

# Prediction of Translation Techniques for the Translation Process

Fan Zhou, Vincent Vandeghinste

KU Leuven, KU Leuven

Belgium, Belgium

fan.zhou@student.kuleuven.be, vincent.vandeghinste@kuleuven.be

## Abstract

Machine translation (MT) encompasses various methodologies aimed at enhancing translation accuracy. In contrast, the human-generated translation process relies on diverse translation techniques, which proves essential to ensuring both linguistic adequacy and fluency. This study proposes that these translation techniques can serve as a guide for optimizing machine translation further. However, it's imperative to automatically identify these techniques before they can effectively guide the translation process. The study distinguishes between two translation process scenarios: from-scratch translation and post-editing. For each scenario, a tailored set of experiments has been devised to predict the most suitable translation techniques. The results show that for from-scratch translation, the predictive accuracy reaches 82%, while the post-editing process shows even greater promise, achieving an accuracy rate of 93%.

**Keywords:** translation techniques, translation quality, machine translation

## 1. Introduction

Neural machine translation (NMT) systems, such as Google Translate<sup>1</sup>, DeepL<sup>2</sup>, and recent large language models like ChatGPT<sup>3</sup> (Hendy et al., 2023), have made significant progress but still fall short of achieving human parity (Even-Zohar, 1979; Toury, 1991; Baker, 1993; Baker, 1996; Baker, 2004; Dayrell, 2007; Hassan et al., 2018; Wang et al., 2021). Issues persist, such as word-for-word translation, false friends, ambiguity, information omission or addition, cultural insensitivity, etc. (Puşnei, 2023; Zhang et al., 2023; Bao et al., 2023; Lee et al., 2023; Hendy et al., 2023), resulting in low-quality translations that may not be easily understood.

In the analysis of low-quality translations, most issues stem from inappropriate translation techniques. For example, in Table 1, both source sentences were translated using a literal or word-for-word technique with ChatGPT, resulting in either unnatural Chinese expressions or poor comprehension. In contrast, the first human translation employs the "reduction" translation technique to convey the meaning more concisely, while the second human translation splits the sentence into two clauses, using "modulation" for a more natural and concise translation.

Utilizing translation techniques is crucial for addressing translation problems, improving translation quality, and ensuring contextually appropriate translations. However, there are few existing studies explicitly suggesting such inference processing in current MT systems as to humans' decisions on

translation techniques throughout the translation procedure (Stahlberg, 2020; Dabre et al., 2020; Wan et al., 2022; Ranathunga et al., 2023; Klimova et al., 2023).

These translation techniques can serve as a guide for human translation and, furthermore, for the generation of MT. They can also act as prompts for large language models to produce high-quality translations. In our investigation of both from-scratch translation and post-editing scenarios, our objective is to ascertain whether it is possible to predict the most appropriate translation techniques when provided with the source sentences and bad translations together with its source sentences. Our experiments have shown that pre-trained models, after being fine-tuned for both scenarios, can proficiently predict the most suitable translation techniques. This paves the way for future advancements in the field of machine translation generation.

## 2. Related Work

In the from-scratch translation process, manipulating the NMT architecture can help improve translation accuracy. For instance, structuring hidden states in the encoder as syntactic tree representations has shown performance improvement (Chen et al., 2017). Another approach involves a semantic interface that connects pre-trained encoders and decoders, enabling separate pre-training with a shared language-independent space (Ren et al., 2021). In the post-editing process, Chatterjee et al. (2019) developed automatic post-editing (APE), which automatically corrects errors in the output of an MT system by learning from human corrections. APE can be viewed as a cross-lingual

<sup>1</sup><https://translate.google.com/>

<sup>2</sup><https://www.deepl.com/en/translator>

<sup>3</sup><https://chat.openai.com/>

<b>English Source:</b> He's one of the famous writers.
<b>Chinese (ChatGPT):</b> 他是著名的作家之一。(He's one of the famous writers.)
<b>Chinese (Human):</b> 他是位有名的作家。(He's a famous writer.)
<b>English Source:</b> The dawn was unfolding out of the darkness.
<b>Chinese (ChatGPT):</b> 黎明从黑暗中展现出来。(The dawn unfolds out of the darkness.)
<b>Chinese (Human):</b> 晨曦出现，黑暗渐去。(The dawn begins to appear, and the darkness was fading.)

Table 1: Comparison of English-Chinese Translation Examples: Machine Translation vs. Human Translation

sequence-to-sequence task, which takes a source sentence and the corresponding MT output as inputs and generates the post-edited (PE) output (Lee, 2020). Lee used source texts along with MT translation as input to be processed in the Transformer’s encoder to generate the post-edited output, and then used its decoder to fine-tune PE. However, Carl and Schaeffer (2017) explored the effects of cross-lingual syntactic and semantic distances on translation production times and found that non-literal translations pose challenges in both from-scratch translation and post-editing.

