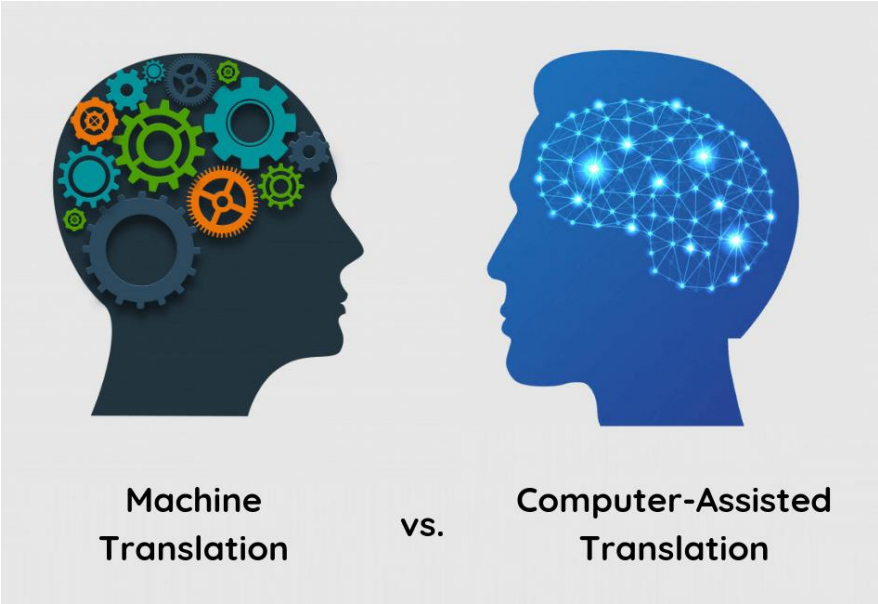
Machine Translation
&
Computer-Aided (Assisted) Translation



Week 3 & 4

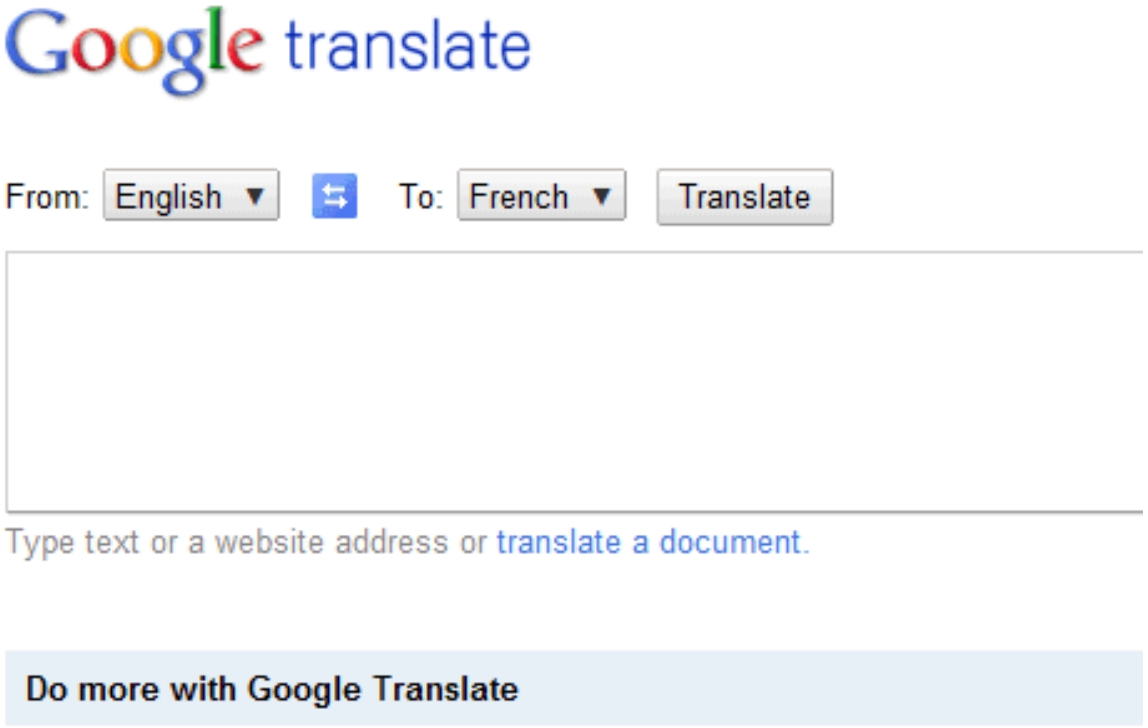
Lect. Dr. Fatma Toköz

What is Machine Translation?

