AAB UNIVERSITY

Lecture 2

The role of technology in translation process
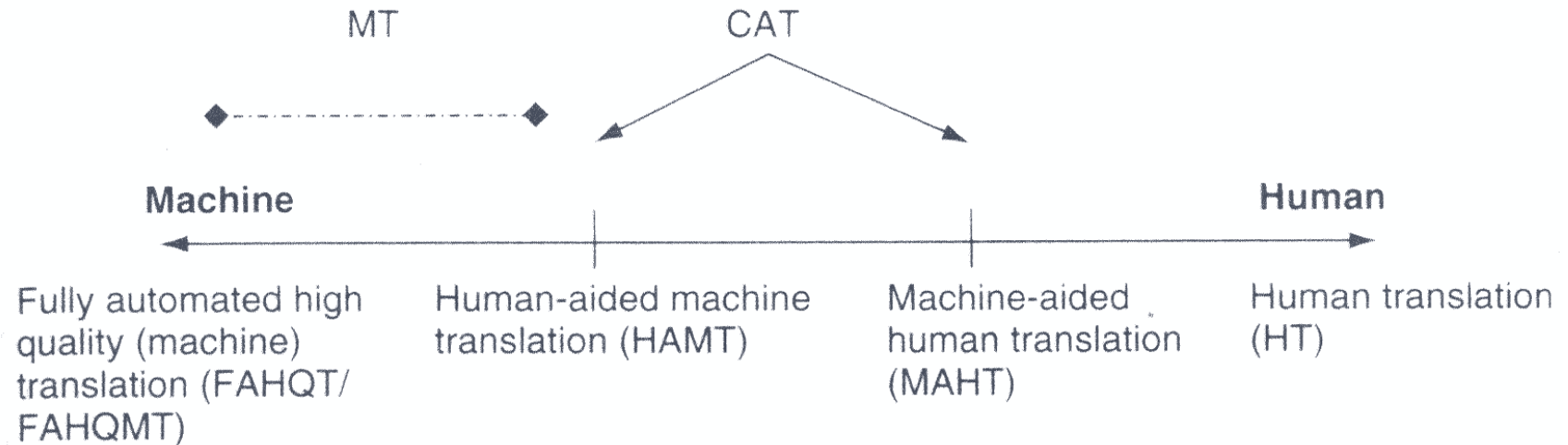
arianitm@gmail.com

Tel: 044 425 159

Table of contents

- Introduction
- Definition
- Characteristics of TMS (Translation Memory Systems)
- The translation workflow
- Reasons for using TMS
- Advantageous and disadvantageous of using TMS!
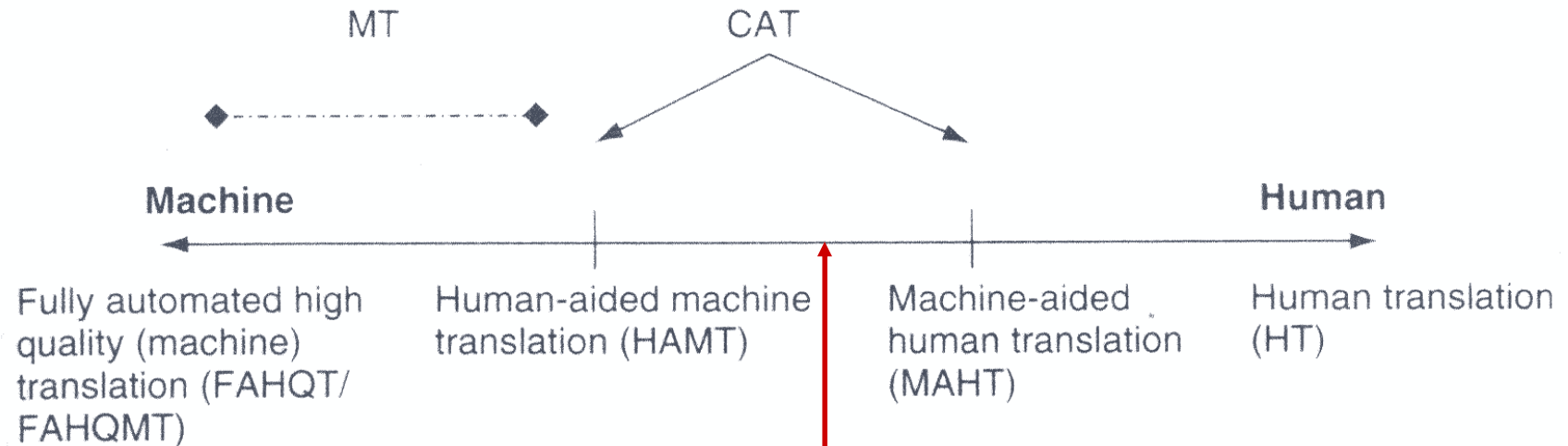
Definition

- Classification of translation types

MT   CAT

Machine   **Human**

Fully automated high quality (machine) translation (FAHQT/ FAHQMT)

Human-aided machine translation (HAMT)

Machine-aided human translation (MAHT)

Human translation (HT)

MT = machine translation; CAT = computer-aided translation

Definition

- Classification of translation types

MT   CAT

Machine   **Human**

Fully automated high quality (machine) translation (FAHQT/ FAHQMT)

Human-aided machine translation (HAMT)

Machine-aided human translation (MAHT)

Human translation (HT)

MT = machine translation; CAT = computer-aided translation

Translation Memory Systems

What is machine translation-MT?

- Machine translation is the translation of text by a computer, with no human involvement.

- Pioneered in the 1950s, machine translation can also be referred to as automated translation, automatic or instant translation.

How does MT work?

- **There are two types of machine translation system:**
  - **Rules-based and**
  - **Statistical:**
- Rules-based systems use a combination of language and grammar rules, plus dictionaries for common words.
- Rules-based systems typically deliver consistent translations with accurate terminology when trained with specialist dictionaries.

How does machine translation work?

- **There are two types of machine translation system:**
  - **Rules-based and**
  - **Statistical:**
- Statistical systems have no knowledge of language rules.
- Instead they "learn" to translate by analyzing large amounts of data for each language pair.

Definition of Machine Translation Systems (MTS)

- Machine translation combines a number of fields of study such as:
  - lexicography,
  - linguistics,
  - computational linguistics,
  - computer science and
  - language engineering.

Translation Memory Systems

Definition of TMS (1)

- A TMS does not create a machine translation; rather, it is a method of re-using previous translations.