Translation techniques assist in selecting appropriate translations, considering factors like naturalness and discourse coherence (Deng and Xue, 2017). Non-literal translations reflect the diversity of human languages and are essential for preserving accuracy and fluency (Zhai et al., 2020). Zhai et al. (2019) automated the classification of translation techniques for English-French pairs, but the number of aligned pairs for the experiment is limited. For the translation techniques used in NMT, most studies focus on idiom translation problems caused by literal translation. The fact is that most of the idioms are translated with non-literal translation techniques, but according to the evaluation of the language model (Dankers et al., 2022), Transformer’s tendency to process idioms as compositional expressions contributes to poor literal translations of idioms. In order to make the idiom translated correctly, Fadaee et al. (2018) retrieved sentences containing idioms and mixed them with non-idiom phrases as a training data set to train the NMT. Apart from idioms, some other literal translation problems are studied. For example, Gamallo and Labaka (2021) suggested that the large amount of literal translation causes transferring the constructions of the source language to the target language, and they proposed to use monolingual corpora instead of parallel ones with unsupervised translation, and this makes the hybrid SMT plus NMT system produce more non-literal translation output on passive voice-structured sentences. Bacquelaine (2023) found word-by-word translation decreases in DeepL and Google Translate. The case study of Raunak et al. (2023) shows that there is a greater tendency towards non-literalness in GPT transla-

tions and GPT systems’ ability to figuratively translate idioms.

In this paper, we use English-Chinese pairs labeled with translation techniques to investigate whether pre-trained cross-lingual language models can be fine-tuned to accurately predict translation techniques so as to provide guidance for producing good translations in both from-scratch translation and PE/APE processes.

### 3. Data

Our study uses English-Chinese aligned pairs at the word and phrase level. Each aligned pair is labeled with a specific translation technique.

#### 3.1. UM Corpus

The English-Chinese parallel sentences employed in our experiments are sourced from the UM-corpus (Tian et al., 2014). This corpus boasts manually verified sentence-level alignments. The parallel sentences we have selected for our study have been screened to ensure non-duplication and feature well-aligned sentence pairs. These specific pairs were chosen due to the presence of at least one individual aligned unit at the sub-sentence level (either a word or a phrase). This selection criterion allows for precise labeling with a specific translation technique, eliminating potential ambiguities where a single aligned unit might be associated with multiple labels.

#### 3.2. Data Preparation

The data preparation process involves three phases: aligning parallel pairs, extracting features for each translation technique, and labeling the pairs.

**Alignment** In the alignment process, we work with parallel sentences, with a specific focus on sub-sentence units. English and Chinese sentences are tokenized using SpaCy<sup>4</sup> and Jieba<sup>5</sup>, respectively. To ensure precise alignment of English-Chinese tokens, we use multiple toolkits, such as

<sup>4</sup><https://spacy.io/>

<sup>5</sup><https://github.com/fxsjy/jieba>

GIZA++ (Och and Ney, 2003), and cross-lingual models like mBERT (Devlin et al., 2018), mBART (Liu et al., 2020), and mT5 (Xue et al., 2021). These models calculate cosine similarity and select the highest score for each word in the source language, significantly improving alignment accuracy. Furthermore, we employ custom algorithms to retrieve data that specifically pertains to translation technique features. To ensure data’s high quality, we incorporate human annotations and leverage ChatGPT’s alignment by prompts.

**Features of Translation Techniques** Our feature selection process adheres to the guidelines outlined in the “Annotation Guidelines of Translation Techniques for English-Chinese”<sup>6</sup>. These guidelines present a comprehensive typology of eleven translation techniques: literal translation, equivalence, transposition, modulation, modulation+transposition, particularization, generalization, figurative translation, lexical shift, explication, and reduction. Notably, we exclude “figurative translation” from this list for three reasons:

- scarcity in the data;
- elusive features for detection;
- the fact that using the other ten techniques can also yield good translations.

From a linguistic perspective, we categorize these features into three types:

- syntax-featured;
- semantics-featured;
- both syntax-featured and semantics-featured.

To collect data based on the features of each translation technique, we employ a range of methods. For translation techniques with syntactically related features, such as transposition and lexical shift, we employ SpaCy for part-of-speech tagging. For techniques associated with semantic features, such as equivalence, particularization, and generalization, we use Synonym<sup>7</sup>’s word similarity, NLTK WordNet<sup>8</sup>’s hyponyms and hypernyms, along with prompt-based ChatGPT for initial processing. We also incorporate human checks to ensure data quality. Translation techniques that involve both syntactic and semantic features, like modulation and modulation+transposition, are handled by combining the aforementioned methods to collect the data.

**Labeling** In the labeling process based on features, we assign each aligned pair a single, un-equivocal translation technique.

