

How Machine Translation Works

Antonio Toral

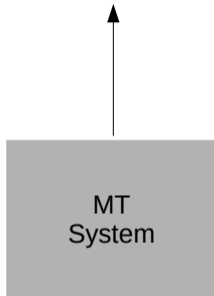
Panel 3: Literary translation and AI (CEATL Strasbourg Conference), 4 October 2024



rijksuniversiteit
groningen

Intro: Opening the Black Box

Comment fonctionne la traduction automatique



MT
System

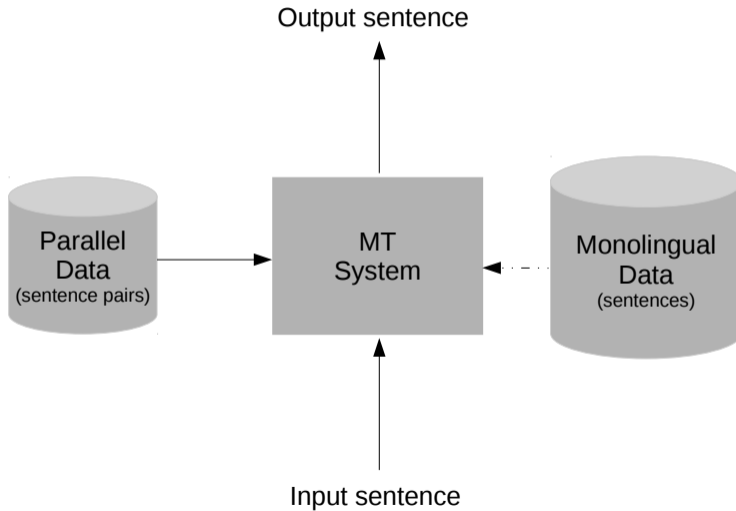
How Machine Translation Works

Table of contents

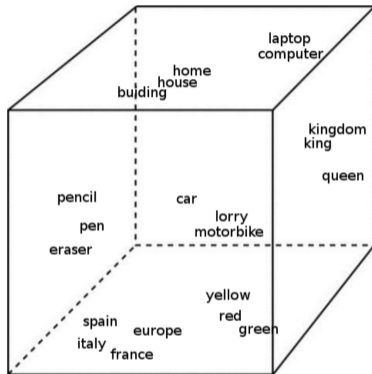
1. NMT 101: Quality
2. LLMs 101: Flexibility
3. Post-editing
4. Technology for Inspiration

NMT 101: Quality

NMT training: predict the translation



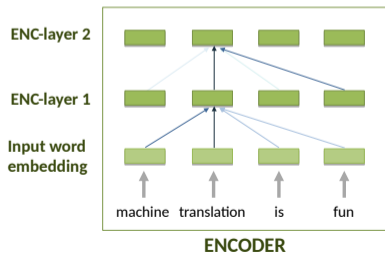
NMT key 1: words as concepts



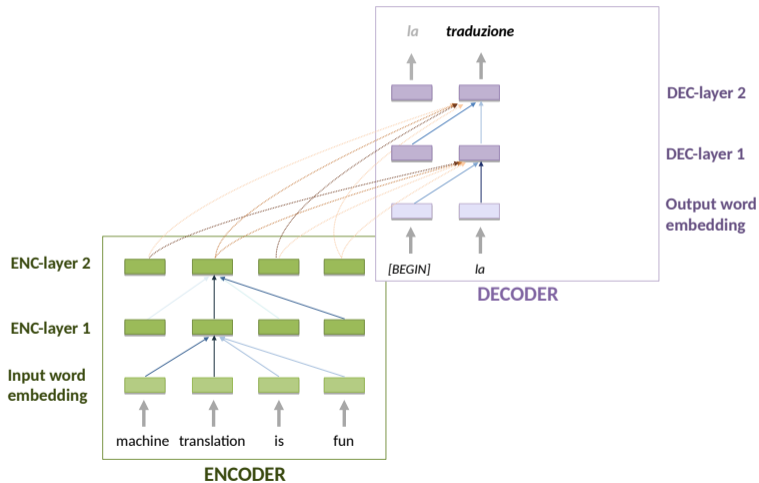
spain = {2, 1, 6}

Forcada (2017)

NMT key 2: all words interconnected



NMT key 2: all words interconnected



NMT vs SMT

NMT vs SMT

EN-DE | EN-FR

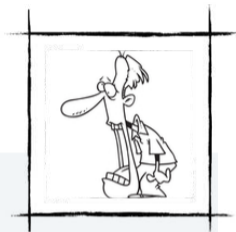
Type text here:

The Budapest Prosecutor's Office has initiated an investigation on the accident.

