Data Augmentation Strategies to Combat Rare Words in Machine Translation

A0244482A, A0235256E, A0217689L, A0217387W, A0220750U
Group 33
Mentored by Luo Yang
{e0865694,e0727256,e0543725,e0543423,e0555942}@u.nus.edu

Abstract

Neural Machine Translation (NMT) models have progressed significantly in recent years, especially since the advent of the transformer architecture. However, there is still room for improvement, especially when it comes to translation of sentences that contain rare or out-of-vocabulary (OOV) words. In this paper, we focus on the former issue of rare words, and identify data augmentation as a potential solution. We evaluate various augmentation strategies that preserve syntactic correctness of the generated data, and our findings on custom datasets show that such techniques achieve significant BLEU score gains on sentences containing rare words, whilst also improving accuracy on the original dataset.

1 Introduction

Data augmentation is widely recognised as a fundamental strategy across numerous branches of machine learning in enhancing both the robustness and efficacy of models. This practice involves artificially increasing the size of training datasets with label-preserving transformations, allowing models to capture more generalised features and perform better on unseen data. This technique is especially prevalent in fields like computer vision, where operations such as translations, reflections, and colour adjustments are routinely applied to images to develop more adaptable and robust image recognition models (Krizhevsky et al., 2012). Similarly, in automatic speech recognition, variations in the speed of audio samples generate new training data (Ko et al., 2015).

However, the adoption of data augmentation in Natural Language Processing (NLP), and more specifically in Machine Translation (MT), has been less immediate. This delay can largely be attributed to the challenges in defining transformation rules that are both universally applicable and maintain the integrity of the data across different linguistic contexts. In the context of MT specifically, these complexities are further compounded as transformations must ensure that any changes to sentences are mirrored accurately in their corresponding translations, preserving both syntactic and semantic correctness. Despite setting new performance benchmarks in the MT task (Vaswani et al., 2017), NMT systems continue to face non-trivial challenges due to the limitations of the training data they are based on. These challenges manifest in two ways: the presence of rare words with minimal occurrences in training data, and the absence of certain words from the entire training corpus. Standard NMT models struggle in generating accurate translations in such contexts (Fernando and Ranathunga, 2022).

In this paper, we concentrate on the specific problem of rare words in the context of the IWLST-2017 English-Chinese dataset (Cettolo et al., 2017). Past efforts to address this challenge have included innovative word replacement methods that incorporate linguistic features such as Part-of-Speech (POS) and morphological analysis (Fadaee et al., 2017).

Our study aims to evaluate data augmentation strategies that preserve syntactic correctness of the generated data by analysis of the resulting parse trees. Our empirical findings reveal that such a method is capable of producing high-quality synthetic data, not only improving the baseline model's translation performance (as indicated by the BLEU score) on a hand-crafted test set containing sentences with rare words, but also on the original test data, thereby enhancing overall translation quality.

2 Preliminaries and Related Work

Fadaee et al. (2017) were among the first to implement a word replacement strategy to enhance data augmentation for addressing the OOV challenge. Their method involved substituting a frequently occurring word in the source sentence with a rarer counterpart and making a corresponding replacement in the target sentence with its translation.

However, the resulting synthetic sentences often lacked fluency.

Duan et al. (2020) attempted to address this limitation by introducing heuristics to validate replacements, utilising dependency parse trees to assess the structural role of words in sentences. They noted that words positioned closer to the root of the tree typically hold greater significance and should not be chosen for replacement over words that are closer to the leaf nodes.

Zhang and Zong (2016) developed two dictionary-based methods aimed at enhancing the performance of NMT systems on rare or unseen words. The first method used a hybrid word/character model to decompose problematic words into characters, which are presumed to be more common in the training set. The second method exploited the vast amounts of monolingual data available online for back-translation, creating synthetic parallel sentences with the aid of a bilingual dictionary through a Statistical Machine Translation (SMT) model. This ensured the frequent inclusion of rare words in the translation lexicon.