<sup>6</sup>[https://yumingzhai.github.io/files/Annotation\\_guide\\_EN\\_ZH.pdf](https://yumingzhai.github.io/files/Annotation_guide_EN_ZH.pdf)

<sup>7</sup><https://github.com/chatopera/Synonyms>

<sup>8</sup><https://www.nltk.org/howto/wordnet.html>

Task	Softmax	Train	Dev	Test
Architecture 1	10	69828	7759	8621
Architecture 2	10	69828	7759	8621
Architecture 3 <sub>subtask1</sub>	3	73958	8218	9131
subtask2	9	54303	6034	6890
Architecture 4	10	73958	8218	9131

Table 2: Data Overview

### 3.3. Data Summary

We have amassed a dataset of more than 100,000 data pairs. Each data pair consists of a source sentence, a target sentence, an aligned word or phrase (a source unit and a target unit), and a corresponding translation technique label. For a comprehensive view of how the data is employed for specific tasks, please refer to the details provided in Table 2. We make both the collected data and the codes for data retrieval and experiments available<sup>9</sup>.

## 4. Experiments

There are four experiments involving two architectures of from-scratch translation (4.1.1 and 4.1.2) and two architectures of PE/APE (4.2.1 and 4.2.2). Since the experiments do not involve translation generation, our focus is solely on the encoder part of NMT architectures.

The four experiments, each employing different architectures, are depicted in Figure 1. The data overview for each architecture can be seen in Table 2. Different architectures utilize distinct data formats as input, encompassing two input formats, denoted as Input1 and Input2 in Table 3.

### 4.1. From-scratch Translation

The sole data available for from-scratch translation consists of texts in the source language. In preparation for the from-scratch translation generation, we employ language models to explore the possibility of predicting appropriate translation techniques. This predictive analysis is carried out as part of the pre-translation process.

We employ two distinct architectures for this prediction task. One architecture is trained exclusively with source language data, while the other incorporates both source and target language data, with the target data serving as supplementary information to train the model. The objective of these two experiments is to forecast both literal and non-literal translation techniques for the source language.

<sup>9</sup><https://anonymous.4open.science/r/translation-technique-436D/>

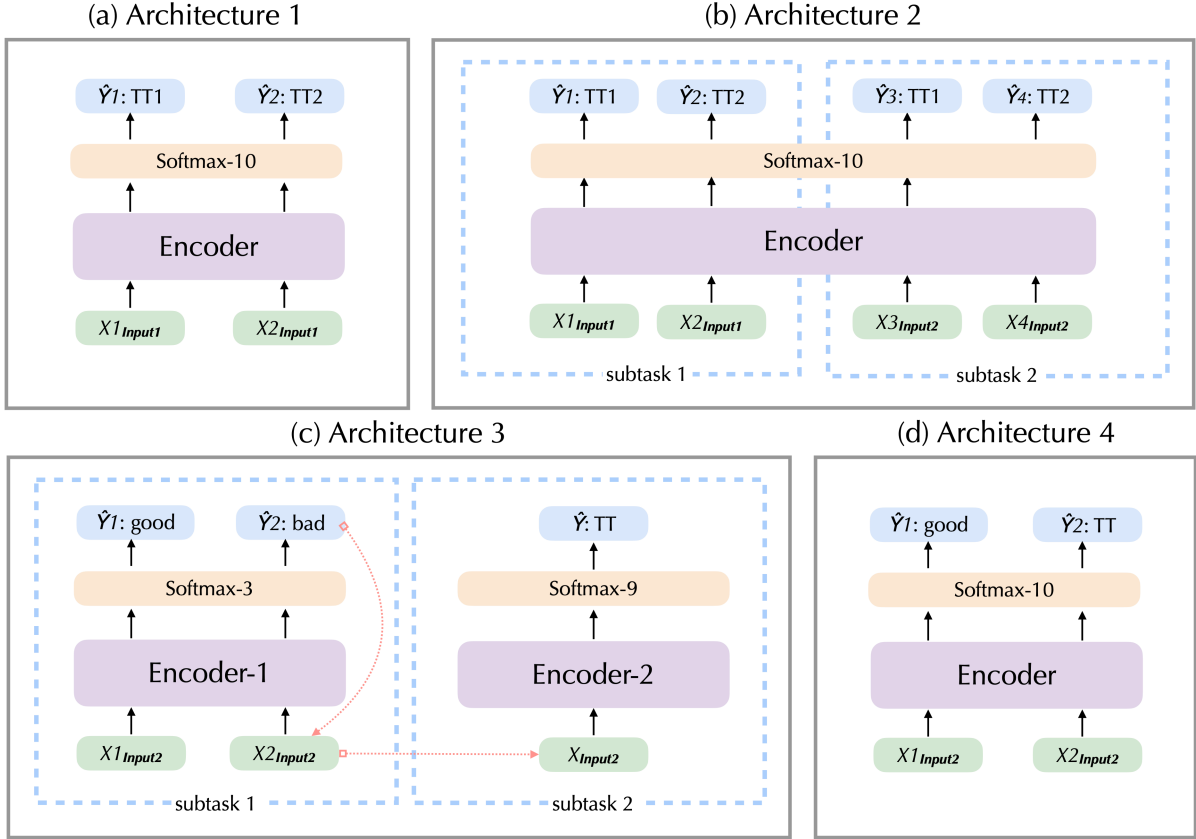


Figure 1: Experiments feature four distinct architectures: (a) and (b) are designated for from-scratch translation, while (c) and (d) are tailored for post-editing tasks. The input data is structured in two formats: Input1 and Input2, with comprehensive specifics available in Table 3. 'n' of the Softmax-n layer corresponds to the number of categories. 'TT' signifies a specific translation technique.

