Task description: WMT 2022 Shared Task: Word-Level AutoCompletion
Data: https://github.com/lemaoliu/WLAC
This project is a shared task which has well defined data and evaluation metric. It is more about the approach striking for the best performance.

Possible tasks:
- Self-supervised learning on multi-lingual textual data, e.g., translation language model
- Neural aligner, e.g. Mask-Align
Project 1 - References

References
- GWLAN: General Word-Level Autocompletion for Computer-Aided Translation
- TRANSMART: A PRACTICAL INTERACTIVE MACHINE TRANSLATION SYSTEM
- Cross-lingual Language Model Pretraining
- Mask-Align
Project 2 - Sentence-Level AutoCompletion

Task description, see Figure 1: TRANSMART paper

Data:

- Same dataset as word-level autocompletion with appropriate adjustment
- or dataset used in the references
Project 2 - What to expect?

This project is more about the design of the experiment.

Possible tasks:

- Design of experiments, e.g., data and evaluation metrics
- Constrained search (hard vs soft)
- Auto regressive vs Non-Auto regressive translation with constraints
References

- TRANSMART: A PRACTICAL INTERACTIVE MACHINE TRANSLATION SYSTEM
- GWLAN: General Word-Level Autocompletion for Computer-Aided Translation
- Fast Lexically Constrained Decoding with Dynamic Beam Allocation for Neural Machine Translation
- EDITOR: an Edit-Based Transformer with Repositioning for Neural Machine Translation with Soft Lexical Constraints