GOOGLE TRANSLATE

Google Translate (Groves & Mundt, 2015)

- Web-based machine translation (MT) (Groves & Mundt, 2015)
- It is a statistics-based translation tool:
  - the system calculates probabilities of various translations of a phrase being correct, rather than aiming for word-for-word translation.
- Interactivity:
  - users are able to correct the original translation, and this information being absorbed into the database.
- https://youtu.be/_GdSC1Z1Kzs

The three generations of MT Architecture

1. **The first generation** (1960s to 1980s) was based on **direct translation**
2. **The second generation** (1980s to present) consists of **rule-based** systems such as the transfer and interlingua systems
3. **The third generation** (1990s to present) includes corpus-based systems that are either statistical based or example based.

While **direct translation** systems employed a "word-for-word translation… with no clear built-in linguistic component", the **rule-based** and corpus-based systems are far more complex (Banitz, 2020).

Rule-based Machine Translation

- **Why (and to what extent) translation is hard for a computer**
  - As all translators know, **translation is not simply a matter of finding the target words that correspond to the words in the source text, and then getting the target grammar right.** But even if this was all there was to it, it would still be a difficult task for a computer program. Let us start by **suggesting that 'translation' involves understanding the meaning of the source text and rendering it in an appropriate form in the target language.** Although 'understanding' and 'meaning' are vague terms, we can agree that at the least it involves selecting the correct sense of each individual word, and recognizing the relationship between the words, as expressed by the syntax of the source text.

Rule-based Machine Translation

- ➔ It involves the application of morphological, syntactic and/or semantic rules to the analysis of a source-language text and synthesis of a target language text requiring linguistic knowledge of both the source and the target languages as well as the differences between them" (Banitz, 2020, p. 56)

- Interlingua Systems: work with an abstract intermediate representation of the source text out of which the target text is generated "without 'looking back' to the original text" - *limited version due to the need of building a whole model for the translation at hand, so it is used in some contexts.*

- The more common approach to rule-based MT are transfer systems - it *analyzes a source text sentence by sentence identifying the part of speech of each word and its possible meanings*

Statistical Machine Translation (SMT)

➔ SMT is a predominant method in Machine Translation (MT) that relies on large bilingual corpora to find the most probable target-language sentence.

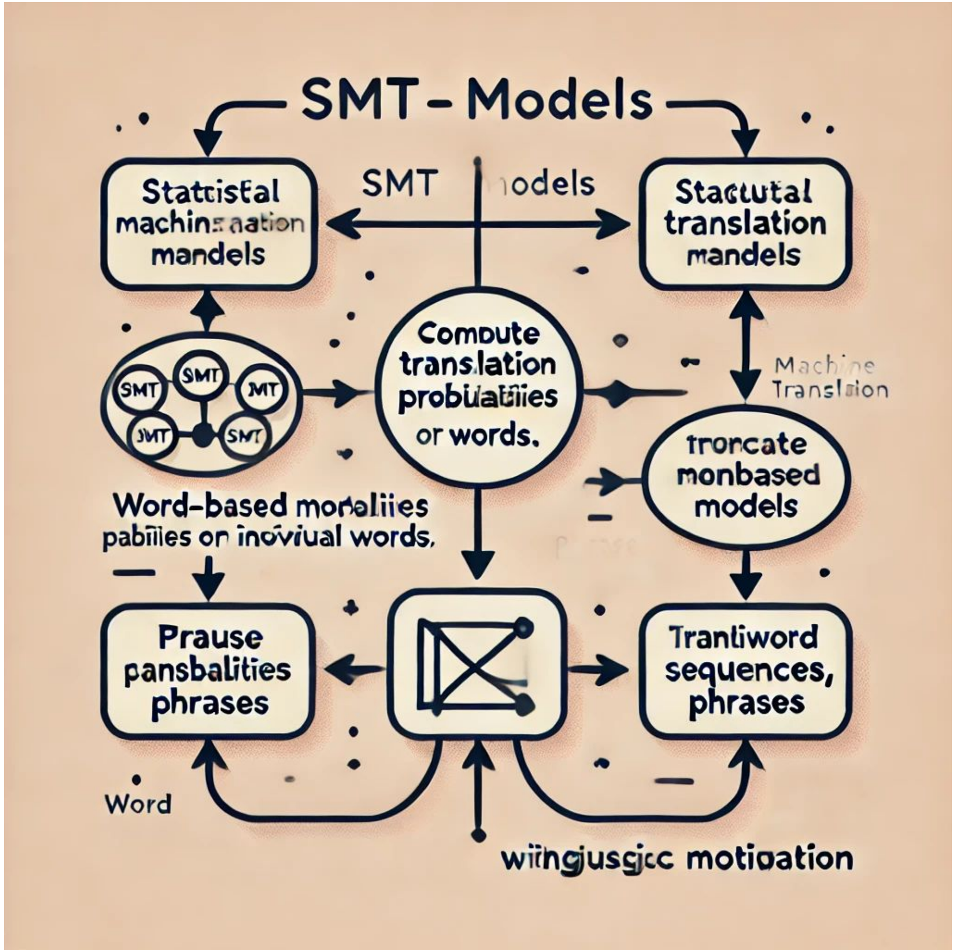
> Working with massive bilingual corpora, the system looks for the target sentence with the highest probability match (Banitz, 2015, p. 57)