- TM - also called **DB (Data Base) of translations**

- TM contains sections of text in the source language stored together with the corresponding text in the target language as "segment pairs".

Definition of TMS (2)

- The text sections are first divided into segments according to previously defined rules

- Markers used as segment ends are for example full stops (end of sentence) or paragraph markers

- Segments are therefore usually sentences, lists or table elements

Definition of TMS (3)

- Any new text is compared with the segments previously stored in the database.

- The TMS determines matching source language (**SL**) segments and offers the saved, equivalent segment in the target language (**TL**) as an existing translation.

- Similar segments are also recognized and offered as so-called fuzzy matches.

Definition of TMS (4)

- It is based on the hypothesis that natural languages can be fully described, controlled and mathematically coded.

Definition of TMS (5)

- Unlike **MTS** which generate translations automatically, **TMS** allow professional translators to be in charge of the decision-making whether to accept or reject a term or an equivalent phrase suggested by the system during the translation process.

Definition of TMS cont.

- Translators can also build their own 'memory'.

Why TMS? (1)

- The key challenge today in translation is managing your translation costs and volumes.

- Deciding how best to translate, whilst maintaining consistent high quality, is a tough challenge.

- Translation memory tools deliver on quality and in terms of cost savings, and help manage the process for you.

Why TMS? (2)

- Imagine if you could simply import your files for translation, translate the files with suggestions from your previous projects and then export your translated files, delivering consistent quality results time and again…

- NEVER TRANSLATE THE SAME SENTENCE TWICE.

Characteristics of TMS

- When a new document has been analyzed against an existing translation memory, the resulting segments are divided into following categories:
  - Perfect matching
  - Fuzzy matching
  - No matches
  - Repetitions
  - Filter
  - Segmentation
  - Alignment

Perfect matching

- **Perfect Matches (100 % match):** A segment in the new document is identical to a segment stored in the translation memory.

Fuzzy matching

- **Fuzzy Matches (75-99 % match):** Unlike a perfect match, a fuzzy match occurs when an old and a new SL (Source Language) segment are similar but not exactly identical.

- Even a very small difference such as punctuation leads to a fuzzy match.

Repetitions

- **Repetitions:** The same segment exists several times in the current document.

- This segment only has to be translated once.

- When the segment is repeated, the translation is automatically supplied by the translation memory.

No matches

- **No Matches (0-74 % match):**

- No similar segment has been found.

- The text must be newly translated.

Filter

- It is a feature that converts a SL text from one format into another, giving the translator the flexibility to work with texts of different formats:

  - Text without graphics
  - Text without HTML code signs

Filter example

HTML code signs

No HTML code signs

<!-- Side bar structure ends here --><!-- End of Math Side Bar --
><!-- Allow menubars in non-printable versions. --><!--
Secondary Menubar handling--><!-- End of Secondary Menu
Bar Handling -->
<div id="CreatorContent"><!-- Content Section begins here --
><!-- Math Body Prefix file ends here -->
<h2> Teaching Assignment Winter 04
</h2>
<a
href="http://www.stats.uwaterloo.ca/%7Epmarriott/STAT231/">
STAT
231
Empirical Problem Solving

Office Hours for STAT 231

Tuesday 10-11
Thursday 10-11

Figure 4.1 Example of HTML code in a web page

Teaching Assignment Winter 04
STAT 231 Empirical Problem Solving
Office Hours for STAT 231
Tuesday 10–11
Thursday 10–11

Figure 4.2 Example of the web page without HTML code

Segmentation

- It is a process of breaking a text up into units consisting of a word or a string of words that is linguistically acceptable.

Segmentation

- **Segmentation is especially useful for:**

  - Headings,

  - Lists,

  - Bullet points

Segmentation example

Segment	English
4.16	the translation of English affixes into Malay
4.17	in Malaysia
4.18	the terminology committee
4.19	the planning of the Malay language
4.20	scientific and technological terms
4.21	the translation of English affixes

one segment

Alignment

- It is a process of binding a SL segment to its corresponding TL segment.

*SL = Source Language
**TL = Target Language

Alignment

- It is a process of binding a SL segment to its corresponding TL segment.

- The purpose is to create a new translation memory base or to add to an existing one.

Alignment

- It is a process of binding a SL segment to its corresponding TL segment.

- The purpose is to create a new translation memory base or to add to an existing one.

- The corresponding pairs of SL & TL are called 'translation units'.

Alignment example

Segment	English (source)	Segment	French (target)
4.16	the translation of English affixes into Malay	4.16a	la traduction d'affixes anglais en malais
4.17	in Malaysia	4.17a	en Malaisie
4.18	the terminology committee	4.18a	la commission de terminologie
4.19	the planning of the Malay language	4.19a	la planification de la langue malaise
4.20	scientific and technological terms	4.20a	termes scientifiques et technologiques
4.21	the translation of English affixes	4.21a	la traduction des affixes anglais

TM and translation process

- During the translation process, the TMS marks the analyzed segments in the target text and labels them with the match percentage.

- The translator can choose to accept the suggested translation, adapt it or discard it altogether and enter a new translation.