	Format	Example
<b>Input1</b>	source unit[SEP]source sentence	started[SEP]My wound has been aching ever since it started to rain.
<b>Input2</b>	source unit[SEP]target unit [SEP]source sentence[SEP]target sentence	started[SEP] 开始 [SEP]My wound has been aching ever since it started to rain.[SEP] 自从开始下雨以来, 我的伤口一直在疼。

Table 3: Data Input Formats

#### 4.1.1. Architecture 1

We adopt the architecture showcased in Figure 1 (a). In contrast to the approach employed by Zhai et al. (2020), where they localized aligned pairs and extracted the last hidden state based on token positions, our methodology takes a different path. Our model's input combines the source part of the aligned pair with the source sentence by inserting the special token [SEP] (end-of-sentence token) between them, referred to as Input1 (as detailed in Table 3). This format effectively unifies both parts into a single input, allowing us to consider how the sole source part within the source sentence interacts with the entire context, resembling the human translation process.

For fine-tuning, we employ encoder models, which encompass BERT-base, BART-large's encoder, and T5-large's encoder. These models vary in terms of structure and the number of trainable parameters (110M, 410M, and 770M parameters, respectively). The fine-tuning process stops when accuracy reaches its peak on the validation dataset. We utilize the Adam optimizer with an initial learning rate of 1e-5, and the learning rate is dynamically adjusted. Additionally, dropout or L2 regularization is dynamically incorporated in different training epochs. Typically, fine-tuning converges at the sixth epoch.

Model	Architecture 1	Model	Architecture 2	Architecture 3	Architecture 4	
BERT	0.802	mBERT	0.808	0.912 <sub>subtask1</sub>	<b>0.929</b> <sub>subtask2</sub>	0.901
BART	0.774	mBART	0.813	0.911 <sub>subtask1</sub>	0.922 <sub>subtask2</sub>	0.907
T5	<b>0.811</b>	mT5	<b>0.819</b>	<b>0.915</b> <sub>subtask1</sub>	0.920 <sub>subtask2</sub>	<b>0.914</b>

Table 4: Overall performance by different models in four architectures, measured by accuracy.

Architecture 1	LIT	TRA	EXP	RED	LEX	EQU	GEN	MOD	PAR	MOT
BERT-Base	0.725	0.672	1.0	<b>0.923</b>	0.784	<b>0.900</b>	0.728	<b>0.773</b>	0.775	0.579
BART-Large	0.681	0.657	1.0	0.901	0.736	0.891	0.714	0.748	0.746	0.540
T5-Large	<b>0.741</b>	<b>0.699</b>	1.0	0.920	<b>0.804</b> †	<b>0.900</b>	<b>0.751</b>	0.764	<b>0.782</b>	<b>0.623</b> †
Architecture 2	LIT	TRA	EXP	RED	LEX	EQU	GEN	MOD	PAR	MOT
mBERT	0.740	0.710	1.0	<b>0.936</b> †	0.794	<b>0.903</b> †	0.728	0.761	0.773	0.585
mBART-Large	0.741	0.708	1.0	0.929	0.793	0.901	<b>0.760</b> †	0.769	<b>0.806</b> †	<b>0.605</b>
mT5-Large	<b>0.762</b> †	<b>0.727</b> †	1.0	0.932	<b>0.799</b>	0.890	0.753	<b>0.781</b> †	0.793	<b>0.605</b>
Architecture 3	BAD	TRA	EXP	RED	LEX	EQU	GEN	MOD	PAR	MOT
mBERT	0.905	<b>0.962</b> *	1.0	0.984	0.979	<b>0.905</b> *	<b>0.850</b> *	0.922	<b>0.880</b> *	<b>0.752</b> *
mBART-Large	0.907	0.957	1.0	0.982	<b>0.981</b> *	0.888	0.840	0.921	0.868	0.720
mT5-Large	<b>0.916</b>	0.938	1.0	<b>0.986</b> *	0.979	<b>0.905</b> *	0.847	<b>0.925</b> *	0.854	0.700
Architecture 4	GOOD	TRA	EXP	RED	LEX	EQU	GEN	MOD	PAR	MOT
mBERT	0.938	0.904	1.0	<b>0.980</b>	0.938	0.832	0.692	0.726	0.705	0.525
mBART-Large	0.943	<b>0.919</b>	1.0	0.961	0.902	0.792	0.706	0.686	<b>0.736</b>	0.570
mT5-Large	<b>0.954</b> *	0.893	1.0	0.974	<b>0.955</b>	<b>0.839</b>	<b>0.712</b>	<b>0.742</b>	0.710	<b>0.604</b>

Table 5: The models’ predictions encompass various translation techniques, including LIT (literal translation), TRA (transposition), EXP (unaligned explicitation), RED (unaligned reduction), EQU (equivalence), GEN (generalization), MOD (modulation), PAR (particularization), MOT (modulation+transposition), BAD (bad translation), and GOOD (good translation). These predictions are assessed using the F1-score as the performance metric.