➔ SMT does not retrieve exact past translations but instead calculates probabilities for translations based on aligned texts.

Types of SMT Models

Word-Based Models: Compute translation probabilities for individual words.

Phrase-Based Models: Translate multiword sequences (phrases) without linguistic motivation.



ChatGPT. (2025, February 5). *Types of Statistical Machine Translation (SMT) Models* [AI-generated image]. OpenAI.

Translation Process in SMT

1. **Segmentation**: Source text is divided into phrases.

2. **Probability Calculation**: The system compares phrases with a bilingual corpus to determine the most likely translation.

3. **Translation Model**: Matches source-language words with target-language words based on probability.

4. **Target-Language Model**: Ensures fluency and grammatical correctness in the target language.

5. **Final Output**: Translated text is generated by assembling the best-matching translated segments.

NEURAL MACHINE TRANSLATION

Neural Machine Translation

In 2016, the world of translation and multilingualism changed with the development and spread of a **novel approach to machine translation (MT)** implementing the findings of **artificial intelligence, big data, and neuroscience.** The combination of these new scientific endeavors led to the so-called **Neural Machine Translation (NMT)** (Wu et al., 2016 as cited in Klimova et al, 2023).

ChatGPT. (2025, February 5). *Futuristic depiction of Neural Machine Translation (NMT) evolution* [AI-generated image]. OpenAI.

How does Neural Machine Translation work?

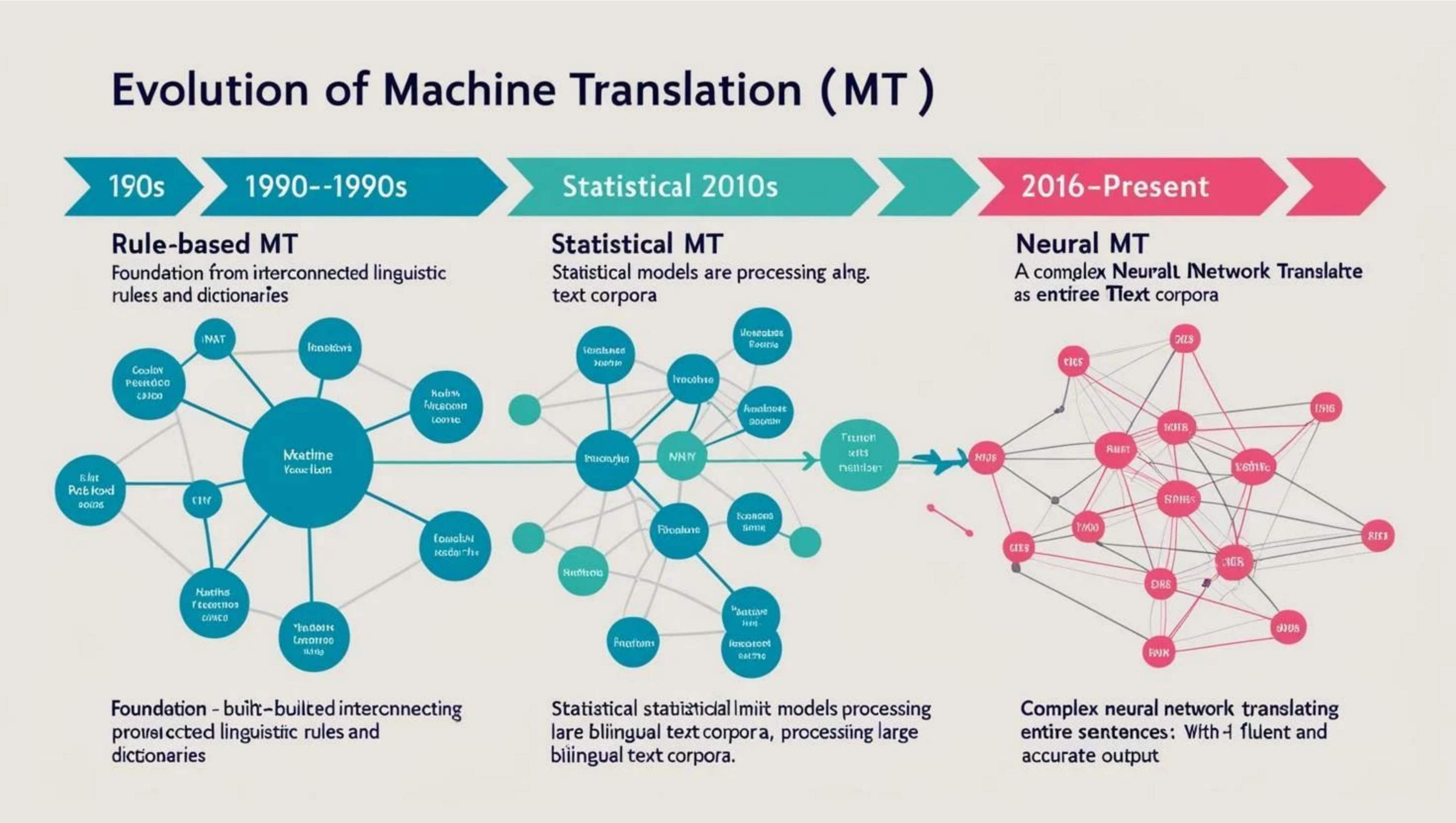
NMT is a form of machine translation that uses artificial neural networks to predict the likelihood of a sequence of words, typically modeling entire sentences in a single integrated model.