Translation:

Die Budapester Staatsanwaltschaft hat ihre Ermittlungen zum Vorfall eingeleitet.

NMT vs SMT

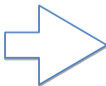


Type text here:

The Budapest Prosecutor's Office has initiated an investigation on the accident.

Translation:

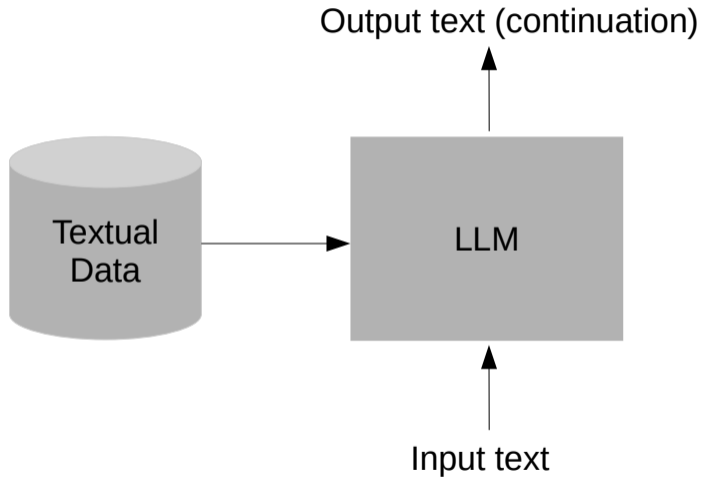
Die Budapester Staatsanwaltschaft hat ihre Ermittlungen zum Vorfall eingeleitet.



LLMs 101: Flexibility

LLM Training 1: predict the next word

LLM Training 1: predict the next word



LLM Training 1: predict the next word

Input: teach me how to bake bread

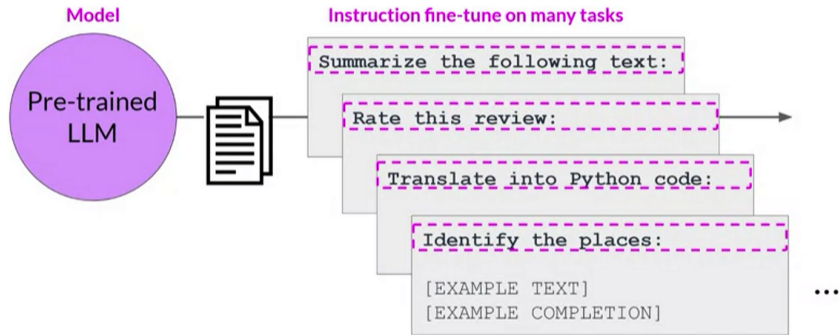
A likely output:

LLM Training 1: predict the next word

Input: teach me how to bake bread

A likely output: in a home oven

LLMs Training 2: instruction tuning



<https://ogre51.medium.com/instruction-fine-tuning-of-llms-a-comprehensive-guide-e2f197e19c36>

LLMs vs NMT for translation

- Much more flexible
- Less literal translations (Raunak et al., 2023)
- Tower (Alves et al., 2024), a LLM then specialised on translation tasks, seems the current state-of-the-art (top rank at WMT 2024)

Post-editing

Post-editing (PE) compared to human translation

- Less creative translations (Guerberof and Toral, 2022)
- Lower lexical variety and higher source language interference (Toral, 2019)
- Translator's voice partially lost (Kenny and Winters, 2020)
- Less enjoyable task, demoted (Moorkens et al., 2018)

Post-editing (PE) compared to human translation

- Less creative translations (Guerberof and Toral, 2022)
- Lower lexical variety and higher source language interference (Toral, 2019)
- Translator's voice partially lost (Kenny and Winters, 2020)
- Less enjoyable task, demoted (Moorkens et al., 2018)

But, in terms of errors (Koponen, 2016), post-editing

- comparable human translation. E.g., Garcia (2010)
- or even better. E.g. Plitt and Masselot (2010)

