AAB UNIVERSITY

Lecture 3

The role of technology in translation process

arianit.maraj@universitetiaab.com

Tel: 044 425 159

Machine Translation

Introduction

- Though research in Machine Translation (MT) has already celebrated its fiftieth birthday, understanding of its successes is still minimal

- The increase in availability of Machine Translation software due to the globalization of the Internet has had little impact.

- User's knowledge of the complexities behind translating remains limited and judgments are based on one off personal experiences.

Overview

- ## What is Machine Translation (MT)?
  - Automated system
  - Analyzes text from Source Language (SL)
  - Produces "equivalent" text in Target Language (TL)
  - Ideally without human intervention

Source Language → 🖥️ → **Target Language**
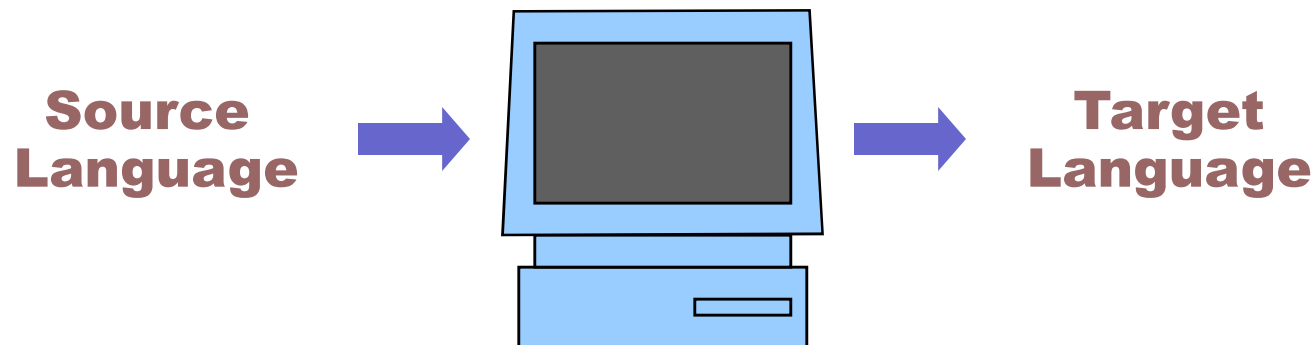
Current Drivers

- Continued improvements in computing power
- Globalization
- Availability of translated works
- Advances in search and relational database technology

Definition of MT

- The European Association for Machine Translation gives the following definition for MT:

  - "*Machine translation (MT) is the application of computers to the task of translating texts from one natural language to another* "

Types of Machine Translation

Unassisted Machine Translation:

- Unassisted MT takes pieces of text and translates them into output for immediate use with no human involvement

- The result is unpolished text and gives only a gist of the source.

- The ultimate aim of this type of MT is sometimes known as Fully Automatic High Quality Translation (FAHQT), perfect translation created solely by a computer

- Examples of this form of MT include IBM alphaworks native search, Babel Fish 2020 , Worldlingo and Dragon systems

Types of Machine Translation

Assisted machine Translation:

- Assisted MT uses a human translator to clean up after, and sometimes before, translation in order to get better quality results.

- Usually the process is improved by limiting the vocabulary through use of a dictionary and the types of sentences/grammar allowed.

- The use of a 'controlled language' has been fairly successful. Some systems have also been set up to learn from corrections.

Quality Levels using MT

Quality levels that should be considered are:

- **Shallow vs. Deep -** Words, phrases, ordering vs. Context, reality etc..

- **Speed vs. Quality -** Real-time vs. High quality

- **Degree of human intervention -** Pre-editing, post-editing

Quality Levels using MT

- **Example: Controlled Language**

- Often used for translation of technical documentation to many target languages

- May include an engine to determine whether a sentence conforms to a controlled language specification (rewrite opportunities)

Methodologies of MT

Two methodologies of MT are:

- Rule-based

- Statistical

Rule based MT

- **Rule-Based Machine Translation** (RBMT; also known as "Knowledge-Based Machine Translation" is a general term that denotes MT systems based on **linguistic information** about source and target languages basically retrieved from **dictionaries** and **grammars**

- **RBMT covers** the main semantic, morphological, and syntactic regularities of each language respectively.

- Having input sentences (in SL), an RBMT system generates them to output sentences (in TL) on the basis of morphological, syntactic, and semantic analysis of both the source and the target languages involved in a concrete translation task.

Methodologies of MT

- **Rule-based**
  1. Lexical – word by word using dictionaries
  2. Grammatical – decompose sentence structure
  3. Semantic – look for meaning of sentence
  4. Statistical/example-based - Searches existing translations for similar phrases

How does MT work?