- A sub-field of computational linguistics
- It investigates the use of computer software to translate text or speech in between languages

Machine translation (MT) is the application of computers to the task of translating texts from one language to another.

Famous Example of Machine Translation



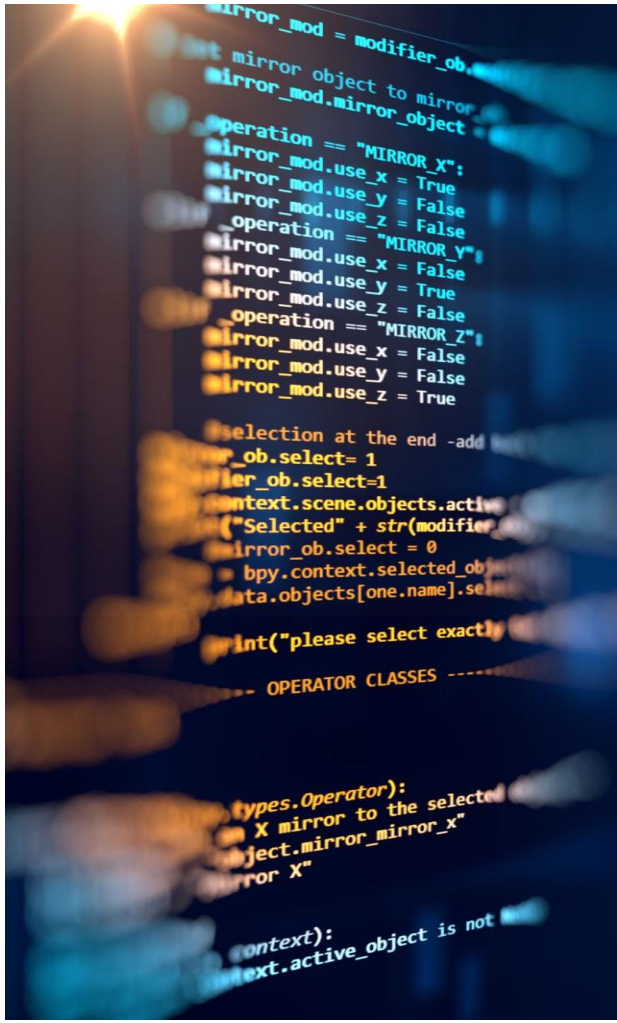
Is it reliable?

Translation process

The translation process can be stated as:

- Decoding the meaning of the source text, and
- Re-encoding this meaning in the target language

=Analyze the source text for meaning and generate the meaning in target language.



Five types of knowledge used in the translation process:

- Knowledge of the source language, which allows us to understand the original text.
- Knowledge of the target language, which makes it possible to produce a coherent text in that language.
- Knowledge of equivalents between the source and target languages.
- Knowledge of the subject field as well as general knowledge, both of which aid comprehension of the source language text.
- Knowledge of socio-cultural aspects, that is, of the customs and conventions of the source and target cultures.

Because, Machine Translation (MT) fails to convey the specific features and meaning of a culture

- For example;
  1) Ben sevdim, eller aldı → *I loved it, hands took*
  2) Gelecekleri varsa, görecekleri de var → *If they have a future, they have something to see.*

In other words, we need…

- ***in-depth knowledge*** of both the grammar, semantics, syntax, idioms and the like of the source language, as well as the culture of its speakers
- the same in-depth knowledge of target language is needed to re-encode the meaning in the target language

The challenge

- How to program a computer to "understand" a text as a human being does
- Also, to "create" a new text in the source language that "sounds" as if it has been written by a human

Some issues

- Languages have different word orders/structures.
- Some languages are morphologically complex.
- Some languages do not have determiners.
- Identifying and finding equivalents of idioms, collocations, phrasal verbs etc.
- Lexical gaps
- Lexical ambiguity
- Structural ambiguity

History

- The Georgetown experiment in 1954 [after the second world war] involved fully automatic translation of more than sixty Russian sentences into English.
- The experiment was a great success and ushered in an era of significant funding for machine translation research.
- ALPAC = Automatic Language Processing Advisory Committee

History (cont’d.)