Building on Zhang and Zong (2016), Peng et al. (2020) proposed a similar dictionary-based method, but considered domain-specific information that was previously overlooked. In what Peng et al. (2020) termed as the "domain information gap", they argued that the lack of domain-specific information in the previous methods are likely to reduce the benefits offered by dictionary-based data augmentation, for example within the medical domain. Peng et al. (2020) utilised in-domain dictionaries to generate synthetic parallel sentences from an out-of-domain parallel corpus through phrase-replacement. This process involved filtering sentences based on the semantic similarity between dictionary terms and the source text, focusing on noun phrase replacements. Nonetheless, in both Zhang and Zong (2016) and Peng et al. (2020), the methods primarily focus on preserving semantic content, perhaps at the cost of syntactic accuracy.

3 Method

To combat rare words in the MT task, we looked into constituent substitution as our main strategy for augmenting the training data. We made use of the OpusMT en-zh pre-trained model (Tiedemann and Thottingal, 2020) to fine-tune with our dataset.

As with the intricate rules of natural languages, we took extra caution to ensure that the augmented data remain grammatically sound, to abide by the grammar rules. This was to ensure that the rules are well picked up by our model as it trains on the data. This was accomplished by paying attention to the Part-of-Speech (POS) of each word and ensuring that the parse trees of each sentence when augmenting the training data are maintained.

3.1 Constituent Substitution (`CONST`)

With a list of rare words, we augmented our training data by having new sentences that have constituents containing the rare words. This was carried out in 2 phases, namely to generate new English sentences that contain the rare words and then obtain the corresponding Chinese translations.

To generate new English sentences with each rare word, we identified the POS of the rare word that we are interested in. This was carried out using the Berkeley Neural Parser (Kitaev and Klein, 2018). This is a crucial first step to ensure that the English sentences generated are grammatically sound.

After obtaining the POS tags for each rare word, we randomly selected 50 sentences that contains at least one word which shares the same POS as the rare word. To ensure the grammatical soundness of the new sentences, we generated the parse trees of each sentence to identify the constituents in the sentences. The sentences are then grouped based on the smallest constituent that contains a word with the same POS as the rare word (see Figure 1).

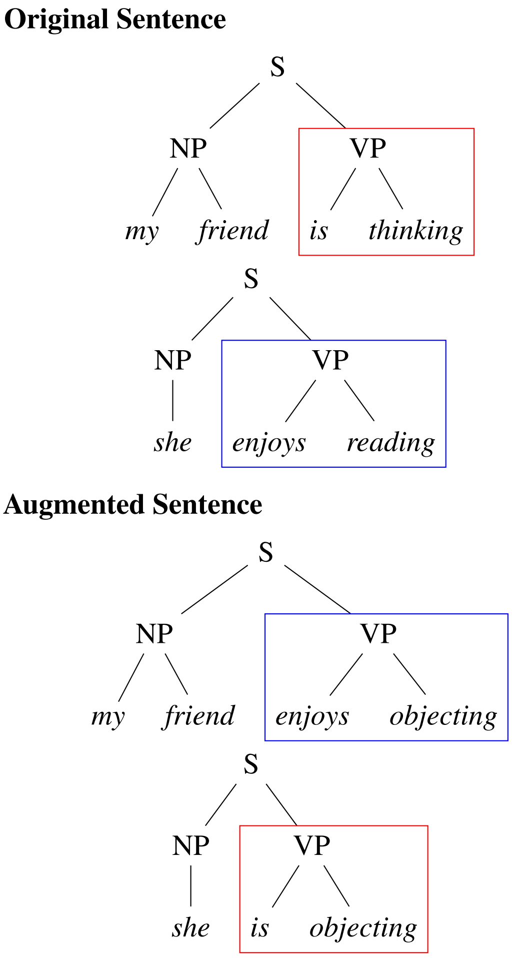

Figure 1: Augmenting the Original Sentence via Constituent Substitution with OOC Word "objecting"

Subsequently, within each group of sentences that has at least 2 sentences, the smallest constituents that contains a word with the same POS as the rare word were pairwise substituted, forming new sentences with newly substituted constituent from another sentence along with a rare word.

3.2 Chinese Translation Generation

Using a baseline model that was fine-tuned on the un-augmented training data, we generated the Chinese translations for the new English sentences from the augmentation steps described in the previous sub-section. These translations would then form sentence pairs with the new English sentences which we can then add to our original training data to form an augmented training dataset.