Rules-based

- Rules-based systems use a combination of language and grammar rules plus dictionaries for common words.

- Specialist dictionaries are created to focus on certain industries or disciplines.

- Rules-based systems typically deliver consistent translations with accurate terminology when trained with specialist dictionaries.

Rule based MT

- There are three different types of RBMT systems:
- **Direct Systems** (Dictionary Based Machine Translation); map input to output with basic rules.
- **Transfer BMT Systems** (Transfer Based Machine Translation); employ morphological and syntactical analysis.
- **Interlingual BMT Systems** (Interlingua); use an abstract meaning

Transfer Based MT

- Transfer-based MT needs n(n-1) transfer modules for n languages.
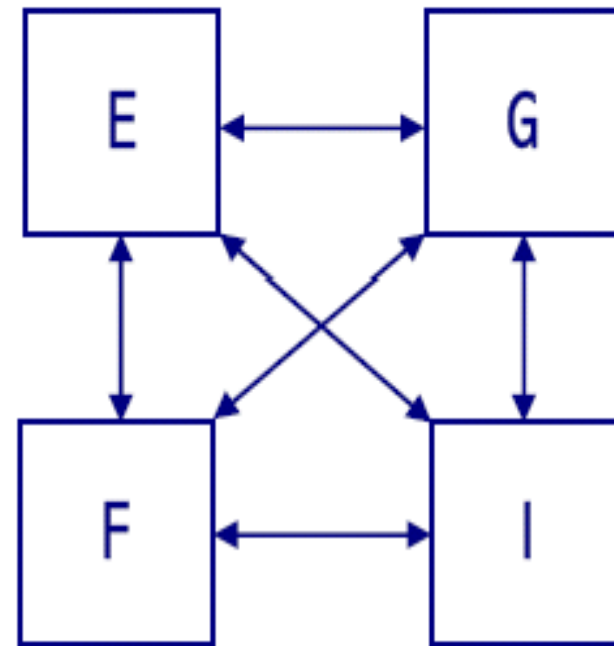
Interlingua MT

- An alternate to the transfer component is an Interlingua, a type of intermediate language
- A translation is made from the **SL** into the **Interlingua** and then into the **TL (SL-IL-TL)**
- The benefits of using an Interlingua are that only one part is required for each language and therefore further languages can be added easily

Transfer based vs. Interlingua MT

- **Transfer-based** and **Interlingua-based** MT have the same idea: to make a translation it is necessary to have an intermediate representation that captures the "meaning" of the original sentence in order to generate the correct translation.

- In **interlingua-based** MT this intermediate representation must be independent of the languages in question, whereas in **transfer-based** MT, it has some dependence on the language pair involved.

How does machine translation work?

Statistical

- Statistical systems have no knowledge of language rules.

- Instead they "learn" to translate by analyzing large amounts of data for each language pair.

- They can be trained for specific industries or disciplines using additional data relevant to the sector needed.

- Typically statistical systems deliver more fluent-sounding but less consistent translations.
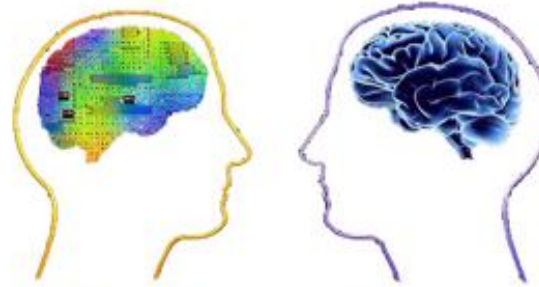
Machine vs Human Translation

MACHINE TRANSLATION TODAY

Where are we now?

- Huge potential/need due to the internet, globalization and international politics.

- Quick development time due to Statistical Machine Translation (SMT), the availability of parallel data and computers.

- Translation is reasonable for language pairs with a large amount of resources.

- Start to include more "minor" languages.

What is MT good for?

- Rough translation: web data

- Computer-aided human translation

- Translation for limited domain

What is MT good for?

