The more general the TM, the more specific the fields should be. You can define field values by subject, year, genre, client name. Fields can be formatted as a picklist, text entry, number format or date/time.

What do I need to know about working with TMs in practice?

- You can add several translation memories to a translation project.

- Use the AnyTM feature to add TMs that have different language variants or are in the reverse language direction from your project.

- It's best to update just one TM with newly translated content, and keep the others for look-up and concordance.

- Apply a penalty to TMs that you know are less relevant or less reliable. A penalty can range from 1%-25% and is reflected in the fuzzy match percentage that is returned from a TM look-up. If you think you might get identical 100% matches from more than one TM, apply a 1% penalty to all TMs except one.

- The default threshold for returning hits from TM look-up is 70%. New users may be tempted to lower this figure, but these additional results will be distracting and need a lot of rewriting. A higher percentage will produce fewer but more meaningful results.

- A warning triangle in a TM symbol means your TM needs upgrading to benefit from all the latest features. In Studio 2019, type "Upgrade" in the 'Tell me what you want to do' box to call up the wizard.

Should I change segmentation rules?

You can decide exactly how to segment a file. Segmentation rules are set in your translation memory. By default, a segment will end with a full stop, but you can add rules, for example, to start a new segment after a soft return, a tab or a colon. Similarly, you can add abbreviations where you don't want a new segment to start, such as Tues. or rec'd.

Decide which tokens (word or string) should be transferred directly or localised in the target segment. You can enable the following tokens:

dates

times

numbers

acronyms

variables

measurements

alphanumeric strings

Time, date and user details are automatically added to a translation unit. These fields are known as system fields.

The default threshold for returning hits from TM look-up is **70%**

Who owns the TM? The agency or me?

- If you create and update a translation memory, it clearly belongs to you.

- An agency may ask you to deliver your TM together with your translation. The solution in this case is to deliver the bilingual file or to export it into a new TM. The agency may prefer a translation memory in .tmx format, an industry standard translation memory format, so that it can be used in other tools.

- If an agency sends you a TM to work with, that TM is owned by the agency.

- Never share a translation memory that contains confidential data from different clients.

How do I maintain a TM?

- Your own translation memories will need little maintenance work if you've used field values carefully, have reviewed your work for errors, and haven't stored mismatched segments or content in other languages, which would contaminate your TM.

- Batch editing is useful for performing 'find and replace' operations in several segments at once, or to change field values.

- To trim a very big TM, filter and delete segments by "last used" rather than "created on" date, to make sure you keep the most useful segments.

- Always make a copy of your TM before editing it in any way.

Always make a copy of your TM before editing it in any way

Latest developments in translation memory technology

A brief history of translation memories at the start of this book examined the emergence and early development of TMs. Moving on several decades, here is a selection of the latest features that have been added to translation memory technology in recent years.

upLIFT

Introduced in SDL Trados Studio 2017, upLIFT technology takes translation memory to the next level by instantly providing you with sub-segment level matches and automatically improving the fuzzy matches you receive. upLIFT technology consists of two features: **Fragment Recall** and **Fuzzy Match Repair**.

Fragment Recall

upLIFT is implemented through AutoSuggest, so fragment suggestions appear as you type. Segments saved to a TM are available immediately through upLIFT, without needing to build a dictionary or have a minimum number of segments in the TM.

You can easily control the minimum number of words you would like in a match as well as the minimum number of significant words. Significant words are the ones that matter most for a correct match. Words like 'the' or 'a' are not significant for example.

Simply go to *File > Options > Language Pairs > All Language Pairs > Translation Memory and Automated Translation > Search >* and in the Fragment Matching Options for upLIFT, choose the minimum words for a match.

Fuzzy Match Repair

SDL Trados Studio analyzes fuzzy matches and repairs them from various resources, such as TMs, termbases and machine translation providers. The translator remains in control right through the process, from seeing which resource has been used, to further editing the segment or simply applying the translation without Fuzzy match repair.

Learn more by visiting our dedicated upLIFT webpage

LookAhead

Translation search results are usually retrieved in the active segment. With LookAhead, SDL Trados Studio performs a lookup (in TMs, termbases and machine translation providers) on upcoming segments, giving speedy results as the translator moves from one segment to the next.

To benefit from these state-of-the-art features, older translation memories must be upgraded. To do so, run the Upgrade Translation Memories wizard on the Welcome Screen or type "Upgrade" in the 'Tell me what you want to do' box in Studio 2019.

A glossary of TM terms

100% match
A segment that is an exact match, including formatting

Alignment
Automatic matching of source and target translations, segment by segment

Auto-propagation
Automatic propagation of a translation to other identical segments in a file

AutoSuggest
A feature in SDL Trados Studio that provides suggestions as you type that can be added into a translation quickly and easily

Concordance
A search for one or more words in the TM

Context match
A segment that is an exact match, and the preceding segment is also the same

Field
Supplementary data about a segment

Fuzzy match
A match that is less than 100%

LookAhead
A lookup (in TMs, termbases and machine translation providers) that is applied to upcoming segments to give instantaneous results as you work

Penalty
A deduction from a percentage match, defined by formatting differences, aligned source, user criteria, etc.

Pre-translation
An automated task where matches from the TM are applied to the source file

Segment
Text in a translation unit, often corresponding to a sentence

TMX
Translation Memory Exchange file format (an industry standard TM format)

Translation unit
A paired source and target segment with field data

upLIFT
Segment leverage to improve translation speed by recalling and repairing segment fragments

SDL*

About the author

Emma Goldsmith, RGN, DipTrans, MITI, is a freelance Spanish to English medical translator who made a career switch from nursing some 25 years ago. She learnt about CAT tools the hard way, throwing herself in the deep end, sceptically testing an empty TM and soon giving up.

Later, she embraced SDL Trados Studio by attending in-person and on-line training sessions and asking a lot of questions. She enjoys keeping abreast of translation technology developments and writes a blog called "Signs & Symptoms of Translation" about medical translation and SDL Trados Studio.

For more information about translation memory technology:

Freelancers and LSPs: **sdltrados.com/translation-memory**
Corporations: **sdl.com/translation-memory**

SDL*

SDL (LSE:SDL) is the global leader in content creation, translation and delivery. For over 25 years we've helped companies communicate with confidence and deliver transformative business results by enabling powerful experiences that engage customers across multiple touchpoints worldwide.

Are you in the know? Find out why the top global brands use SDL at **sdl.com** and follow us on **Twitter**, **LinkedIn**, **Facebook**, **YouTube**, **Google+** and **sdltrados.com**