- Starting in the late 1980s, as computational power increased and became less expensive, more interest began to be shown in statistical models for machine translation.
- Today, there are many software programs for translating natural language, several of them online, such as the SYSTRAN system which powers both Google translate and the AltaVista's Babelfish.

Compromise

- It would be absurd to claim that a machine could produce a target text of the same quality as that of a human being
- A machine can do the first stage of translation automatically and then human beings can revise and edit the translation.
  - Machine/Computer Aided Translation systems
  - Human Aided Machine Translation systems

Machine Translation Today

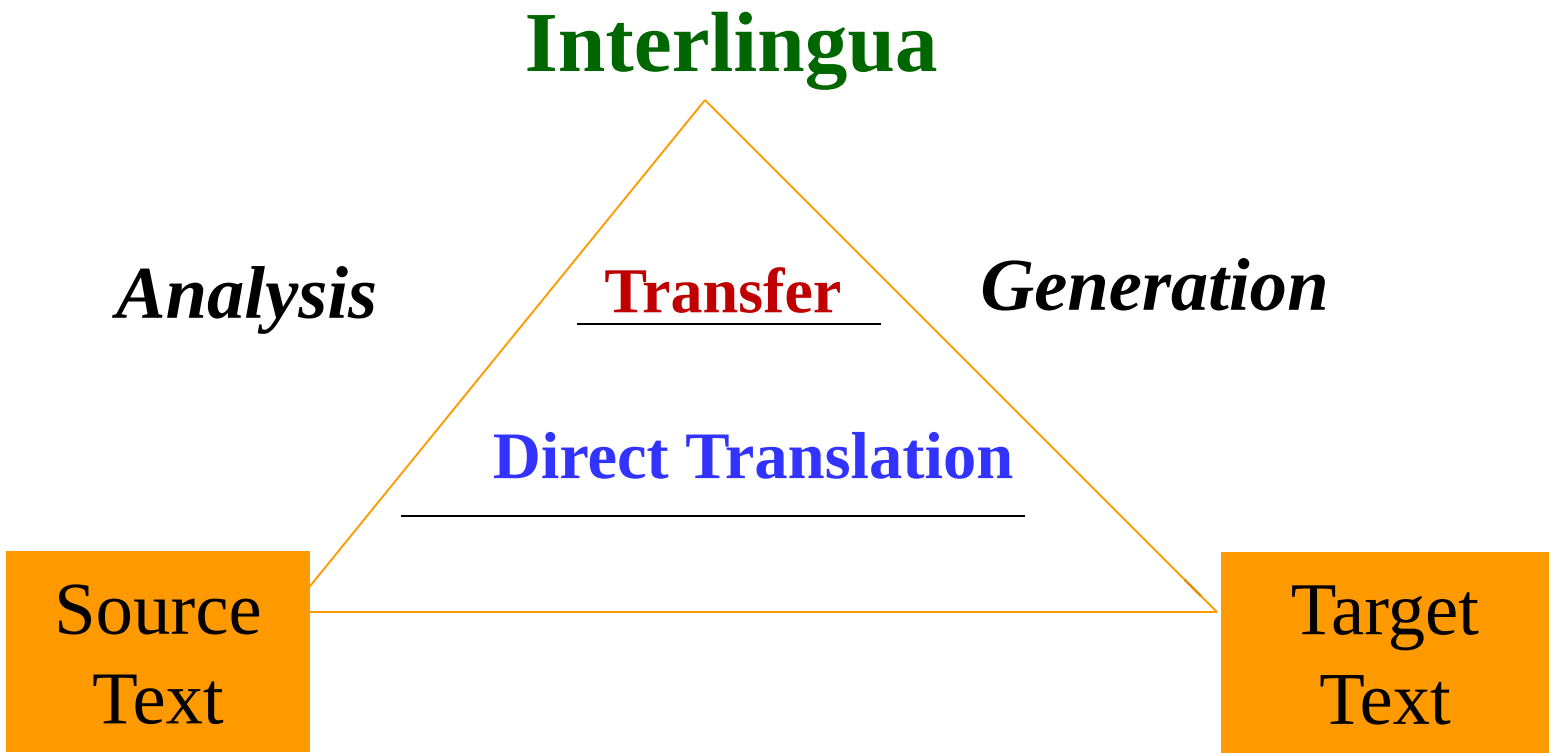
- The product of machine translation is sometimes called a "gisting translation”.
- MT will often produce only a rough translation that will at best allow the reader to "get the gist" of the source text.
- It may **not** convey a complete understanding of it.
- The user may find the raw translation sufficiently useful as it is.
- Despite their inherent limitations, MT programs are currently used by various organisations around the world.