By relying on the training data we have, this approach of generating the Chinese translation ensures the consistency and coherence of the translations, leveraging the patterns and structures learned during training.

3.3 Rare Words Generation

The term frequency of each word in the training set was used to determine if it was rare. The vocabulary of the training set is compared against the vocabulary of the baseline model. Words in the vocabulary of the model that did not appear in the training set or appeared only once are compiled into a list. Rare words are chosen from the list in a manner that different type of words are selected. We refer to such words as out-of-corpus (OOC) words.

To evaluate the effectiveness of our approach in combating rare words in MT, we created a custom test set with sentence pairs that contain at least one of the rare words identified. These sentences were vetted through by the team, who are native speakers of both English and Chinese, to ensure their correctness.

4 Experiments

4.1 Setup

We evaluate data augmentation via constituent substitution, along with a host of other strategies (Section 4.2) on the task of English-Chinese MT.

The metric used for evaluation is the Bilingual Evaluation Understudy (BLEU) metric (Papineni et al., 2002), where a perfect translation match gives a score of 1.0 and a perfect translation mismatch gives a score of 0.0. In Section 4.3, we will tabulate and analyse the BLEU scores obtained by each model that was fine-tuned on the unique dataset that corresponds to an augmentation strategy.

In all experiments, we exclusively augment the training set while keeping the validation split untouched. Using each augmentation strategy, we generate 20000 more examples containing rare word(s) into the original training set. For each augmented English sentence, we obtain the corresponding Chinese sentence training pair by feeding the synthetic English sentence back into the baseline model to generate a translation. As per Section 3.3, a separate, hand-crafted test set was created to evaluate the models' performance on specifically translating sentences containing rare words.

The models were initialized using the OpusMT en-zh pre-trained model (Tiedemann and Thottingal, 2020), configured with a batch size of 64. Due to constraints on the available compute resources, training was limited to three epochs. Model performance was monitored on the validation set after every 1000 batches, selecting the checkpoint with the highest BLEU score for final evaluations on the test dataset. All other hyper-parameters remained aligned with the defaults specified in the provided code bases[1].

During pre-processing, each training batch underwent dynamic padding or truncation to match the longest sentence in the batch, followed by tokenisation into numerical identifiers corresponding to the model's vocabulary. Pad tokens were substituted with a special token to exclude them from the loss calculation process.

During inference, beam search with a beam width of 4 was chosen as the decoding strategy. This method maintains multiple hypotheses at each decoding step, selecting the sequence with the highest cumulative probability. This approach is preferred over a greedy algorithm as it allows for the consideration of high-potential sequences that may begin with less probable initial tokens.

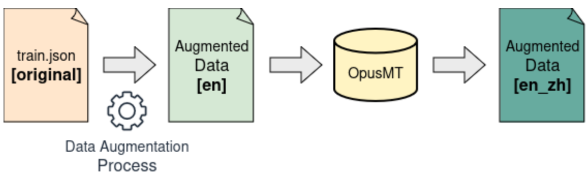

Figure 2: Data Augmentation Pipeline

4.2 Baselines

In addition to constituent substitution (CONST) as detailed in Section 3.1, the following data augmentation strategies were attempted:

4.2.1 No Augmentation (NoAUG)

The original training and validation sets are used.

4.2.2 Token Substitution (TOK & TOK+SYN)

One simple approach to augmenting the dataset involves substituting words in the training corpus with OOC words. This substitution process specifically targets words sharing the same POS tag as the OOC words, thereby ensuring grammatical consistency within the augmented data. For each OOC word, we repeatedly sample random sentences from the training corpus. For every sentence, we perform POS tagging, and the words that have the same POS tag as the current OOC word is swapped. These new sentences will be added to our training dataset.

[1] https://github.com/cs4248-33/project

(*Original*) And I think a lot of scientists have this attitude.

(*Augmented*) And I think a lot of scientists have this commonwealth.

Figure 3: Data Augmentation using TOK with OOC Word "commonwealth"

There are several reasons why token substitution can improve model performance compared to NoAUG. Firstly, token substitution generally introduces variations in the input data. This increased variation exposes the model to a broader range of linguistic structures and patterns, which can then help to improve its ability to generalise and make accurate translations. Secondly, since we are performing token substitution with OOC words, this helps to alleviate data sparsity issues and provide the model with more examples of rare or unseen patterns. Lastly, augmentation as a whole, acts as a form of regularisation as it adds noise to the training data. This prevents over-fitting by encouraging the model to learn more robust and generalisable representations.