- Each new translation is in turn stored in the translation memory and is therefore available for use in future projects
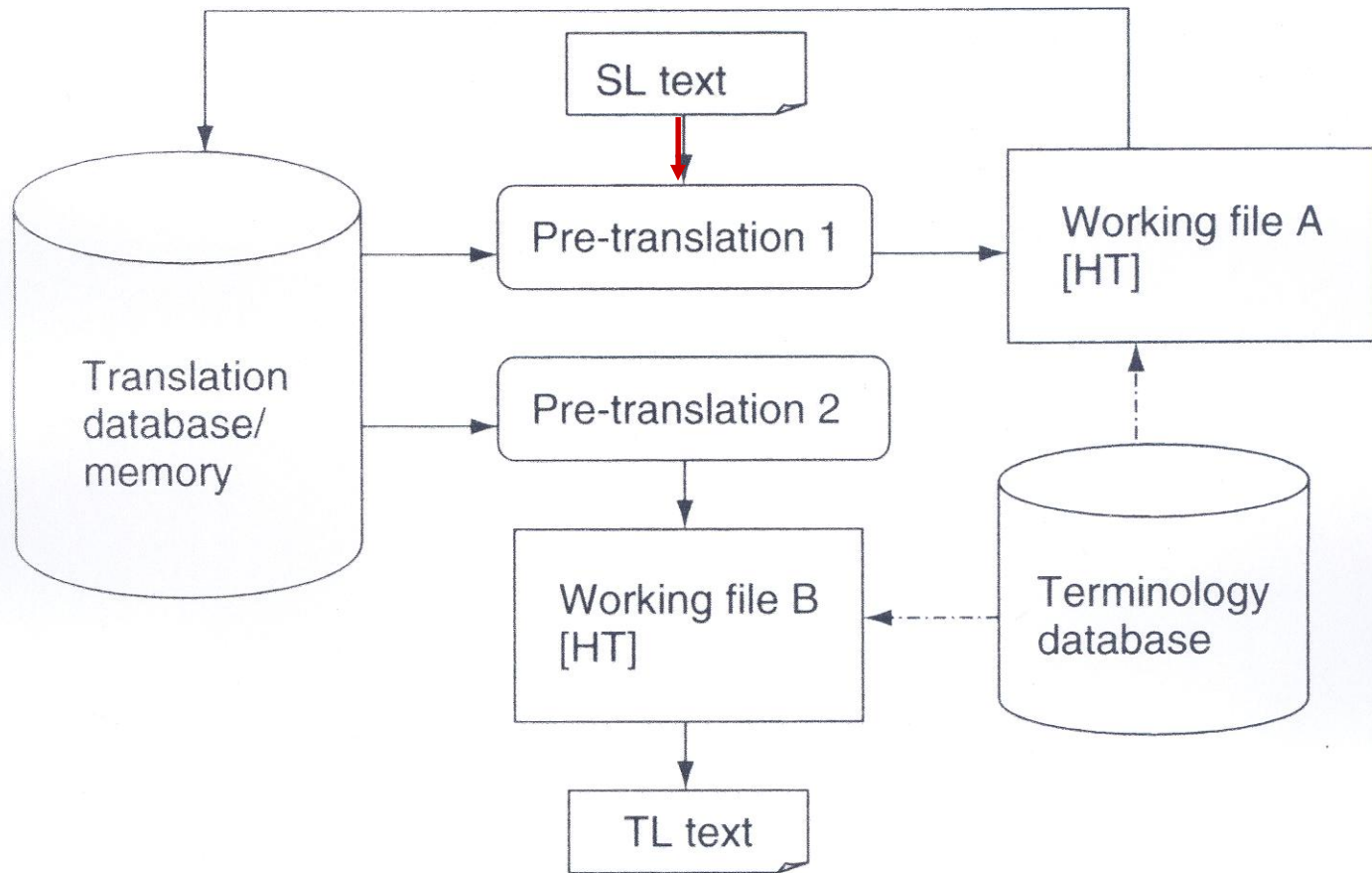