Very useful for…

- Translating user manuals
- Translating UN, EU documents
- Translating instructions
- Translating web pages

- BUT, ***limited domain***

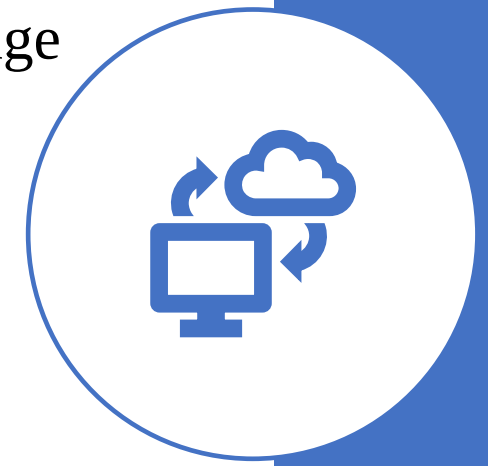
Approaches



Source: http://www.multilingual.com/articleDetail.php?id=1082

Types of Machine Translation

- **Rule-based**
  - **dictionary based** - EASIEST - also called word-based or word-for-word approach
  - **transfer rule approach** - try to use the meaning of source language to output the same meaning in the target language
  - **interlingual** - translate via a language neutral intermediate form
- **Statistical**
  - calculate most likely translation by using pairs of translation (bilingual text corpora)
  - **neural networks** - HARDEST - state of the art, really a variation of statistical
- **Example-based**
  - use simple examples in one language to generate the same thing in another language



Evaluation

- Evaluating the MT system is essential.
- There are various methods for evaluating the performance of machine translation systems:
  - The oldest method is by using human judges to tell the quality of a translation,
  - Automated methods include BLEU, NIST and METEOR.