Technology for Inspiration

Inspiration. Example 1 (Youdale, 2019)

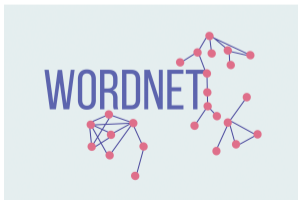
- Use of technology to support **distant reading**
- Corpus tools to complement close reading
 - Measure style: sentence length, repetitions...
 - Better informed translations



Inspiration. Example 2 (Kolb and Miller, 2022)

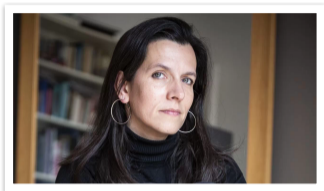
PunCAT: a tool to assist with the translation of puns

- Aims: facilitate brainstorming and provide inspiration
- Automatically translate each sense of the pun
- Allow user to explore the semantic fields of these translations



Inspiration. What about MT? (Guerberof and Toral, 2022)

Inspiration. What about MT? (Guerberof and Toral, 2022)



*The [MT] proposal gave me an adjective, **I wouldn't have thought of.***

*The problem when you translate is that you have lots of words in your head **but you don't know how to reach them.***

Carlota Gurt
literary translator and writer

Inspiration. What about MT? (Farrell, 2022)

Survey: 86% of translators that use MT, use it for inspiration

Example

“I translate passages or sentences myself and then use the MT on the source text to see what it comes up with, and I may adjust my translation on that basis or indeed completely ignore the MT text. The MT never takes the lead but can sometimes be useful as a supplement.”

Merci de votre attention!

Antonio Toral
a.toral.ruiz@rug.nl

<https://antoniotor.al/ceatl24.pdf>

References

- D. M. Alves, J. Pombal, N. M. Guerreiro, P. H. Martins, J. Alves, A. Farajian, B. Peters, R. Rei, P. Fernandes, S. Agrawal, et al. Tower: An open multilingual large language model for translation-related tasks. *arXiv preprint arXiv:2402.17733*, 2024.
- M. Farrell. Do translators use machine translation and if so, how? results of a survey held among professional translator. In *Proceedings of the 44th Conference Translating and the Computer*. Tradulex, 2022.

- M. L. Forcada. Making sense of neural machine translation. *Translation spaces*, 6(2):291–309, 2017.
- I. Garcia. Is machine translation ready yet? *Target. International Journal of Translation Studies*, 22(1):7–21, 2010.
- A. Guerberof and A. Toral. Creativity in translation: Machine translation as a constraint for literary texts. *Translation Spaces*, 11(2):184–212, 2022.
- D. Kenny and M. Winters. Machine translation, ethics and the literary translator's voice. *Translation Spaces*, 9(1):123–149, 2020.

- W. Kolb and T. Miller. Human–computer interaction in pun translation. In J. Hadley, K. Taivalkoski-Shilov, C. S. C. Teixeira, and A. Toral, editors, *Using Technologies for Creative-Text Translation*. Routledge, 2022. To appear.
- M. Koponen. Is machine translation post-editing worth the effort? A survey of research into post-editing and effort. *Journal of Specialised Translation*, 25 (25):131–148, 2016. ISSN 0169-2607. URL <https://sites.google.com/site/wptp2015/>.
- J. Moorkens, A. Toral, S. Castilho, and A. Way. Translators’ perceptions of literary post-editing using statistical and neural machine translation. *Translation Spaces*, 7(2):240–262, 2018.

References iv

- M. Plitt and F. Masselot. A productivity test of statistical machine translation post-editing in a typical localisation context. *The Prague bulletin of mathematical linguistics*, 93:7–16, 2010.
- V. Raunak, A. Menezes, M. Post, and H. Hassan. Do gpts produce less literal translations? In *The 61st Annual Meeting Of The Association For Computational Linguistics*, 2023.
- A. Toral. Post-editeese: an exacerbated translationese. In *Proceedings of the XVII Machine Translation Summit*, Dublin, Ireland, August 2019.
- R. Youdale. *Using computers in the translation of literary style: challenges and opportunities*. Advances in Translation and Interpreting. Routledge, United Kingdom, June 2019. ISBN 9780367141233.