

Translation 2.0: Equipping linguists for the machine translation future

Alina Karakanta

Leiden University Centre for Linguistics

a.karakanta@hum.leidenuniv.nl

Vasilis Kalogiannis

Leiden University Centre for Linguistics

v.p.kalogiannis@umail.leidenuniv.nl

Abstract

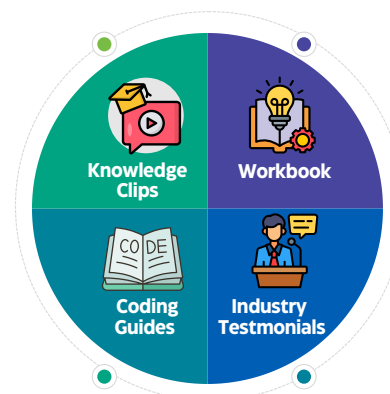
Translation 2.0 addresses a critical gap in accessible, up-to-date educational resources on recent developments in Machine Translation and Large Language Models for students of linguistics and translation. It develops an online module with open-access learning materials, including knowledge clips, a workbook with incremental exercises to consolidate conceptual understanding, practical coding guides, and videos featuring industry professionals. The module aims to build both subject knowledge and computational literacy, freeing up contact hours for deeper engagement and critical discussions on practical, professional and ethical aspects. Translation 2.0 is funded through an Educational Innovation grant by the Faculty of Humanities at Leiden University and ECOLe (Expert Centre for Education and Learning) and runs from February to December 2026.

1 Introduction

Translators and linguists increasingly require training in language technologies and machine translation (MT) to meet market demands (ELIS, 2025), yet accessible educational resources remain scarce. Academic literature on MT tends to presuppose a computational background that many linguistics and translation students lack, creating a significant gap between available materials and the needs of this audience. Notable exceptions include the textbooks by Koehn (2020) and Kenny (2022), as well

© 2026 The authors. This article is licensed under a Creative Commons 4.0 licence, no derivative works, attribution, CC-BY-ND.

Translation 2.0



From neural MT to LLMs
Corpora and other resources
Human and automatic evaluation
Large Language Models for MT
MT customisation
Multi-agent workflows

Figure 1: Structure of the Translation 2.0 module

as collaborative projects such as MultiNMT¹ and LT-LiDER², which made NMT more approachable for translators.

The advent of Large Language Models (LLMs) has further widened this gap. LLMs have fundamentally transformed how MT systems are built and deployed, rendering much of the existing accessible literature obsolete. This leaves linguistics and translation programmes without adequate teaching materials on the current state of the field, and places teachers in a difficult position, as they are expected to deliver up-to-date instruction on a rapidly evolving technology while keeping pace

¹<https://www.multitrainmt.eu/index.php/en/>

²<https://lt-lider.eu/>

with those advancements. Compounding the problem, previous course materials have tended to prioritise declarative content over active learning: they offer limited incremental exercises to consolidate conceptual understanding, little opportunity for students to reason critically about MT implementation, and few practical guides that would equip them to work with open-source tools.

2 Project objectives

This project aims to develop a set of open-access learning materials for courses in translation technology and machine translation, addressed at students of both linguistics and translation programmes. The materials are designed to be flexible across curricula and thus intended for BA and MA students alike. Their goal is not only to convey subject knowledge, but also to build towards digital literacy and internalisation of essential computational principles, e.g. understanding where data is stored, how to estimate computational requirements, and how to structure computational steps and processes. Rather than relying solely on cloud services such as Google Colab or Kaggle, students will also learn to work on local infrastructure and on high performance computing (HPC) clusters³, in order to develop foundational skills necessary for careers as computational linguists or translation technologists.

The structure of the module is shown in Figure 1. Each unit will include 2-4 knowledge clips (5-10 min duration each) covering core topics in MT and LLMs. All clips will be designed based on Web Content Accessibility Guidelines (WCAG)⁴. A workbook with incremental exercises serves improving conceptual understanding and developing hands-on procedural knowledge. Alongside these, step-by-step practical guides will walk students through the use of open-source tools to train and use MT systems in the terminal. An example of a practical coding guide is shown in Figure 2. Lastly, videos featuring industry professionals will discuss real-world applications of MT and language technology, bridging the gap between academic training and professional practice.

All materials will be housed in a module on an

³For Leiden University, students have access to the HPC ALICE computer cluster for running computationally heavy experiments. We acknowledge this may not be the case with many other programs.

⁴<https://www.w3.org/WAI/standards-guidelines/wcag/>

online e-learning platform (Articulate Rise 360⁵). The underlying structure is flexible: as new engines, metrics, and LLM-based features emerge, individual components can be updated without disrupting the overall structure. Students can work on the platform independently before class. This flipped classroom approach aims to free up contact hours for substantive discussion, assignment review, and individual feedback. The guides and clips will be piloted iteratively with students during development to ensure clarity and usability. The evaluation of the final module includes feedback by students and translator trainers. After completion, the module will be made publicly available as open educational resources.

Translating with Tower+ 2B

Contents

1	Introduction	1
2	Setting Up	2
2.1	HuggingFace Account	2
2.2	The Model: Tower+ 2B	2
2.3	Interacting with the Model	2
3	Environment Setup	2
3.1	Create a Virtual Environment	2
3.2	Activate the Virtual Environment	2
3.3	Install Required Packages	3
4	Downloading the Model	3
5	Running a Translation	3
5.1	Request a GPU	3
5.2	Run the Example Script	3
5.3	Understanding the Output	3
6	Translating a File	4
6.1	Prepare the Data	4
6.2	Script Arguments	4
6.3	Run the Translation	4

1 Introduction

In this practical we move from training our own model to using a **pre-trained large language model (LLM)** for translation. Rather than spending days training a system from scratch, we can download a model that has already been trained on large amounts of multilingual data and immediately use it to translate text.

Figure 2: Example taken from the practical guide to inference with a pre-trained LLM on the University HPC Cluster.

References

- ELIS, Research. 2025. European language industry survey 2025: Trends, expectations and concerns of the european language industry. https://elis-survey.org/wp-content/uploads/2025/03/ELIS-2025_Report.pdf.
- Kenny, Dorothy, editor. 2022. *Machine translation for everyone*. Number 18 in Translation and Multilingual Natural Language Processing. Language Science Press, Berlin.
- Koehn, Philipp. 2020. *Neural machine translation*. Cambridge University Press.

⁵<https://360.articulate.com/rise>