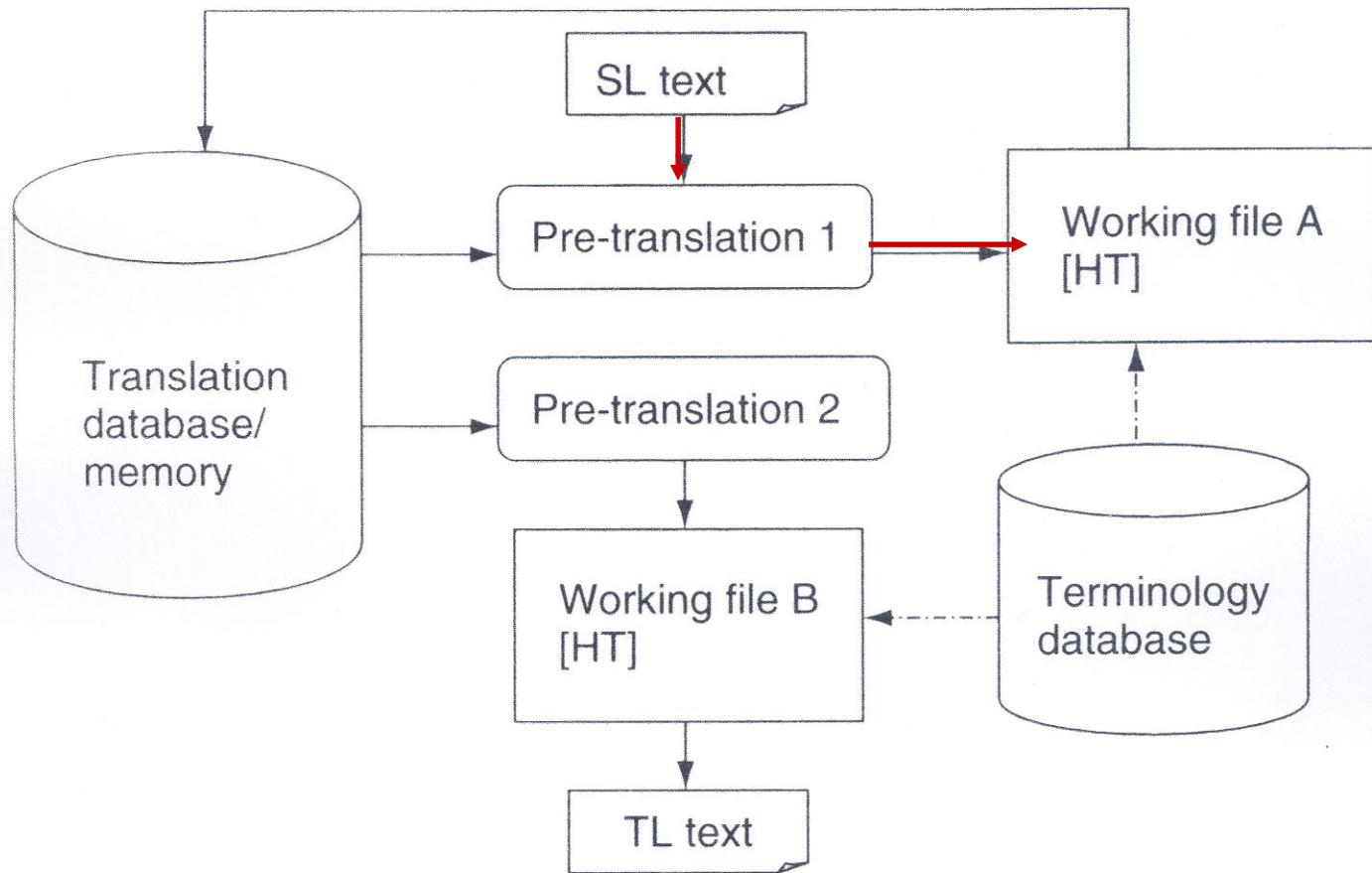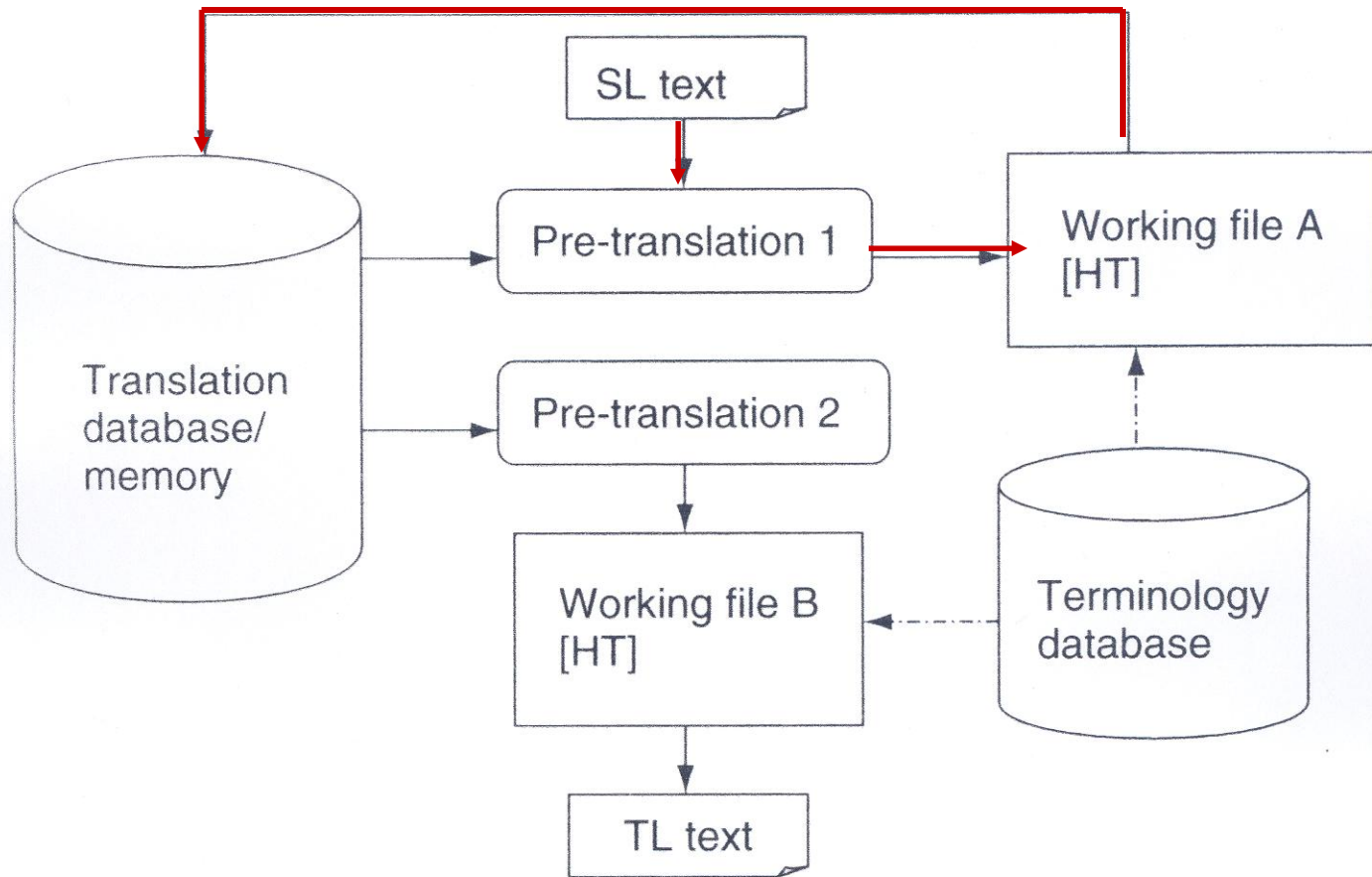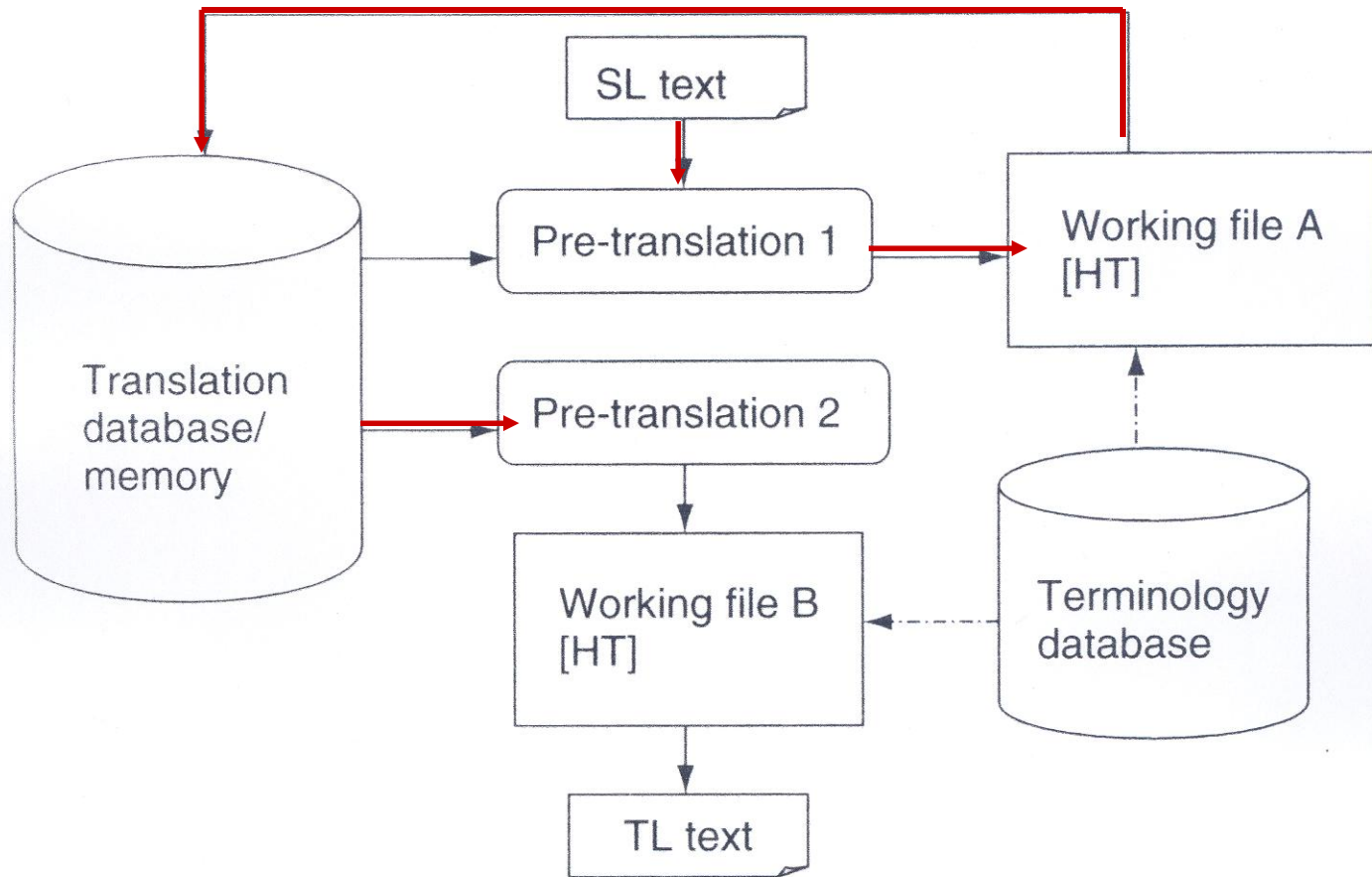
The translation workflow