- **Machines beat humans at:**
  - **Speed:** much faster than humans
  - **Memory:** can easily memorize millions of word/phrase translations.
  - **Manpower:** machines are much cheaper than humans
  - **Fast learner:** it takes minutes or hours to build a new system.
  - Never complain, never get tired, …

Interest in Machine Translation (1)

- **Commercial interest:**
  - U.S. has invested in machine translation (MT) for intelligence purposes
  - MT is popular on the web—it is the most used of Google's special features
  - EU spends more than $1 billion on translation costs each year.
  - (Semi-)automated translation could lead to huge savings

Interest in Machine Translation (2)

- **Academic interest:**
  - One of the most challenging problems in NLP (Natural Language Processing) research
  - Requires knowledge from many NLP sub-areas, e.g., lexical semantics, syntactic parsing, morphological analysis, statistical modeling,…
 - Being able to establish links between two languages allows for transferring resources from one language to another

Natural Language Processing

- **Natural language processing** (**NLP**) is a field of computer science, artificial intelligence, and linguistics concerned with the interactions between computers and human (natural) languages.

- NLP is related to the area of **Human–Computer** interaction

- NLP parses sentences and determines their underlying meaning in order for databases to answer SQL queries entered in the form of a question

Goals & Uses

- Translating
- Summarizing
- Communicating
- Pre-editing
- Grammar analysis
- Analyzing text
- Understanding text and images

Languages on the Internet
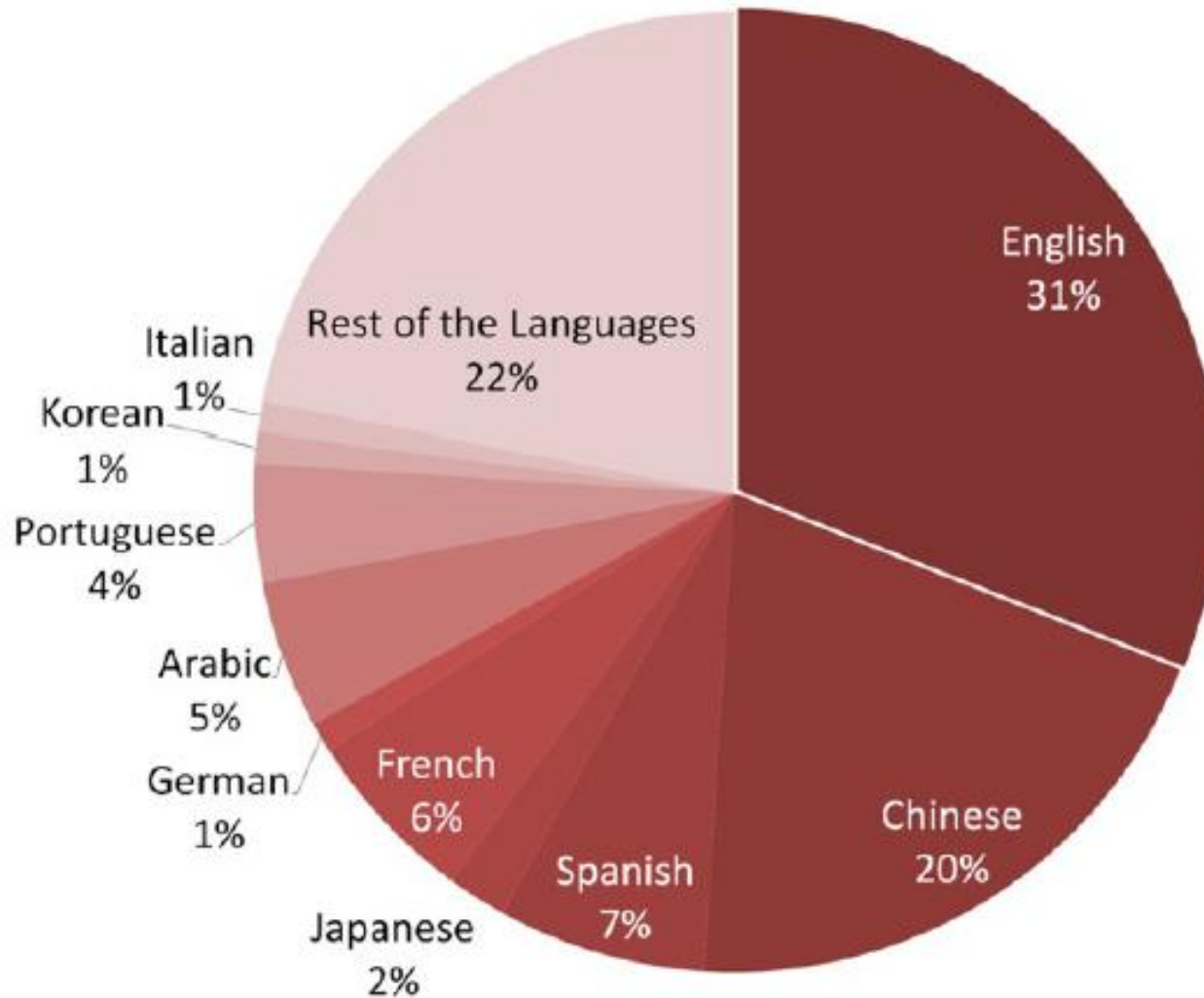
Why Machine Translation is Difficult?