#### 4.1.2. Architecture 2

To improve the accuracy of classifying and predicting translation techniques, we introduce a multi-task architecture, which is illustrated in Figure 1 (b). This architecture comprises two subtasks. Subtask 1 is focused on predicting translation techniques based on data from the source language, and its input follows the Input1 format, as outlined in Table 3. On the other hand, Subtask 2 is designed to identify patterns associated with specific translation techniques, utilizing Input2 (Table 3) as its input. Subtask 2 encodes data from both the source and target languages, providing the model with insights into how translation techniques are utilized. Both subtasks share the same encoder, employing identical architecture and parameters.

Given that the primary goal of the experiments related to the from-scratch translation process is to use source language data exclusively for predicting the most suitable translation technique for each unit within a source language sentence, subtask 1 holds the principal role, with subtask 2 serving as a supportive auxiliary task. Consequently, during the back-propagation process, we calculate a

weighted loss summation for the optimization strategy, as outlined in Formula (1). The weight configuration, with  $\alpha$  set at 0.8 for subtask 1 loss and  $\beta$  at 0.2 for subtask 2 loss, yields optimal results. This weight distribution emphasizes that the second task is the primary one and should carry more weight during the encoder training.

$$L = \alpha L1 + \beta L2 \quad (1)$$

In this setup, we evaluate cross-lingual encoder models, including mBERT, mBART-large, and mT5-large, which have 110M, 410M and 770M parameters, aligning with parameter numbers of models used in Architecture 1. The training process stops when the best validation accuracy is achieved.

#### 4.2. PE/APE

Our investigation aims to determine whether models can accurately predict suitable translation techniques for poor translations in the PE/APE process. In practice, post-editors frequently need to differentiate between good and poor translations



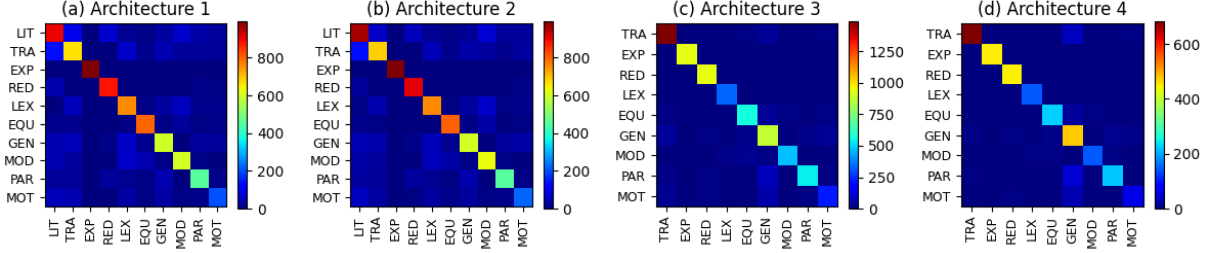


Figure 2: Translation techniques' prediction heatmap. The statistical number of each translation technique is averaged from 3 models in each architecture. Non-diagonal X-axis represents false positive and non-diagonal Y-axis represents false negative.

prior to the PE/APE process. As a result, our experiments unfold in two steps:

1. Identifying poor translations from the provided dataset;
2. Predicting appropriate translation techniques for these poor translations.

We conduct two distinct experiments for the prediction task that precedes the PE/APE process. The first architecture dissects these two steps, while the second integrates them into a single process, where the system distinguishes poor translations and simultaneously predicts the required translation techniques.

Bad translations can arise from various factors, with one of the most common causes being the application of the literal translation technique. This translation technique often lead to translations that are rigid or entirely incorrect. Therefore, in these two experiments, we introduce "bad literal translations" as the exclusive category of bad translations, which are generated from the existing good translations by replacing good non-literal translations with bad literal translation.

#### 4.2.1. Architecture 3

In this architecture, we employ two encoders, and the structure, depicted in Figure 1 (c), consists of two consecutive subtasks. Both encoders utilize the Input2 format, as specified in Table 3.

In subtask 1, we employ Encoder-1 to segregate poor translations from the good translations which encompass both good literal and good non-literal translations. After this filtering process, good translations remain, while the poor ones are directed to subtask 2.

Within subtask 2, Encoder-2 is utilized to assign suitable translation techniques to the identified poor translations. Encoder-2 is trained using data that includes all poor translations, specifically, rigid, poor literal translations. These bad literal

translations are fed into Encoder-2, and its predictions include one of nine non-literal translation techniques. This setup enables us to rectify bad literal translations by applying non-literal translation techniques during the PE/APE phase.