Key Features:

- ➔ Based on deep learning and artificial neural networks.
- ➔ Processes large amounts of data to learn patterns and improve translation quality.

Examples:

- ➔ Google Translate, DeepL.



Evolution of Machine Translation (MT)
190s
1990--1990s
Statistical 2010s
2016-Present
Rule-based MT
Foundation from interconnected linguistic rules and dictionaries
Statistical MT
Statistical models are processing alng. text corpora
Neural MT
A complex Neurall Network Translate as entiree Tlext corpora
Foundation - built-builted interconnecting provelccted linguistic rules and dictionaries
Statistical statistical lmiit models processing lare bliingual text corpora, processing large biiingual text corpora.
Complex neural network translating entire sentences: With -l fluent and accurate output

How NMT works:

Neural Networks:

➔ Composed of interconnected artificial neurons that process data.

➔ Words or sub-words are treated as units, similar to neurons in the brain.

Translation Process:

1. Input sentence is processed by the neural network.
2. The network predicts the most probable sequence of words in the target language.
3. Output is generated based on the highest probability.

Example:

Input: "The cat is on the mat."

Output: "Le chat est sur le tapis."

Challenges and Limitations of NMT

- **Algorithmic Bias:**
  - Gender and regional biases in translations.
- **Contextual Limitations:**
  - NMT struggles with translating entire texts as a single unit.
- **Error Detection:**
  - Advanced learners can detect errors, but beginners may struggle.

GENDER BIAS

Since women are often underrepresented in training data, neutral or female forms are repeatedly replaced by male forms in machine-translated texts. In particular, jobs and titles are translated according to current stereotypes (e.g., nurses will be female, while surgeons are translated using male pronouns).

E X A M P L E S

- **Example: In many languages, professions are often automatically assigned a gender based on societal stereotypes. For instance:**
  - **English to Spanish**: The sentence "The nurse is caring for the patient" might be translated as *"La enfermera está cuidando al paciente"* (using the feminine form for "nurse"), while "The surgeon is operating" might be translated as *"El cirujano está operando"* (using the masculine form for "surgeon").
  - **English to French**: "The teacher is in the classroom" might be translated as *"Le professeur est dans la salle de classe"* (masculine form), even if the teacher is female.
- **Example: When the gender of a person is unspecified, machine translation systems often default to male pronouns:**
  - **English to German**: "Someone left their bag" might be translated as *"Jemand hat seine Tasche vergessen"* (using the masculine pronoun *"seine"* instead of a gender-neutral option).
  - **English to Russian**: "A doctor came to see me" might be translated as *"Ко мне пришел врач"*(using the masculine form of "came" and "doctor").