- A single word can have more than one meaning

- Lexical gaps: single-word concepts with no simple translation

- Idioms

- Different languages use different syntactic structures

- Some syntactic structures are not possible in some languages

Why Machine Translation is Difficult?

- We need to find the correct interpretation

- Literal translation does not produce fluent text

- Literal translation does not preserve semantic information

- Literal translation does not preserve pragmatic information

Human and Machine Translation

- HT and MT differ in two main points:

  - **Mode of process**
  - **Mode of product**

- based on different specifications and theoretical positions
- both modes are used for comparison

Human and Machine Translation

Mode of process

- By comparing the modes of process, you can:
    1. gain knowledge about the respective stages and intersections
    2. make decisions about choices of alternative methods
    3. … and about new designs of translation methods

Human and Machine Translation

Mode of product

- By comparing the modes of product you can:
  1. check the appropriateness of the translation
  2. figure out the most efficient method

→ the MT product must be usable in the same way as the human product

→ secure a basis of equality

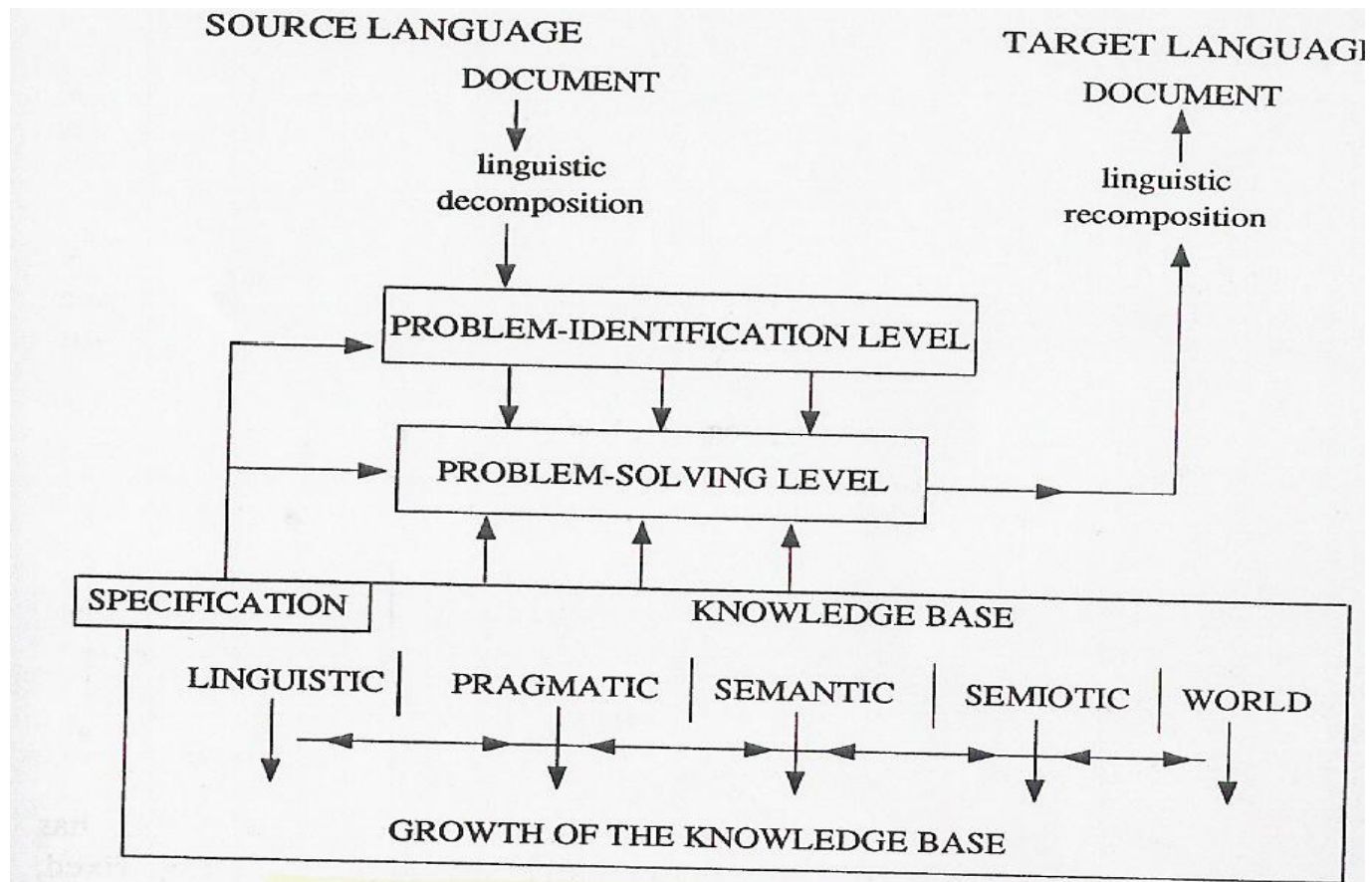# Human and Machine Translation

Another criterion for comparison:

- text input must be a constant so that the products are comparable
- → help to formulate guidelines for HT or MT texts

Human and Machine Translation
- translation processes -

Translation as problem solving

SOURCE LANGUAGE
DOCUMENT

linguistic
decomposition

TARGET LANGUAGE
DOCUMENT

linguistic
recomposition

PROBLEM-IDENTIFICATION LEVEL

PROBLEM-SOLVING LEVEL

SPECIFICATION

KNOWLEDGE BASE

LINGUISTIC | PRAGMATIC | SEMANTIC | SEMIOTIC | WORLD

GROWTH OF THE KNOWLEDGE BASE

Human and Machine Translation
- translation processes -

- Four major steps:

  - SL linguistic de-composition

  - Problem identification at the SL linguistic level

  - Problem solution at the TL linguistic level (knowledge base)

  - TL linguistic re-composition
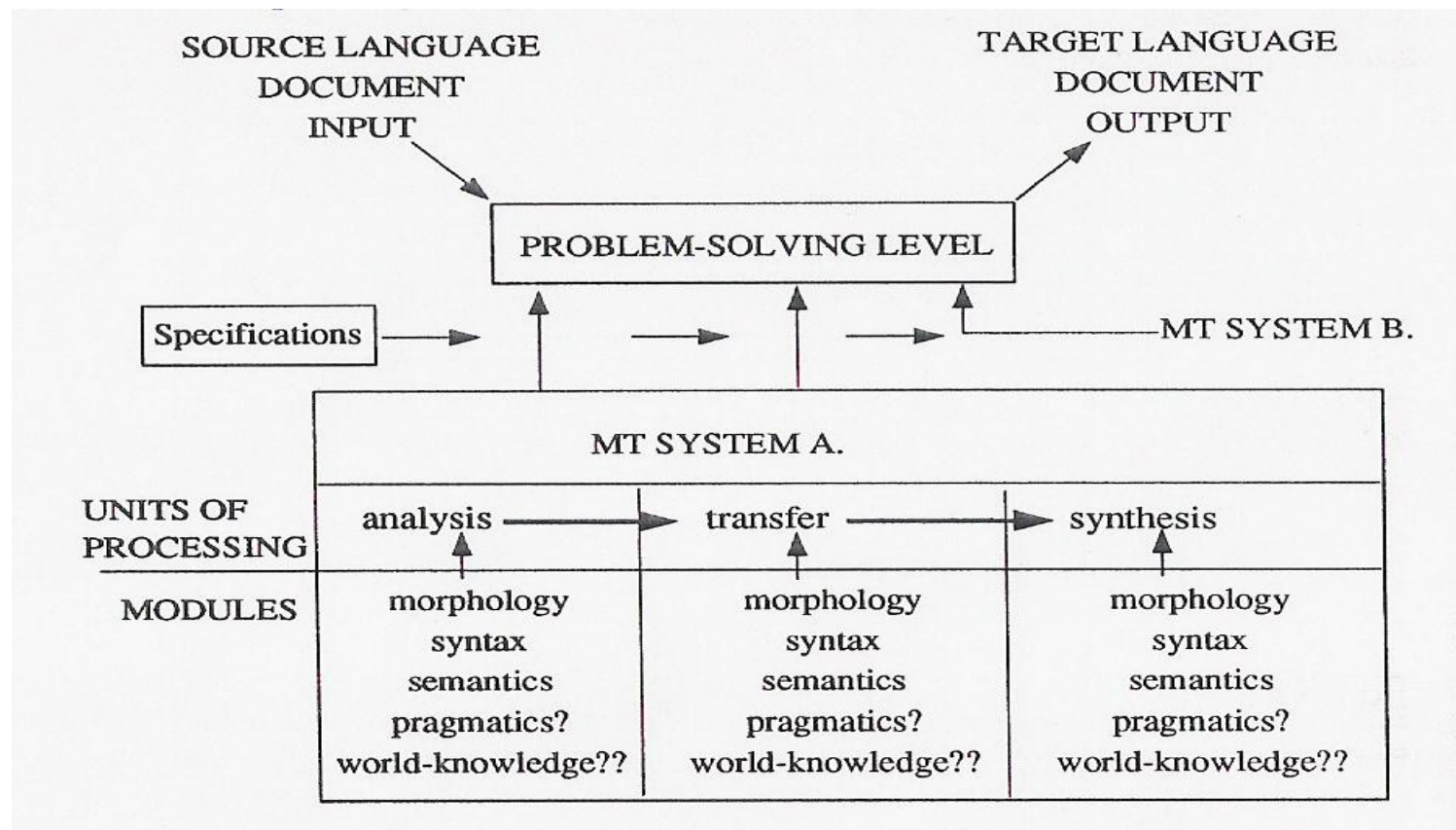
Human and Machine Translation
- translation processes -