To facilitate PE, post-editors initially need to identify bad translations within the source texts and their corresponding translations. We kick-start this process by fine-tuning a pre-trained model for this purpose. The encoders used in the two subtasks of this experiment are of the same type of cross-lingual language models, including mBERT, mBART-large's encoder, and mT5-large's encoder. Encoder-1 converges at the fourth epoch of the training process, while Encoder-2 reaches convergence at the fifth epoch.

#### 4.2.2. Architecture 4

The architecture utilized in this experiment, illustrated in Figure 1 (d), consolidates the two-step process of Architecture 3 into a single step, featuring just one encoder. All data, formatted as Input2 (as shown in Table 3), which includes both good and bad translations, is fed into this encoder. The bad translations receive appropriate translation techniques, while the output for the good translations is labeled as "good." For this experiment, we employed the encoders of mBERT, mBART-large, and mT5. The encoder reaches convergence at the fifth epoch of the training process.

## 5. Results

Table 4 showcases the overall performance of the models, as assessed by their accuracy. Table 5 presents the models' predictions, measured by the F1-score, for different translation techniques. Moreover, Figure 2 provides an illustration of the likelihood of incorrect predictions for the translation techniques.

In both Architectures 1 and 2, the overall accuracy hovers at approximately 81%. However, the latter displays slightly higher average prediction

accuracy. Models fed with the even distribution of input data results in an uneven prediction output. Translation techniques with syntax features achieve prediction F1-score exceeding 80%, surpassing those with semantic features which range between 70-80%. Both architectures face challenges when predicting modulation+transposition, achieving a F1-score of only around 61%.

In Architecture 3, the accuracy for subtask 1, which involves models distinguishing bad literal translations from good translations, surpasses 90%. The overall accuracy reaches as high as 92% in subtask 2. When predicting appropriate translation technique labels for bad translations, most syntactic-related translation techniques (over 94%) and certain semantics-related translation techniques (equivalence and modulation at 90%, generalization and particularization at 85%) achieve high F1-scores in their predictions. However, the modulation+transposition technique only reaches the best F1-score of 75.2% when using mBERT.

In Architecture 4, the models' performance achieves an accuracy above 90%. Specifically, they can identify "good translations" with an impressive 95% F1-score when utilizing mT5. When predicting translation techniques for bad translations, syntax-related techniques achieve over 90% on F1-score, surpassing semantics-related techniques. Notably, generalization (71.2%, best F1-score achieved with mT5), particularization (73.6%, best F1-score achieved with mBART), modulation (74.2%, best F1-score achieved with mT5), and modulation+transposition (60.4%, best F1-score achieved with mT5) exhibit comparatively lower F1-scores.

Both Architectures 3 and 4 achieve high accuracy. However, they differ in structures and dataset used. Whether the difference in prediction performance is due to the architectural design or the dataset composition remains to be explored.

## 6. Conclusions

Using translation techniques is useful in the human translation process to ensure translation accuracy and fluency, yet it remains an under-discussed aspect in current neural machine translation. In this study, we explore the feasibility of predicting translation techniques for different words and phrases within their contexts. This prediction not only aids in human translation but also serves as a means to provide prompts to large language models for the generation of precise translations.

However, it's important to note that our current work primarily focuses on the phases preceding the from-scratch translation and PE/APE process, without directly incorporating translation

techniques into NMT systems for translation generation. Our future research will focus on enhancing the decoder component to achieve optimal translations, guided by translation techniques. Additionally, we acknowledge the challenges associated with obtaining substantial dataset for experiments, considering the scarcity of open-sourced parallel sentence corpora with aligned lexical or phrasal units as well as inaccurate alignment by existing tools. Our forthcoming efforts will also focus on the automation of sub-sentence parallel unit alignment to streamline research and reduce the burden on human resources.

## 7. Bibliographical References

- Jeffrey Allen. 2003. Post-editing. *Benjamins Translation Library*, 35:297–318.
- John R Anderson. 2013. *The architecture of cognition*. Psychology Press.
- Françoise Bacquellaine. 2023. DeepL and Google Translate translating Portuguese multiword units into French: progress, decline and remaining challenges (2019-2023). *Human language technologies as a challenge for computer science and linguistics-2023*.
- Mona Baker. 1993. Corpus linguistics and translation studies implications and applications. *Text and Technology: In honour of John Sinclair*, page 233.
- Mona Baker. 1996. Corpus-based translation studies: The challenges that lie ahead. In *Terminology, LSP and Translation*, page 175. John Benjamins.
- Mona Baker. 2004. A corpus-based view of similarity and difference in translation. *International journal of corpus linguistics*, 9(2):167–193.
- Keqin Bao, Yu Wan, Dayiheng Liu, Baosong Yang, Wenqiang Lei, Xiangnan He, Derek F. Wong, and Jun Xie. 2023. [Towards fine-grained information: Identifying the type and location of translation errors](#).
- Akshay Batheja and Pushpak Bhattacharyya. 2023. [Improving machine translation with phrase pair injection and corpus filtering](#).
- Michael Carl and Moritz Jonas Schaeffer. 2017. Why translation is difficult: A corpus-based study of non-literality in post-editing and from-scratch translation. *HERM ES-Journal of Language and Communication in Business*, (56):43–57.