4.2.3 Statistical Parsing (PARSE)

The key idea behind augmentation using statistical parsing is to generate the best sentences containing the OOC words to augment the training dataset, in order to expose the model to OOC words and new contexts while preserving the quality of training data. The measure of the quality of a sentence is determined using the BLLIP parser (Charniak and Johnson, 2005), which uses a generative parser that has obtained a good score on the Penn Treebank, to generate the maximum probability parse trees and select the best parse among all of them. Similar to the previous method of token substitution, only words in the original sentence which share the same POS tag as the OOC word is replaced with the OOC word, to preserve the grammatical integrity of the sentence, as well as reduce the computation required as compared to checking through all words for replacement. The new sentence with the best parse score among all new sentences created by substituting in an OOC word is then added to the training dataset. One challenge with this method is the high computational load of obtaining the parse scores using the BLLIP parser, which is not optimised for use with GPUs, which limits the practicality of this method. To speed up this process, multiprocessing was implemented. Some reasons why this augmentation method may not be optimal is the reliance on upstream systems trained on general datasets for pre-processing, such as spaCy's en_core_web_md to get the POS tags, as well as using the BLLIP parser to score and get the sentence with the best parse tree, and the fact that the best parse tree among all of the new sentences does not necessarily mean that a sentence is grammatically valid or realistic. This may affect the model's performance since it is trained on mostly grammatically valid data. Hence, while OOC words are introduced to the model, they may occur within unlikely contexts and may not help the model to learn useful representations and generalise.

4.3 Results

As indicated in Table 1, all of our data augmentation strategies resulted in BLEU scores that are higher than the baseline NoAUG.

Notably, there is a significant disparity in BLEU scores between the custom test set (**OOC**) and the **IWLST-2017** test split. This variation can be attributed primarily to differences in the datasets' composition. **OOC** was meticulously crafted by humans, and lacks the "noisiness" that is typical of the conversational style found in **IWLST-2017**, which consists of transcriptions from TED talks. Consequently, achieving higher BLEU scores on **OOC** tends to be less challenging compared to **IWLST-2017**.

Regardless, the nature of **OOC** is such that the dataset contains a high concentration of out-of-corpus words, and our empirical findings show that models trained on augmented data improved on the NoAUG baseline by a greater amount (in both absolute and relative terms) in the context of **OOC** as compared to **IWLST-2017**. This outcome aligns with our intuitions and underscores the effectiveness of augmentation strategies to improve model performance on text containing rare words.

Model	**OOC**	**IWLST-2017**
NoAUG	45.664	26.755
CONST_4	48.007 (+2.343)	27.166 (+0.411)
PARSE	47.726 (+2.062)	26.940 (+0.185)
TOK	48.215 (+2.551)	27.197 (+0.442)

Table 1: BLEU Scores for Each Augmentation Strategy

From our experimentation, we were surprised to see that TOK obtained the highest BLEU scores when tested against both the OOC and original test data. Specifically, it is slightly better than the CONST_4 model. This is surprising as we hypothesised CONST_4 to be superior because of its augmentation process of replacing at least 4 contiguous words each time, as opposed to TOK, which replaces a word each time. By replacing more contiguous words, we hypothesised that it would result in a more diverse training data which would empower the model to learn patterns more effectively and be able to perform better. However, such diversity seem to introduce more noise to the training data, which is a factor that affects its performance (Belinkov and Bisk, 2018).

By comparing the augmentation strategy implemented by TOK and CONST_4, though both are grammatically sound based on their parse trees, we noticed that the sentences generated for the former are semantically more sound. While both strategies do not account for the semantics of the sentences during the augmentation process, the TOK model would have a greater chance of obtaining a synthetic sentence after augmentation than the CONST_4 model because it only replaces a word each time with another that has the same POS. However, CONST_4 replaces constituents of at least 4 words each time with another. As constituents are combinations of words that function together as a single unit within a sentence, they inherently carry more meaning than individual words. Thus, it would be considerably harder for constituent replacements to be carried out while ensuring the semantical soundness of the sentence.