Online available MT software

- Alpha Works Emerging Technologies

http://www.alphaworks.ibm.com/aw.nsf/html/mt

- Systran Language Translation Technologies

http://www.systransoft.com/index.html

- Free Professional Translation

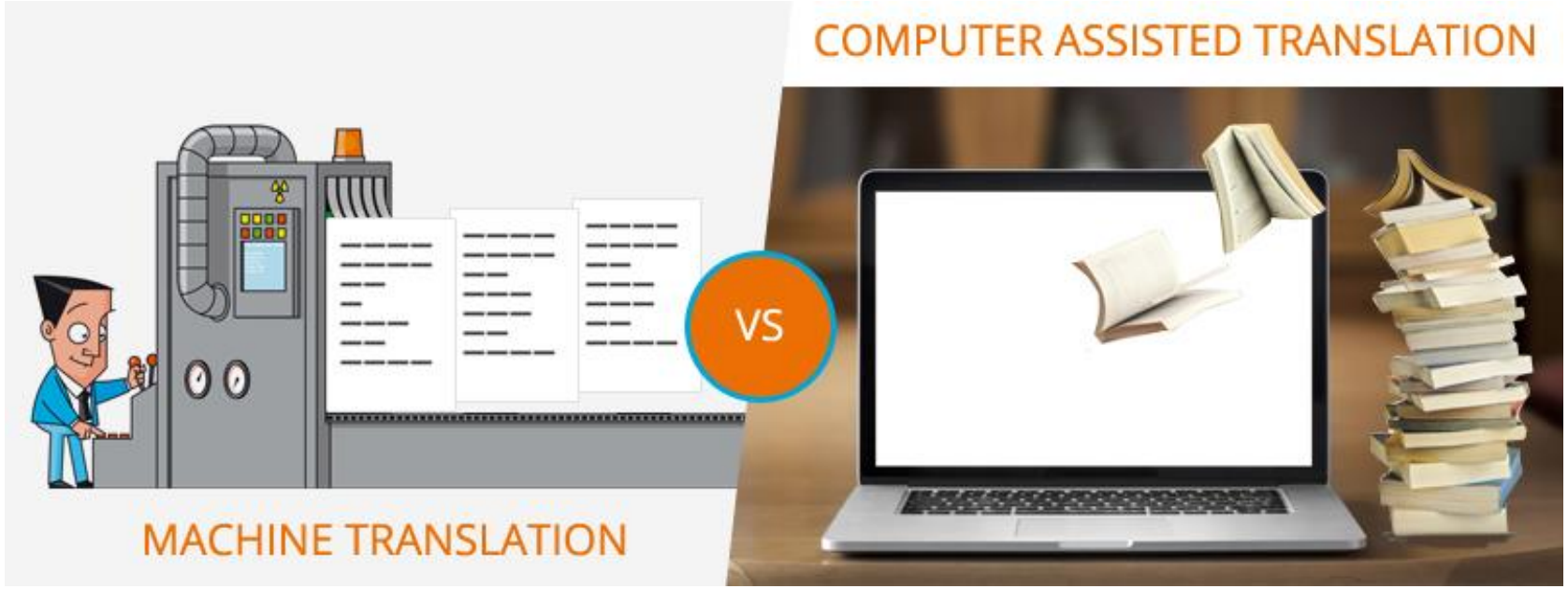
http://www.freetranslation.com/

- 100 Links to Online Translators and Machine Translation Software

http://www.bultra.com/mtlinks.htm

Conclusion

- Relying on machine translation exclusively ignores the fact that
  - communication in human language is context-embedded and that
  - it takes a person to comprehend the context of the original text with a reasonable degree of probability.
- even purely human-generated translations are prone to error.
  - such translations must be reviewed and edited by a human



Computer- Aided Translation

What do you know about CAT-Tools?

Computer Assisted Translation

- Computer-aided translation (CAT), also referred to as machine-assisted translation (MAT) or machine-aided human translation (MAHT)
- It is the use of software to assist a human translator in the translation process
- The translation is created by a human, and certain aspects of the process are facilitated by software
- CAT is in contrast with machine translation (MT), in which the translation is created by a computer, optionally with some human intervention.

Why CAT?

- Translation companies require experience with CAT tools
- CAT training programs, real or online, virtually non-existent
- Most translators must learn on own
- Almost all source texts are given in digital format; most prevalent format: PDF

Computer Aided Translation Tools

- CAT tools can support the translator by avoiding repetitive work, automating terminology search activities, and reusing previously translated texts with the assistance of translation memory (TM).
- A CAT tool makes it easier to translate a document between languages by using a number of features such as:
  - translation memory
  - automatic translation following glossaries
  - automatic translation quality checks
  - machine translation, and other automation technology

Computer- Aided Translation

CAT ≠ MT

CAT is not the same as Machine Translation

- **MT performs the translation task for the translators**
- **CAT Tools support the translators in performing their tasks**

Target Audience - Who is/should be using CAT Tools?

- Professional Translators
- Translation departments in companies (manufacturing, banking, finance, administration, …)
- Language Service Providers (Translation agencies)
- Freelance Translators

How can CAT Tools help to increase efficiency and to reduce cost thus allowing for a higher turn- around?

CAT Tools contain a translation memory (TM)

Translations are saved in the TM together with the source text.

When a sentence occurs that has already been translated (or a similar sentence) the stored translation are suggested to the translator.

Translators can use these suggestions, adapt them if necessary or decide to translate from scratch.