- Rajen Chatterjee, Christian Federmann, Matteo Negri, and Marco Turchi. 2019. Findings of the wmt 2019 shared task on automatic post-editing. In *Proceedings of the Fourth Conference on Machine Translation (Volume 3: Shared Task Papers, Day 2)*, pages 11–28.
- Huadong Chen, Shujian Huang, David Chiang, and Jiajun Chen. 2017. [Improved neural machine translation with a syntax-aware encoder and decoder](#).
- Hugh A. Chipman, Edward I. George, Robert E. McCulloch, and Thomas S. Shively. 2022. [mBART: Multidimensional Monotone BART](#). *Bayesian Analysis*, 17(2):515 – 544.
- Rahzeb Choudhury and Brian McConnell. 2013. Taus–translation technology landscape report. *Published by TAUS BV, De Rijp*, 12.
- Hélène Chuquet and Michel Paillard. 1987. *Approche linguistique des problèmes de traduction anglais-français*. Editions Ophrys.
- Raj Dabre, Chenhui Chu, and Anoop Kunchukuttan. 2020. A survey of multilingual neural machine translation. *ACM Computing Surveys (CSUR)*, 53(5):1–38.
- Verna Dankers, Christopher G Lucas, and Ivan Titov. 2022. Can transformer be too compositional? analysing idiom processing in neural machine translation. *arXiv preprint arXiv:2205.15301*.
- Carmen Dayrell. 2007. A quantitative approach to compare collocational patterns in translated and non-translated texts. *International Journal of Corpus Linguistics*, 12(3):375–414.
- Dun Deng and Nianwen Xue. 2017. Translation divergences in chinese–english machine translation: An empirical investigation. *Computational Linguistics*, 43(3):521–565.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Itamar Even-Zohar. 1979. Polysystem theory. *Poetics today*, 1(1/2):287–310.
- Marzieh Fadaee, Arianna Bisazza, and Christof Monz. 2018. Examining the tip of the iceberg: A data set for idiom translation. *arXiv preprint arXiv:1802.04681*.
- Pablo Gamallo and Gorka Labaka. 2021. Using dependency-based contextualization for transferring passive constructions from english to spanish. *Procesamiento del Lenguaje Natural*, 66:53–64.
- Jonas Gehring, Michael Auli, David Grangier, Denis Yarats, and Yann N Dauphin. 2017. Convolutional sequence to sequence learning. In *International conference on machine learning*, pages 1243–1252. PMLR.
- Pamela Gerloff. 1986. Second language learners’ reports on the interpretive process: Talk-aloud protocols of translation pamela gerloff, harvard university. *Interlingual and intercultural communication: Discourse and cognition in translation and second language acquisition studies*, 272:243.
- Behrooz Ghorbani, Orhan Firat, Markus Freitag, Ankur Bapna, Maxim Krikun, Xavier Garcia, Ciprian Chelba, and Colin Cherry. 2021. [Scaling laws for neural machine translation](#).
- Hany Hassan, Anthony Aue, Chang Chen, Vishal Chowdhary, Jonathan Clark, Christian Federmann, Xuedong Huang, Marcin Junczys-Dowmunt, William Lewis, Mu Li, et al. 2018. Achieving human parity on automatic chinese to english news translation. *arXiv preprint arXiv:1803.05567*.
- Amr Hendy, Mohamed Abdelrehim, Amr Sharaf, Vikas Raunak, Mohamed Gabr, Hitokazu Matsushita, Young Jin Kim, Mohamed Afify, and Hany Hassan Awadalla. 2023. [How good are gpt models at machine translation? a comprehensive evaluation](#).
- Ke Hu and Patrick Cadwell. 2016. A comparative study of post-editing guidelines. In *Proceedings of the 19th annual conference of the European association for machine translation*, pages 34206–353.
- Wenxiang Jiao, Wenxuan Wang, Jen tse Huang, Xing Wang, and Zhaopeng Tu. 2023. [Is chatgpt a good translator? yes with gpt-4 as the engine](#).
- Blanka Klimova, Marcel Pikhart, Alice Delorme Benites, Caroline Lehr, and Christina Sanchez-Stockhammer. 2023. Neural machine translation in foreign language teaching and learning: a systematic review. *Education and Information Technologies*, 28(1):663–682.
- Hans P Krings. 2001. *Repairing texts: Empirical investigations of machine translation post-editing processes*, volume 5. Kent State University Press.
- Dongjun Lee. 2020. Cross-lingual transformers for neural automatic post-editing. In *Proceedings of the Fifth Conference on Machine Translation*, pages 772–776.