REGIONAL BIAS

Regional language variations are not always taken into consideration: in pluricentric languages, the overrepresentation of one variety in the training data leads NMT systems to produce only outputs in this variety (e.g., the dominating German variety of the German language in contrast to the varieties of German used in Austria or Switzerland).

Input (English): "I am going to buy a roll."

German (Germany) Translation: *"Ich werde ein Brötchen kaufen."*

Austrian German: The correct term would be *"Semmel"* instead of *"Brötchen."*

Swiss German: The correct term would be *"Weggli"* instead of *"Brötchen."*

CONTEXTUAL LIMITATIONS

To date, common systems are not capable of translating whole texts as a single unit, but only isolated sentences. This results in several problems at once because the **text type and the related textual features are not considered in the translation process**, which will not result in a proper translation.

For example, no distinction is made between technical and lay language when it comes to the selection of terminology:

Technical Language:

English: "The system requires a firmware update to resolve the compatibility issue."

Brazilian Portuguese: *"O sistema requer uma atualização de firmware para resolver o problema de compatibilidade."*

Lay Language:

English: "You need to update the system's software to fix the problem."

Brazilian Portuguese: *"Você precisa atualizar o software do sistema para corrigir o problema."*

Issue: A machine translation system might use *"firmware"* (a technical term) instead of *"software"*(a more general term), confusing non-technical users.

CONTEXTUAL LIMITATIONS

Another example are forms of address and cultural conventions that are handled arbitrarily since the machine has no knowledge of intercultural differences.

For example, languages can vary considerably with respect to the formality and politeness they require in emails and letters.

Bom dia, Bruno.

As orientações da Proen relativas ao calendário de reposição da greve deixam claro que o aluno não é penalizado por eventuais faltas em aulas extras para reposição de carga horária (seja no sábado ou em qualquer outro dia/horário que não seja o habitual da UC). Deste modo, faltas nos sábados letivos serão deduzidos do total de faltas do aluno no fechamento do semestre.

Atenciosamente,
A Coordenação

Good morning, Bruno.

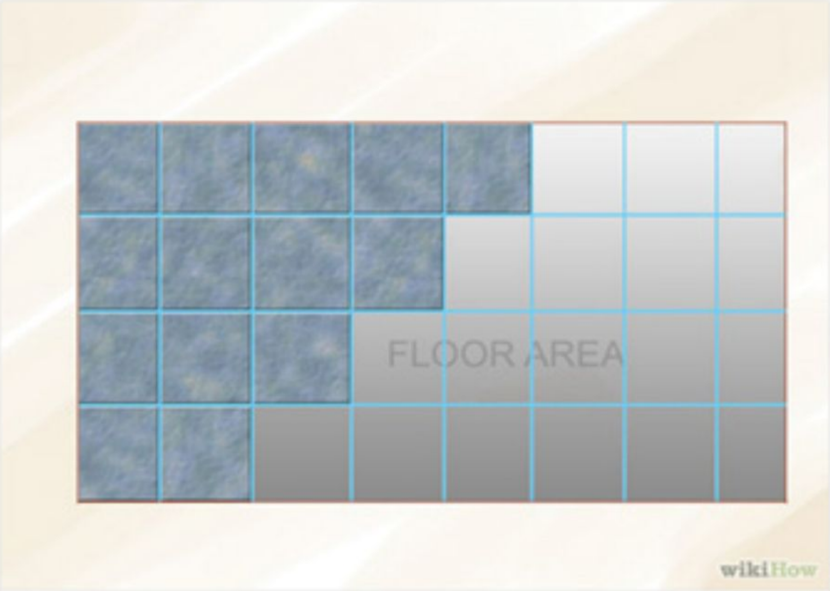
Proen's guidelines regarding the strike make-up schedule make it clear that students will not be penalized for any absences from extra classes to make up for class hours (whether on Saturday or any other day/time other than the usual UC time). Therefore, absences on Saturdays will be deducted from the student's total absences at the end of the semester.