Typical components of CAT-Tools

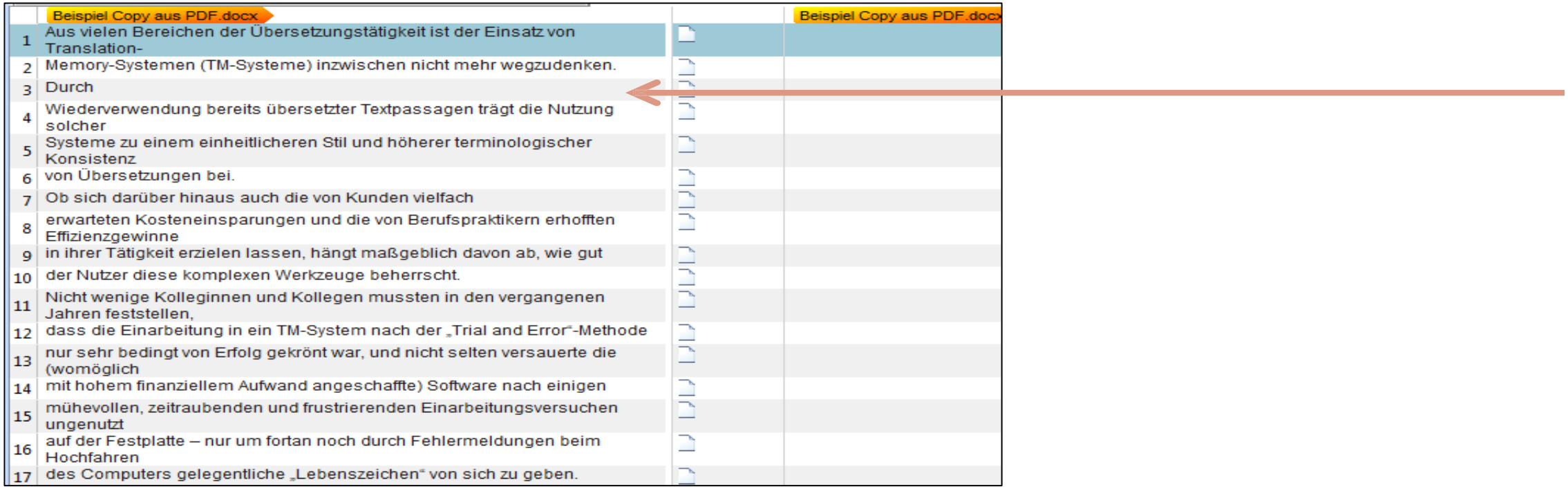
Translation Memory	• Database in which translations are stored (typically as sentence pairs)
Termbase	• Database in which terminology is stored and managed
Editor	• Writing environment to create and to edit translations
Alignment	• Application to recycle legacy data
Filter Tool	• to convert various file formats in translatable file formats
Project or Workflow management	

Translation memory (TM)

- (TM) programs store previously translated source texts and their equivalent target texts in a database and retrieve related segments during the translation of new texts.
- Such programs split the source text into manageable units known as "segments"
- A source-text sentence or sentence-like unit (headings, elements in a list) is considered a segment.
- As the translator works through a document, the software displays each source segment in turn and provides a previous translation for re-use, if the program finds a matching source segment in its database.
- If it does not, the program allows the translator to enter a translation for the new segment.

Segmentation

Examples for segmentation issues: Copy & Paste from PDF



	Beispiel Copy aus PDF.docx		Beispiel Copy aus PDF.docx
1	Aus vielen Bereichen der Übersetzungstätigkeit ist der Einsatz von Translation-		
2	Memory-Systemen (TM-Systeme) inzwischen nicht mehr wegzudenken.		
3	Durch		
4	Wiederverwendung bereits übersetzter Textpassagen trägt die Nutzung solcher		
5	Systeme zu einem einheitlicheren Stil und höherer terminologischer Konsistenz		
6	von Übersetzungen bei.		
7	Ob sich darüber hinaus auch die von Kunden vielfach		
8	erwarteten Kosteneinsparungen und die von Berufspraktikern erhofften Effizienzgewinne		
9	in ihrer Tätigkeit erzielen lassen, hängt maßgeblich davon ab, wie gut		
10	der Nutzer diese komplexen Werkzeuge beherrscht.		
11	Nicht wenige Kolleginnen und Kollegen mussten in den vergangenen Jahren feststellen,		
12	dass die Einarbeitung in ein TM-System nach der „Trial and Error"-Methode		
13	nur sehr bedingt von Erfolg gekrönt war, und nicht selten versauerte die (womöglich		
14	mit hohem finanziellem Aufwand angeschaffte) Software nach einigen		
15	mühevollen, zeitraubenden und frustrierenden Einarbeitungsversuchen ungenutzt		
16	auf der Festplatte – nur um fortan noch durch Fehlermeldungen beim Hochfahren		
17	des Computers gelegentliche „Lebenszeichen" von sich zu geben.		