SL = source language; TL = target language; HT = human translator

The translation workflow

SL = source language; TL = target language; HT = human translator

The translation workflow
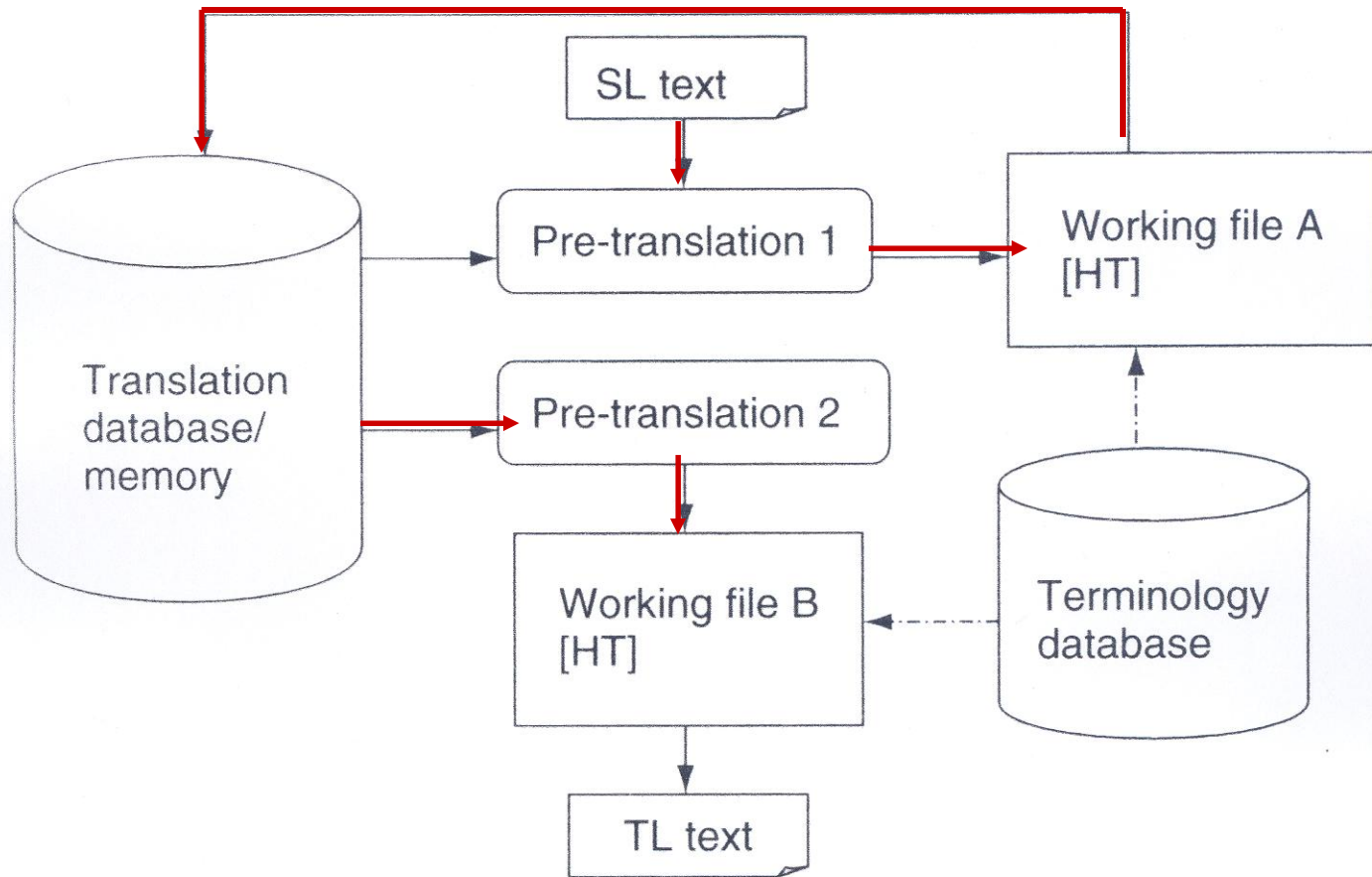