With regards to PARSE, a closer analysis of the augmented data reveals some grammatically invalid or grammatically valid yet nonsensical sentences. For example, (*Original*) These are the spaces that are not just luxurious spaces for some of us, but are important for everybody in this world.

(*Augmented*) These concurrent the spaces that are not just luxurious spaces for some of us, but are important for everybody in this world.

Figure 4: Grammatically Incorrect Data Augmentation using `PARSE` with OOC Word "concurrent"

(*Original*) We need to do for the ocean what Al Gore did for the skies above.

(*Augmented*) programmes need to do for the ocean what Al Gore did for the skies above.

Figure 5: Grammatically Correct but Nonsensical Data Augmentation using `PARSE` with OOC Word "programmes"

Such sentences may disrupt the model's ability to learn meaningful representations due to syntactic and semantic inconsistencies, and may hence degrade the model's performance.

5 Discussion

Our discussion seeks to evaluate how certain variables would affect the accuracy of translation. In particular, how would the length of constituents affect the quality of the model when using `CONST`, and to what extent are synonyms helpful in preserving sentence meanings when performing `TOK`. Through these questions, we hope to gain better insights into better ways of structure-preserving data augmentation.

5.1 Length of Constituents

Model	**OOC**	**IWLST-2017**
$\texttt{CONST}_2$	47.728	27.223
$\texttt{CONST}_3$	47.911	27.009
$\texttt{CONST}_4$	48.007	27.166
$\texttt{CONST}_5$	47.723	26.947

Table 2: BLEU Scores of Constituent Substitution Variants

To address the question on how the length of constituents affect the resulting BLEU scores, we repeated our experiment using the constituent substitution method for constituents of other lengths, namely 2, 3 and 5.

As discussed in Section 4.3, constituent substitution introduces greater diversity to the dataset, but at the same time, noise as well. From Table 2, we observe that $\texttt{CONST}_4$ scores the highest BLEU score on the OOC test, which suggests that there is a trade-off between diversity and noise. Compared to smaller sizes of constituents, such as in $\texttt{CONST}_2$ and $\texttt{CONST}_3$, $\texttt{CONST}_4$ would have a better diversity of training data, which explains why it has a higher BLEU score. Compared to having a larger constituent size, such as in $\texttt{CONST}_5$, $\texttt{CONST}_4$ would not have a training data that is as diverse, but it would have noise introduced to its dataset. It seems to suggest that the amount of noise introduced in $\texttt{CONST}_5$ outweighs the advantages brought about by its diversity, resulting in $\texttt{CONST}_4$ having a higher BLEU score.

5.2 Synonyms

As our first attempt in `TOK` did not consider synonyms, it may lead to sentences that are grammatically correct but semantically inconsistent. For example, replacing a noun with another noun of a completely different meaning could result in a nonsensical sentence which no longer conveys the intended message. In addition, it may also fail to capture important syntactic patterns or linguistic structures present in the data, which can be preserved by substituting with synonyms instead.

(*Original*) Some believe he wasn't the villain that he is commonly perceived to be.

(*Augmented*) Some believe he wasn't the scoundrel that he is commonly perceived to be.

Figure 6: Data Augmentation using `TOK+SYN` with OOC Word "scoundrel"

Hence, we conducted further investigation on `TOK` to obtain a new augmentation strategy (`TOK+SYN`) which only performs substitution of words that are synonymous, in addition to the original constraint of it having the same POS. WordNet (Miller, 1994) was used to check whether two words were synonymous.

Due to the restrictive nature of `TOK+SYN`, we trained it with a smaller augmented dataset and took appropriate actions to TOK to ensure a fair comparison.

Model	OOC	IWLST-2017
TOK (10K)	46.344	27.263
TOK+SYN (10K)	45.868	27.273

Table 3: BLEU Scores for TOK and TOK+SYN

While it is intuitive to think that synonyms would help in ensuring the quality of the training data, we observed that TOK+SYN performed worse than its counterpart which does not enforce the criteria of substituted words to be synonyms (see Table 3).