Sincerely,
The Coordination

Multimodal Translation Problems

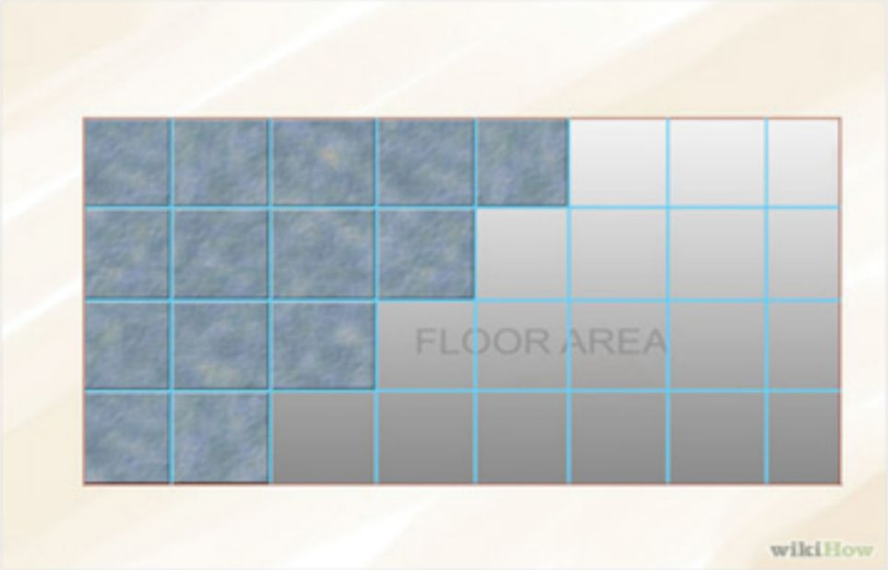
- Intersemiotic Texture: How text and images interact
- Intersemiotic Mismatches: Text and images do not align
- Example:
  - English: “Lay the tile on the floor.”
  - MT Output (Portuguese): “Coloque a telha no chão.”
  - Issue: “Tile” (floor tile) translated as “Telha” (roof tile)

Part 2 of 3: Laying the Tiles



1 **Place the tiles in the pattern.** Place your tiles within the grid you have created. This dry-run allows you to identify areas where you will need to cut tiles to fit and will help you determine the best place to start laying the tile based on your pattern and the shape of the area you plan to tile.

Parte 2 de 3: colocar as telhas



1 **Coloque as telhas no padrão.** Coloque suas telhas dentro da grade que você criou. Este prazo seco permite identificar as áreas onde você precisa para cortar telhas para caber e irá ajudá-lo a determinar o melhor lugar para começar, que a telha com base no seu padrão ea forma da área que você pretende telha.

Extracted from Pires & Espindola (2021)

5 **Stick the dipstick back into the transmission fluid and lift it out again for your reading.** You should now be able to see what level your transmission fluid reaches. Remember to read the "hot" level on the transmission dipstick.

5 **Furar a vareta de volta para o fluido de transmissão e tirá-la de novo para a sua leitura.** Agora você deve ser capaz de ver o que o seu nível de transmissão atinge fluidos. Lembre-se de ler o nível de "quente" na vareta do óleo de transmissão.

4 **Pull the dipstick out and wipe it on a rag.** This will help give you an accurate reading.

4 **Puxe a vareta para fora e limpe-o em um pano.** Isso ajudará a dar-lhe uma leitura precisa.

REFERENCES

Banitz, B. (2020). Machine translation: a critical look at the performance of rule-based and statistical machine translation. *Cadernos de Tradução*, *40*(1), 54–71. https://doi.org/10.5007/2175-7968.2020v40n1p54

Ducar, C, Schocket, DH. Machine translation and the L2 classroom: Pedagogical solutions for making peace with Google translate. *Foreign Language Annals*. 2018; 51: 779– 795. https://doi-org.ez130.periodicos.capes.gov.br/10.1111/flan.12366

Groves, M., & Mundt, K. (2015). Friend or foe? Google Translate in language for academic purposes. *English for Specific Purposes*, *37*, 112–121. https://doi.org/10.1016/j.esp.2014.09.001

Klimova, B., Pikhart, M., Benites, A.D. *et al.* Neural machine translation in foreign language teaching and learning: a systematic review. *Educ Inf Technol* 28, 663–682 (2023). https://doi.org/10.1007/s10639-022-11194-2

Somers, Harold L. “Machine translation: History, development, and limitations.” In: Malmkjaer, Kirsten; Windle, Kevin (Ed). The Oxford handbook of translation studies. Oxford: Oxford University Press, 2011. p. 427-440.

Pires, T. B., & Espindola, A. V. dos S. (2021). Aproximando resultados de tradução automática e imagens em documentos multimodais. *Cadernos De Tradução*, *41*(2), 85–108. https://doi.org/10.5007/2175-7968.2021.e75483