SL = source language; TL = target language; HT = human translator

The translation workflow

SL = source language; TL = target language; HT = human translator

The translation workflow
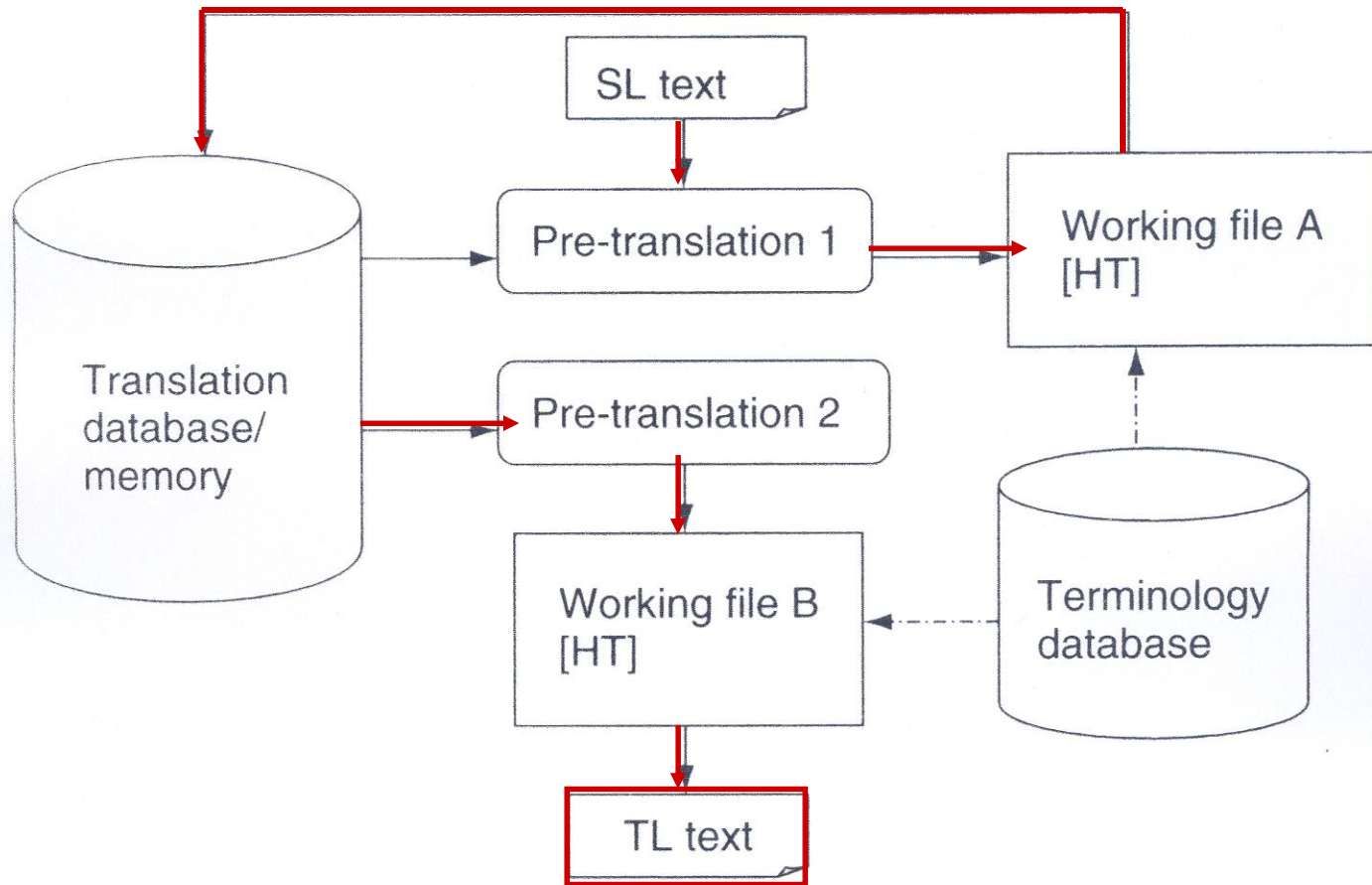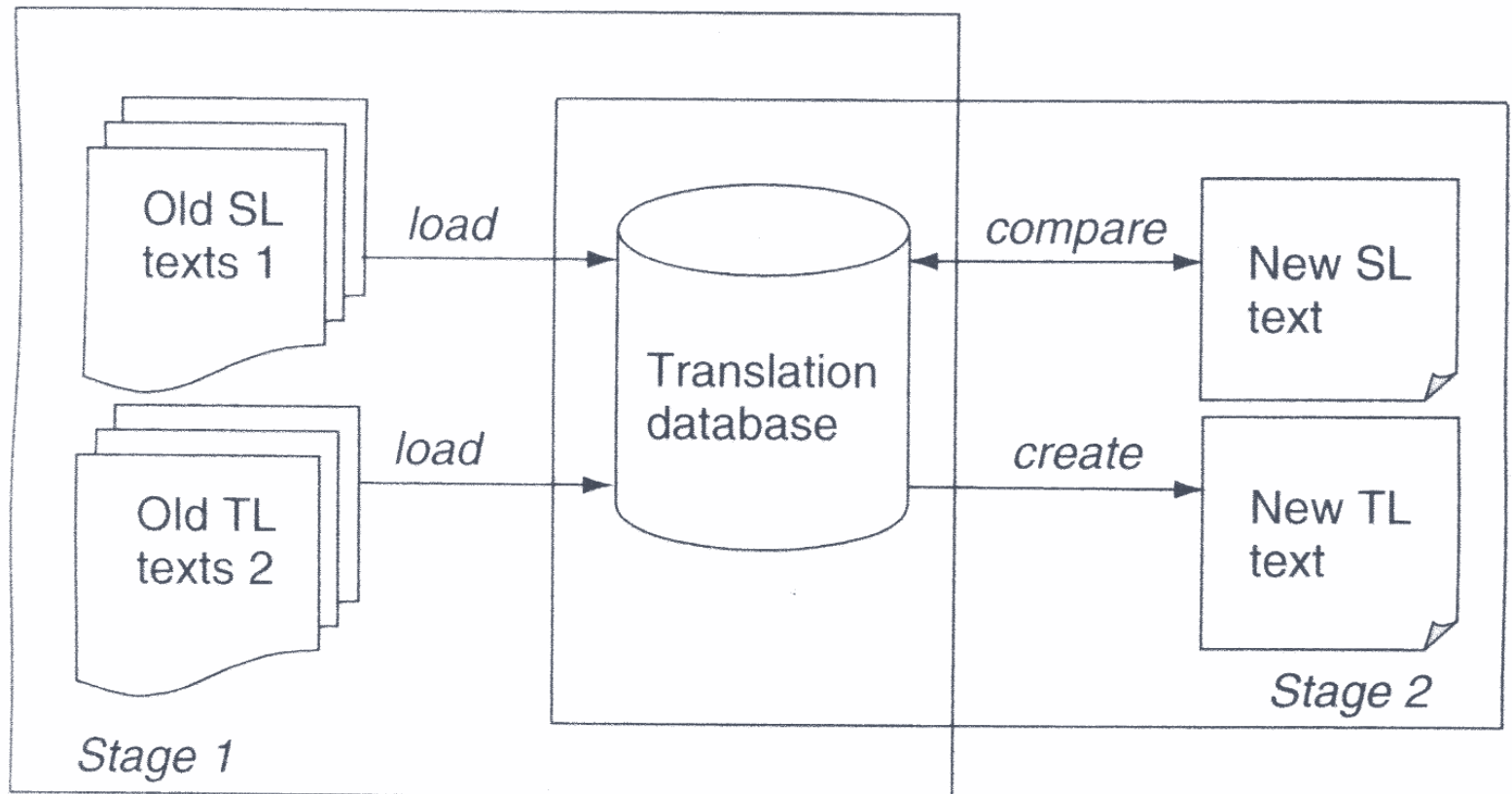