- Nayeon Lee, Chani Jung, and Alice Oh. 2023. Hate speech classifiers are culturally insensitive. In *Proceedings of the First Workshop on Cross-Cultural Considerations in NLP (C3NLP)*, pages 35–46.
- Xiner Liu, Jianshu He, Mingzhe Liu, Zhengtong Yin, Lirong Yin, and Wenfeng Zheng. 2023. A scenario-generic neural machine translation data augmentation method. *Electronics*, 12(10):2320.
- Yinhan Liu, Jiatao Gu, Naman Goyal, Xian Li, Sergey Edunov, Marjan Ghazvininejad, Mike Lewis, and Luke Zettlemoyer. 2020. Multilingual denoising pre-training for neural machine translation. *Transactions of the Association for Computational Linguistics*, 8:726–742.
- Amparo Molina, Lucía ; Hurtado Albir. 2002. [Translation techniques revisited: A dynamic and functionalist approach](#). *Meta*, 47(4):498–512.
- Peter Newmark. 1981. *Approaches to translation (Language Teaching methodology series)*. Oxford: Pergamon Press.
- Franz Josef Och and Hermann Ney. 2003. A systematic comparison of various statistical alignment models. *Computational linguistics*, 29(1):19–51.
- Juan Ignacio Pozo and Yolanda Postigo. 1993. Las estrategias de aprendizaje como contenido del currículo. *Estrategias de aprendizaje: procesos, contenidos e interacción*. Barcelona: Domenech.
- Irina Puşnei. 2023. Source language and culture interference in translation. *ACROSS*, 7(4):5–15.
- Surangika Ranathunga, En-Shiun Annie Lee, Marjana Prifti Skenduli, Ravi Shekhar, Mehreen Alam, and Rishemjit Kaur. 2023. Neural machine translation for low-resource languages: A survey. *ACM Computing Surveys*, 55(11):1–37.
- Vikas Raunak, Arul Menezes, Matt Post, and Hany Hassan Awadalla. 2023. [Do gpts produce less literal translations?](#)
- Felix Stahlberg. 2020. Neural machine translation: A review. *Journal of Artificial Intelligence Research*, 69:343–418.
- Liang Tian, Derek F Wong, Lidia S Chao, Paulo Quaresma, Francisco Oliveira, and Lu Yi. 2014. UM-Corpus: A Large English-Chinese Parallel Corpus for Statistical Machine Translation. In *LREC*, pages 1837–1842.
- Gideon Toury. 1991. What are descriptive studies into translation likely to yield apart from isolated descriptions. In *Translation studies: The state of the art*, pages 179–192. Brill.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.
- J.P. Vinay and J. Darbelnet. 1958. [Stylistique comparée du français et de l'anglais: méthode de traduction](#). Bibliothèque de stylistique comparée. Didier.
- Yu Wan, Baosong Yang, Derek Fai Wong, Lidia Sam Chao, Liang Yao, Haibo Zhang, and Boxing Chen. 2022. Challenges of neural machine translation for short texts. *Computational Linguistics*, 48(2):321–342.
- Haifeng Wang, Hua Wu, Zhongjun He, Liang Huang, and Kenneth Ward Church. 2021. Progress in machine translation. *Engineering*.
- Shijie Wu and Mark Dredze. 2020. [Are all languages created equal in multilingual bert?](#)
- Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V. Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, Jeff Klingner, Apurva Shah, Melvin Johnson, Xiaobing Liu, Łukasz Kaiser, Stephan Gouws, Yoshikiyo Kato, Taku Kudo, Hideto Kazawa, Keith Stevens, George Kurian, Nishant Patil, Wei Wang, Cliff Young, Jason Smith, Jason Riesa, Alex Rudnick, Oriol Vinyals, Greg Corrado, Macduff Hughes, and Jeffrey Dean. 2016. [Google's neural machine translation system: Bridging the gap between human and machine translation](#).
- Linting Xue, Noah Constant, Adam Roberts, Mihir Kale, Rami Al-Rfou, Aditya Siddhant, Aditya Barua, and Colin Raffel. 2021. [mT5: A massively multilingual pre-trained text-to-text transformer](#). In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 483–498, Online. Association for Computational Linguistics.

- Yuming Zhai, Lufei Liu, Xinyi Zhong, Gbariel Illouz, and Anne Vilnat. 2020. Building an english-chinese parallel corpus annotated with sub-sentential translation techniques. In *Proceedings of the 12th language resources and evaluation conference*, pages 4024–4033.
- Yuming Zhai, Aurélien Max, and Anne Vilnat. 2018. Construction of a multilingual corpus annotated with translation relations. In *First Workshop on Linguistic Resources for Natural Language Processing*, pages 102–111.
- Yuming Zhai, Pooyan Safari, Gabriel Illouz, Alexandre Allauzen, and Anne Vilnat. 2019. [Towards Recognizing Phrase Translation Processes: Experiments on English-French](#).
- Jinpeng Zhang, Nini Xiao, Ke Wang, Chuanqi Dong, Xiangyu Duan, Yuqi Zhang, and Min Zhang. 2023. [Disambiguated lexically constrained neural machine translation](#).