In a typical CAT Tools (SDL Trados Studio)

Termbase / Terminology management

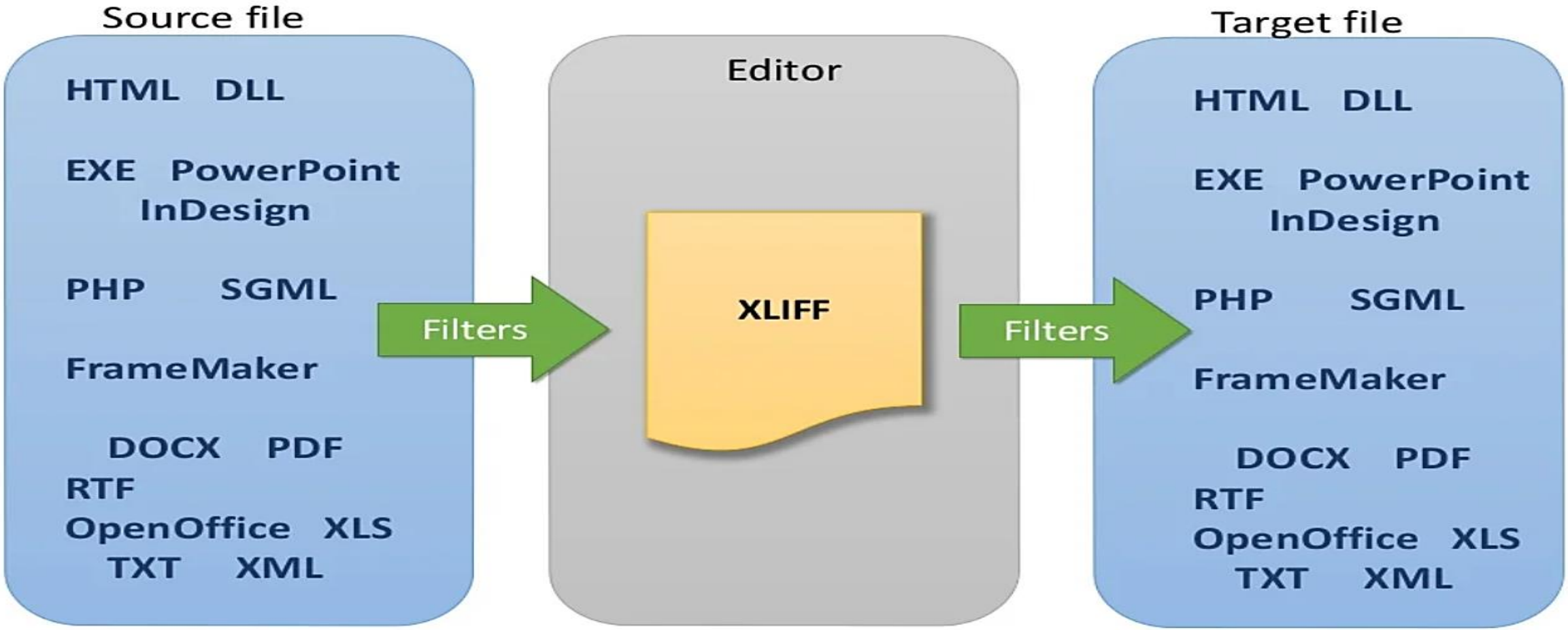
- It is similar to dictionaries or glossaries that are built from frequently occurring words or phrases, such as technical terms and brand names.
- It is used to pre-translate recurring words and phrases, and to assist translators in maintaining consistency.
- Terminology management is a set of activities that ensures correct terms are used consistently in all materials.

Alignment

- It is the process of matching segments in the source text with their translated renditions in order to create new translation memory files.
- Simply, alignment is a way of making use of existing translation materials.
- How does translation alignment work?
  - The alignment tool matches the source and target language files side-by-side, to determine which pairs belong together

The Editor

Editor displays both the document to be translated and its translation **side by side**



CAT Tool

Types of installation

Cloud-based
(Software as a Service)

desktop or client-server mode

Desktop-based CAT Tools

Cloud-based CAT Tools

matecat

Google Translator Toolkit

WORDFAST anywhere

Collaborative translation environment (Cloud-based SaaS)

ABBYY SmartCAT

xtm cloud
Better Translation Technology...

ontram
ONLINE TRANSLATION MANAGEMENT FRAMEWORK

GEOWORKZ

WORDBEE

Localizr.com

Lingotek

collaborative translation tool
crowdin
beta

memsource
TECHNOLOGIES

tx | Transifex BETA
the open translation platform

Conclusion

- ***CAT Tools are useful for:***
  - Enhanced quality
  - Improving productivity
  - Keeping consistency
  - Connectivity

Thank you for listening!

Questions and Comments?

ftokozgoktepe@gmail.com