- **Characteristics of HT:**

  - Knowledge base is flexible

  - Problems can be transferred

  - Intuition/experience of the translator

  - Knowledge base expands constantly

Human and Machine Translation
- translation processes -

MT model of problem solving

Human and Machine Translation
- translation processes -

- **Characteristics of MT:**
  - Knowledge base is relatively limited and rigid
  - Has fixed and pre-established connections
  - Limited possibility of transferring problems
  - Less semantic and pragmatic level experience
  - Lack of essential world-knowledge

Human and Machine Translation - translation processes

Major levels of comparison

Human modules	Machine modules
Comprehension	Analysis
Matching	Transfer
Writing	Generation/Synthesis

Human and Machine Translation
- translation processes –

- **Comprehension vs. Analysis**

Human	Machine
adapts innovations	works retrospectively
high amount of interpretative capacity	limited amount of interpretative capacity

Human and Machine Translation
- translation processes –

- **Matching vs. Transfer**

Human	Machine
compensation of items which cannot be matched in traditional ways	equivalents cannot be pre-planned or incorporated

Human and Machine Translation - translation processes

- **Writing vs. Generation/Synthesis**

Human	Machine
can respond to syntactic or lexical innovations or deviations	works prospectively

Human and Machine Translation
- translation products -

- Products can be compared with regard to:
  - the nature of the output language
  - the produced text

Human and Machine Translation
- translation products -
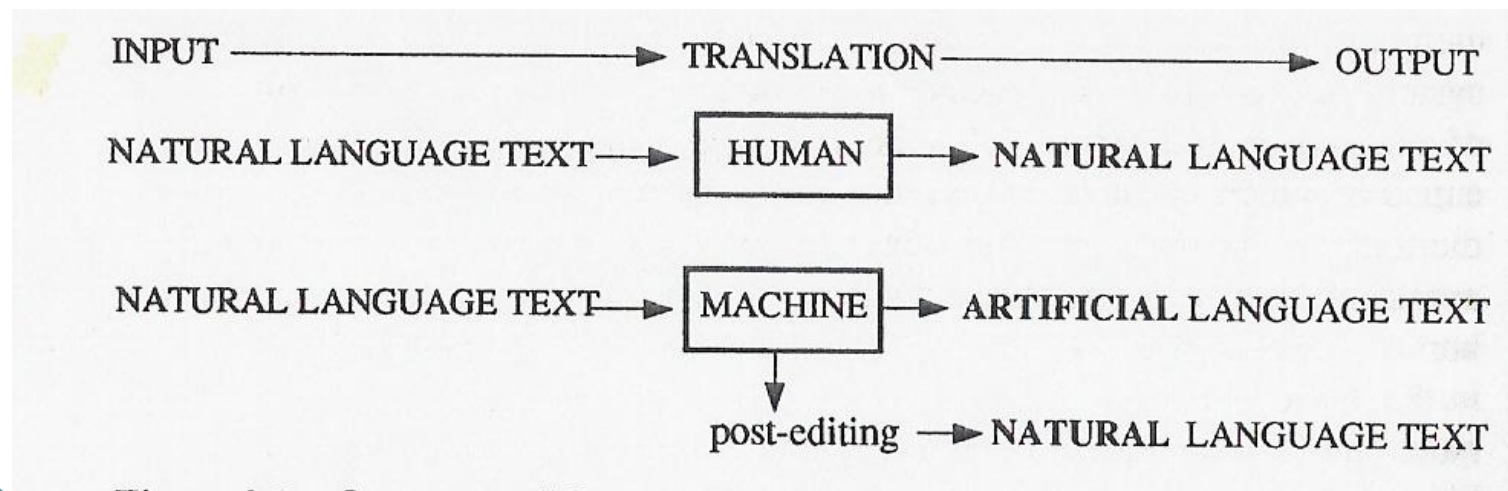
The nature of MT language