SL = source language; TL = target language; HT = human translator

The translation workflow

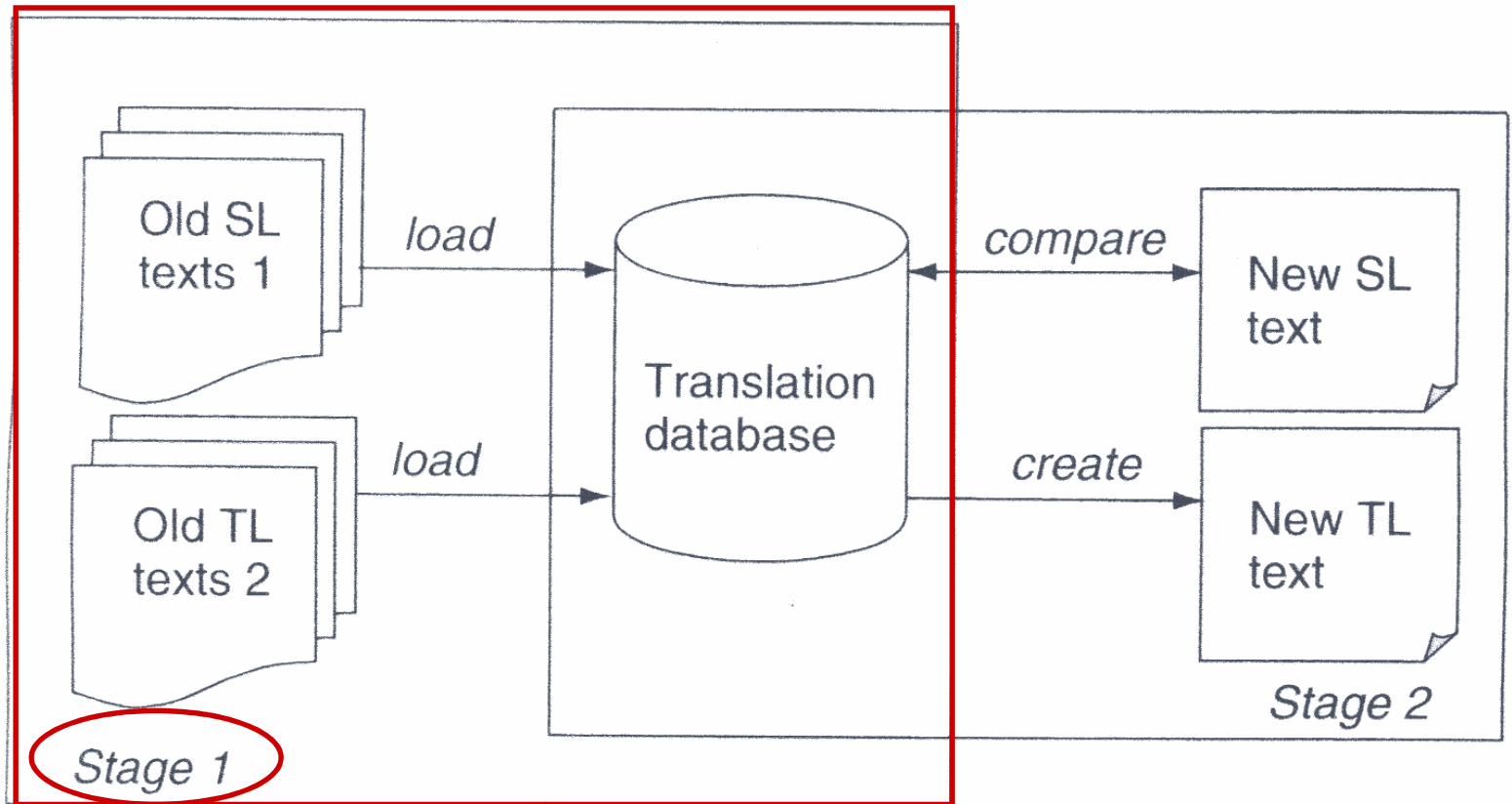SL = source language; TL = target language; HT = human translator

