Boosting machine translation with AI-powered terminology features

Marek Sabo, Judith Klein, Giorgio Bernardinello
STAR AG
Wiesholz 35
CH-8262 Ramsen
{ marek.sabo, judith.klein, giorgio.bernardinello }
@star-group.net

Abstract

Artificial intelligence (AI) is quickly becoming an exciting new technology for the translation industry in form of large language models (LLMs). AI-based functionality could be used to improve the output of neural machine translation (NMT). One main issue that impacts MT quality and reliability is incorrect terminology. This is why STAR is making AI-powered terminology control a priority for its translation products because of the significant gains to be made — greatly improving the quality of MT output, reducing postediting (PE) costs and efforts, and thereby boosting overall translation productivity.

1 Improving terminology accuracy in MT output

The lack of correct terminology in MT output often requires extensive manual postediting to improve accuracy and consistency, which can be addressed at three stages of the translation process:

MT system selection — before translation: A customised MT model could be selected that is trained for more appropriate terminology usage compared to generic systems.

Terminology injection — during translation: More and more MT systems like DeepL or Textshuttle offer terminology support through user-provided bilingual glossaries. While compliance with the term specifications is mostly good, there are still grammatical errors when inserting terms or they do not fit with interdependent words.

Correcting terminology — after translation: The last stage for correcting terminology is on the MT output. This might be mostly necessary if (1) either no customised MT was used or did not perform correctly, (2) the system does not support terminology injection or not for the selected language, (3) the provided glossary term was not used, or it lacked morphological adaptation, or resulted in grammatical errors in other parts of the translation.

2 AI-powered terminology extraction

Numerous in-house translation projects at STAR have confirmed that terminology injection in MT output significantly reduces postediting efforts. The integration of this functionality in CAT tools, such as DeepL within STAR Transit, has become a standard practice in translation workflows, thus expanding its usage at a larger scale yielding growing benefits. With the growing demand for suitable MT glossaries, a fast and reliable terminology extraction method is crucial to boost the production of bilingual glossaries in a translation business driven by MT. This function could significantly enhance the productivity of MTPE.

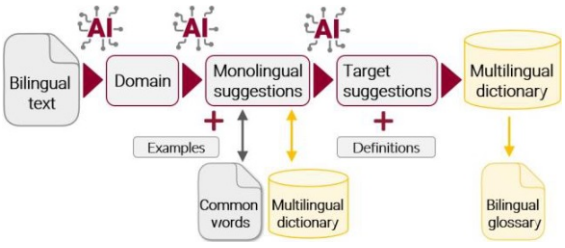

Figure 1: AI-powered bilingual terminology extraction

Powered by AI, STAR Transit in combination with TermStar (integrated terminology management tool) will deliver automatic extraction of

Proceedings of the 25th Annual Conference of the European Association for Machine Translation (Volume 2), pages 25–26
June 24-27, 2024 ©2024 European Association for Machine Translation

bilingual terminology from a translated project or a selected part of the translation memory.

With one click on the AI-powered term extraction feature a carefully-designed prompt is issued to an LLM that identifies the domain as well as candidate source language terms for that domain. These terms are checked against entries in the TermStar dictionary and a user-specific common words list. For the unmatched suggestions the corresponding target terms are extracted from the target language segment. In addition, the source segment is extracted for context, and another prompt is used to create a definition. Before being imported into the dictionary, the proposed entries are validated by human experts.

Some MT systems, e.g. Textshuttle.com, use the TermStar dictionary directly in STAR Transit for term injection while for others, like DeepL.com, the user exports the bilingual list and uploads it in STAR Transit to the MT system.

The seamless integration of the AI features in the terminology extraction function makes them both accessible and easy to use. User feedback and results will lead to improvements for various specific term extractions with additional pre-defined and customized prompts. Initial tests are being carried out with GPT-4-Turbo and GPT-3.5-Turbo for German, English, French, Italian and Spanish but further tests are also planned with smaller, local models.

3 AI-powered terminology correction in MT output

At STAR, development is underway of an AI-enhanced term correction feature, injecting terms to the MT output. As a proof of concept, we have selected English and Swedish as source languages and Slovak, a highly inflected lower-resourced European language as the target language.

We have trained a local model sized under 2 billion parameters for term injection into Slovak. The required MT glossary could be extracted from a domain-specific TermStar dictionary.

The AI term correction function scans and compares source and target text against the glossary, automatically inserting matched terms and ensuring adherence to declension patterns, gender conventions, and inflectional morphemes for each word to maintain syntactic role consistency. The feature will adjust related words like adjectives, verbs, or pronouns associated with the term, ensuring not only accuracy but also fluency in the translation, as shown in this example:

Source	Which screw terminal did you see mentioned in that manual?
DeepL	Ktorý skrutkový terminál ste videli uvedený v tejto príručke?
Terms	manual = manuál screw terminal = hlavičková svorka
Edited	Ktorýú ~~skrutkový~~ ~~terminál~~ hlavičkovú svorku ste videli uvedenýú v ~~tej~~tomto ~~príručke~~ manuáli?
Final	Ktorú hlavičkovú svorku ste videli uvedenú v tomto manuáli?

Table 1: Injected terms and their related words have been adjusted for case and grammar.

A major benefit of this approach is the flexibility of the solution, since the term correction can be applied to any MT output. Our current focus is on (1) finding the ideal model size to attain the best quality and speed, (2) exploring innovative approaches to dataset preparation, and (3) looking into domain matching and improvements in term alignment.

The AI-powered term correction can work with large LLMs, but the smaller models, customized for this specific task, are cost-effective, suitable for running on average laptops without GPU demands, guarantee data privacy over cloud-based LLMs, and function at stable speeds. In addition, it is easier to control their output while commercial LLMs can change at any point, affecting the dependent system.

4 Next steps

The release of the AI-powered terminology features within STAR's language technology products are scheduled for later this year. We will also investigate additional AI features for quality estimation and evaluation, with a focus on terminological accuracy, as automatic terminology correction may have failed or lead to other translation errors. In addition, we will explore the possibility of using AI to generate an image based on the term, its definition and the context-related example that supports the user's understanding of the term and its postediting.