- MT language is constructed and **artificial** (*the computer can't produce sentences on its own*)
- MT corresponds to the designer's perception of SL and TL
- Has no creative potential (it is not as flexible and multifunctional as HT language)
- They exclude emotive, aesthetic of other meanings

Human and Machine Translation
- translation products -

The nature of MT language

- MT systems are one-way converter (they only recognize words that belong to the system)
- MT language often needs post-editing

INPUT ⟶ TRANSLATION ⟶ OUTPUT
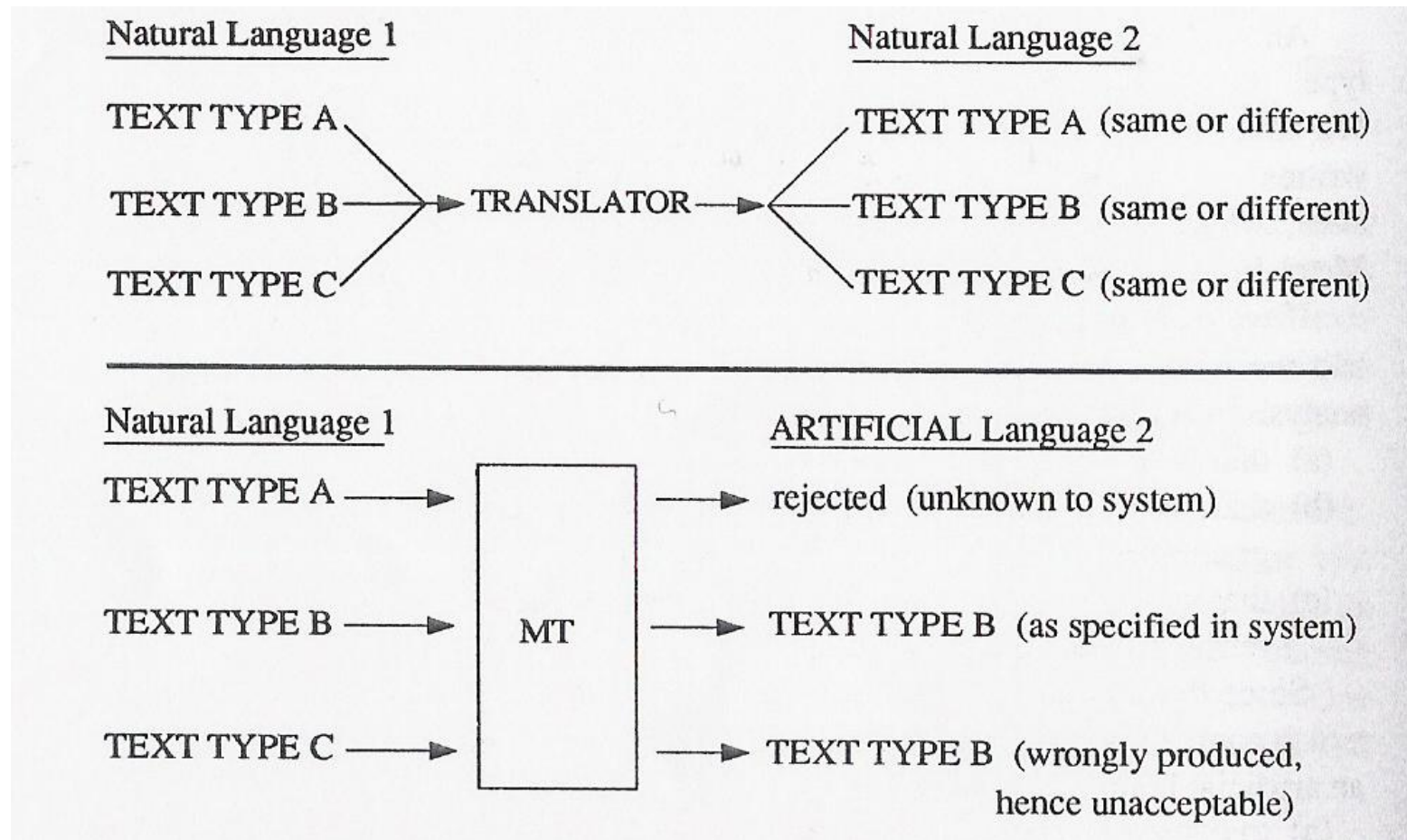
NATURAL LANGUAGE TEXT ⟶ HUMAN ⟶ NATURAL LANGUAGE TEXT