The translation workflow

SL = source language; TL = target language; HT = human translator

The translation workflow

SL = source language; TL = target language

The translation workflow

Old SL texts 1 — load → Translation database ← compare — New SL text

Old TL texts 2 — load → Translation database — create → New TL text

Stage 1

Stage 2
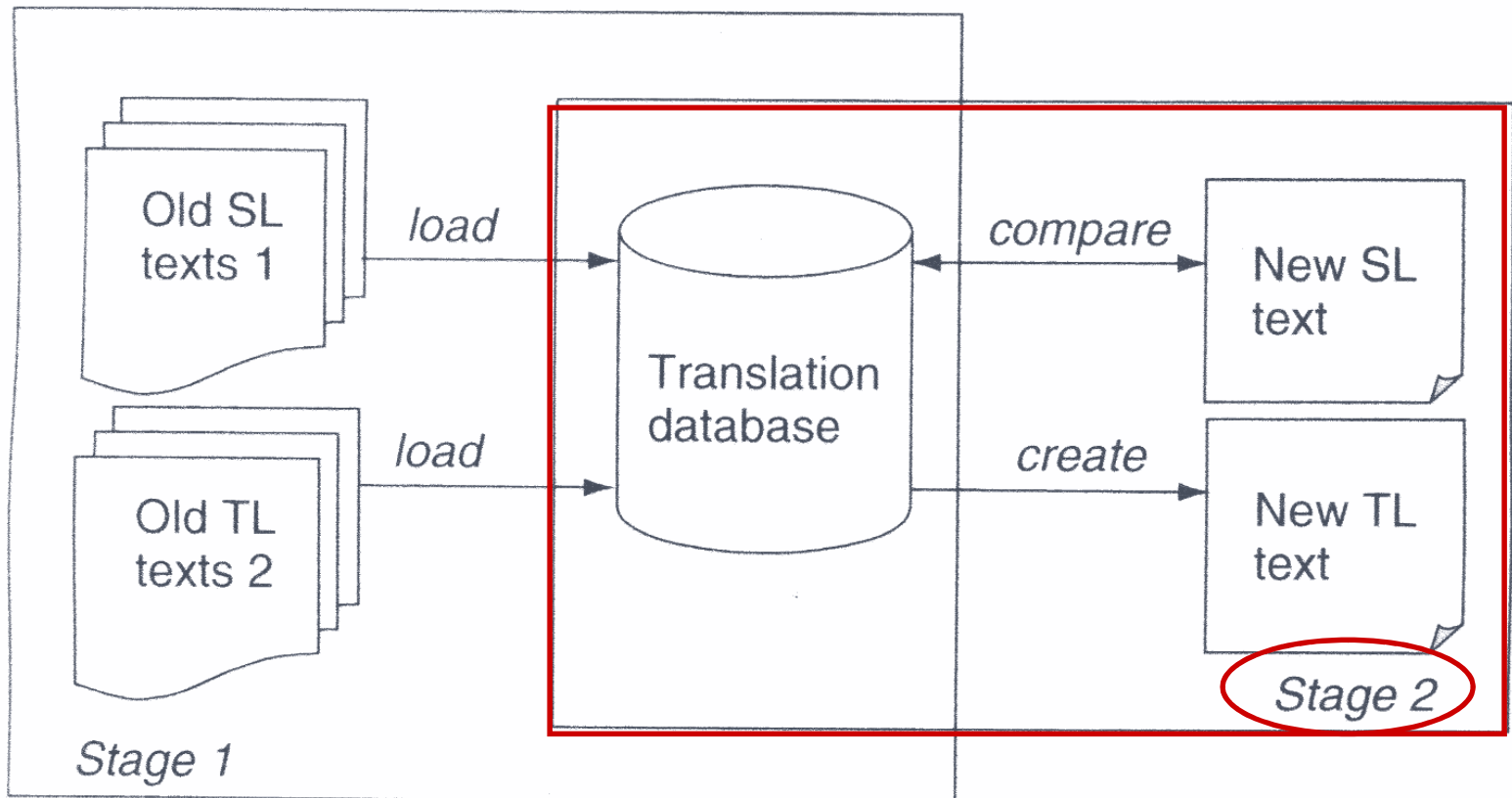
SL = source language; TL = target language

The translation workflow

SL = source language; TL = target language

How is a translation memory created?

- **A translation memory can be populated with data in one of two ways:**

- **FIRST:** Constant adding to the translation memory over the course of the translation.

- The translator creates a new, initially empty translation memory.

- During the translation, every new segment pair is stored in the translation memory and is available for the next translation.

How is a translation memory created?

- **A translation memory can be populated with data in one of two ways:**

- **SECOND:** Creating a TM from existing translations.

- The process of "Alignment" assigns segments in the source and target languages to each other from existing translations and saves them in the translation memory.

- This translation memory can then be used as a basis for future translations.

Advantages of using TMS (1)

- **Using a translation memory system has the following advantages:**

- The translation process is **faster** and update cycles are shorter. In the case of similar texts, such as software updates or documentation, the volume of no matches is continuously reduced.

- Changes to the source text can be reliably tracked and transferred to the translated versions

Advantages of using TMS (2)

- **Using a translation memory system has the following advantages:**

- The translated documentation is more **consistent** because previous wordings are suggested to the translator and because glossaries (e.g. in the form of Excel files) can be imported into the translation memory system.

- When using a TMS, the **quotation** is based on the word count of the source text, allowing the cost to be precisely calculated before translation starts.

Advantages of using TMS (3)

- It is possible to create as many translation memories as needed, for example for different subjects and/or clients, and/or language pairs.

Cost savings for your projects

- A source text analysis with a translation memory system provides the number of words in the four main categories of matches (Perfect Match/ Repetition/ Fuzzy Match/ No Match) and forms the basis of the project-specific quotation.

Cost savings for your projects

- **Perfect Matches or Repetitions:**
- As the text to be translated is identical or almost identical to the previous version, the translator only has to check whether the context is correct and adjust terminology where necessary.
- The text is fully proofread after translation and reviewed according to our quality assurance checklists.

Cost savings for your projects

- **Fuzzy Matches:**
- With 75-94% matches the already translated text of the previous version can largely be re-used.
- The translator edits the suggested text and adapts it to the new context.
- The price includes a full linguistic review.

Cost savings for your projects

- **No Matches:**

- If the match is lower than 75%, the text must be newly translated.

- 100 % of the price per word is applied.

- The price also includes a full linguistic review.

Main obstacles of TMS (1)

- **The main problems hindering wider use of translation memory managers include:**

- The concept of "translation memories" is based on the premise that sentences used in previous translations can be "recycled". However, a guiding principle of translation is that the translator must translate the *message* of the text, and not its component sentences.

- *Translation memory managers do not easily fit into existing translation or localization processes*. In order to take advantage of TM technology, the translation process must be redesigned.

Main obstacles of TMS (2)

- **The main problems hindering wider use of translation memory managers include:**

- Translation memory managers *do not support all documentation formats*, and filters may not exist to support all file types.

- There is a learning curve associated with using translation memory managers, and *the programs must be customized for greatest effectiveness.*

Main obstacles of TMS (3)

- **The main problems hindering wider use of translation memory managers include:**

- Full versions of many translation memory managers can cost from US$500 to US$2,500 per seat, which can represent a *considerable investment* (although lower cost programs are also available).

- However, some developers produce free or low-cost versions of their tools with reduced feature sets that individual translators can use to work on projects set up with full versions of those tools.

Main obstacles of TMS (4)

- **The main problems hindering wider use of translation memory managers include:**

- The costs involved in *importing the user's past translations into the translation memory database*, training, as well as any add-on products may also represent a considerable investment.

- *Maintenance of translation memory databases* still tends to be a manual process in most cases, and failure to maintain them can result in significantly decreased usability and quality of TM matches.

Thank you !