NATURAL LANGUAGE TEXT ⟶ MACHINE ⟶ ARTIFICIAL LANGUAGE TEXT

post-editing ⟶ NATURAL LANGUAGE TEXT

Human and Machine Translation
- translation products -

Flexibility vs. rigidity in text types

- MT language is created on the sentence level

$\rightarrow$ no distinctions on the text type possible

$\rightarrow$ MT systems can only handle text types they have been programmed for

$\rightarrow$ unknown text types cause unacceptable output

Human and Machine Translation
- translation products -

Natural Language 1

TEXT TYPE A

TEXT TYPE B ──► TRANSLATOR ──►

TEXT TYPE C

Natural Language 2

TEXT TYPE A (same or different)

TEXT TYPE B (same or different)

TEXT TYPE C (same or different)

Natural Language 1

TEXT TYPE A ──►

TEXT TYPE B ──► MT ──►

TEXT TYPE C ──►

ARTIFICIAL Language 2

rejected (unknown to system)

TEXT TYPE B (as specified in system)

TEXT TYPE B (wrongly produced, hence unacceptable)

Challenge for MT language

- Construction of user-friendly artificial language (AL)

- Optimum transfer of information from SL/NL to AL

- Convince users that AL is equally efficient as NL (Natural Language)

Some MT systems

- ## ATLAS
  - Japanese system, based on structural transfer, for specialised technical texts
- ## CULT
  - Interactive system, for on-line translation of texts in the field of mathematics from Chinese into English

Some MT systems

- ## METEO
  - The Canadian Federal Government system for the production of bilingual French-English weather reports
- ## SYSTRAN
  - Oldest commercially available MT system, of un-edited output, for post-editing use, for restricted-language document input and for general use in the French Minitel system
  - Largest number of language pairs, all EC languages

Case study: Google translate as MT

- **Seminar paper**: "*An analysis of Google Translate accuracy in translating idioms*"

Students: Laura Bajrami

Pajtesa Mustafa

Faruk Istogu

What is Google Translate?

- Google Translate is a technological resource, through the internet, that offers the chance to benefit from translation materials.

- A document or words are entered into the dialogue box where it is then translated into a different language.

- For example, a language can be changed from French to English and then back to French again with just the click of a few buttons.

This is how it looks like

Advantages of using Google Translate

- Googe translate has the access it provides users. Individuals who have an internet connection and can access Google Translate via their computer are ready to begin working their translations.

- There is also the fact that individuals who use this application find that they can have access to more than fifty languages from which to make their translation

Disadvantages

- An individual that has access to the translator may not put in the time or effort, to fully absorb the language, especially when it comes to high school students.

- Further, the result is a lack of true understanding if the translation is correct and whether or not the words are in the correct order

- GT does not apply grammatical rules, since its algorithms are based on statically analysis rather than traditional rule-based analysis.

An analysis of Google Translate accuracy in translating idioms

- **The text in English**
- My friend suggested that we **join forces**. **"There's safety in numbers,"** he said. Let's **hit the road** together." I **was in two minds** whether to go with him but finally decided to say yes. We travelled together for six months and **had a whale of time**. We spent money **like there was no tomorrow**, so I had **to twist my dad's arm** and persuade him to send me some more money so I could travel further.

Google Translation (the text in Albanian Language)

- Miku im sugjeroi që ne **bashkojmë forcat**. **"Nuk ka siguri në numër,"** ka thënë ai. **Le të goditur në rrugë** së bashku! Unë **kam qenë në dy mendje** nëse do të shkojë me të, **por në fund vendosi për të thënë po**. Kemi udhëtuar së bashku për gjashtë muaj dhe **kishte një balenë të kohës**. Ne shpenzuar paratë **si nuk ka nesër**, kështu që unë kam për të prerë krahun baba im dhe ta bindë atë që të më dërgoni disa para më shumë kështu që unë mund të udhëtonin më tej.

CONCLUSION

- We have seen what is Google Translate and how does it work - Its advantages and disadvantages.

- Google Translate is an electronic device of the modern age that is used as a tool by translators and by everyone who learns foreign languages only by pressing few buttons.

GENERAL CONCLUSION

- The future of MT remains uncertain but with the growth of international trade and the continuing increase in use of MT technologies on the Web, things are looking up.

- It is expected that more MT products will come to market than ever before and a larger number of languages can be tackled.

THANK YOU