We hypothesise that though restricting substitutions to synonyms increases the semantic soundness of the augmented data, it does not bring about much diversity in the training data. With synonyms, the model may be able to learn the different words that can be used in a particular context, but since the overall structure and context of the augmented sentences remain similar, it would not add much benefit to its performance.

6 Conclusion

In this paper, we have explored several data augmentation strategies that focus on improving the MT performance of NMT systems on sentences containing OOC words. Our empirical investigation demonstrates the effectiveness of techniques such as token substitution, constituent substitution, and statistical parsing in generating synthetic data that enriches training sets with OOC words, thereby improving translation accuracy.

However, one key limitation of our approach is its applicability to low-resource languages. The quality and variety of augmented samples heavily depend on the availability of extensive data. Particularly for low-resource languages, our current data augmentation methods may not be practical due to the smaller datasets typically available, which limits the generation of high-quality augmented samples. Moreover, NMT systems are known to be sensitive to both natural and synthetic noise (Belinkov and Bisk, 2018), which could further compromise the quality of translations in such contexts. Addressing these challenges in low-resource scenarios remains a subject for future research.

Nevertheless, our approach proves to be effective in extending the dataset to rare words, which is valuable for niche domains where there are technical terminologies that may not be readily available.

7 Further Improvements

The complexity of data augmentation in MT tasks is multiplied by the requirement to generate accurate parallel sentences. Therefore, an effective augmentation strategy must not only be able to generate high-quality source sentences, but also ensure their translations are equally robust, all while maintaining scalability.

Our research did not delve much into the latter area, in order to maintain a manageable scope for our study given the constraints on time and resources. We opted to generate parallel corpora using our baseline model, given the simplicity of the method and its prevalent use in academic research. Nevertheless, alternative methods warrant exploration to possibly enhance the quality of the generated translations. For instance, Fadaee et al. (2017) employed automatic word alignments trained on bitexts to identify appropriate substitutions in parallel sentences.

Lastly, given that our results did not align with initial expectations, there remains ample scope for further investigation to devise more optimal data augmentation strategies. A promising direction for improving on our CONST strategy could involve integrating additional heuristics to refine the replacement process. For instance, the algorithm could also consider the positional importance of words in a dependency tree analysis, favoring leaf nodes over root-proximal words for substitutions (Duan et al., 2020). Additionally, Fadaee et al. (2017) proposed training a language model to evaluate the suitability of a word replacement within its contextual backdrop. Further work could look into experimenting with a more sophisticated variation of CONST_4 that layers on these heuristics to validate replacements. This multifaceted approach could pave the way for more effective and contextually accurate data augmentation in MT.

References

Yonatan Belinkov and Yonatan Bisk. 2018. Synthetic and natural noise both break neural machine translation.

Mauro Cettolo, Marcello Federico, Luisa Bentivogli, Jan Niehues, Sebastian Stüker, Katsuhito Sudoh, Koichiro Yoshino, and Christian Federmann. 2017. Overview of the IWSLT 2017 evaluation campaign. In *Proceedings of the 14th International Conference on Spoken Language Translation*, pages 2–14, Tokyo, Japan. International Workshop on Spoken Language Translation.

Eugene Charniak and Mark Johnson. 2005. Coarse-to-fine n-best parsing and MaxEnt discriminative reranking. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*, pages 173–180, Ann Arbor, Michigan. Association for Computational Linguistics.

Sufeng Duan, Hai Zhao, Dongdong Zhang, and Rui Wang. 2020. Syntax-aware data augmentation for neural machine translation.

Marzieh Fadaee, Arianna Bisazza, and Christof Monz. 2017. Data augmentation for low-resource neural machine translation. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 567–573, Vancouver, Canada. Association for Computational Linguistics.

Aloka Fernando and Surangika Ranathunga. 2022. Data augmentation to address out-of-vocabulary problem in low-resource sinhala-english neural machine translation.

Nikita Kitaev and Dan Klein. 2018. Constituency parsing with a self-attentive encoder. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2676–2686, Melbourne, Australia. Association for Computational Linguistics.

Tom Ko, Vijayaditya Peddinti, Daniel Povey, and Sanjeev Khudanpur. 2015. Audio augmentation for speech recognition. In *Proc. Interspeech 2015*, pages 3586–3589.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc.

George A. Miller. 1994. WordNet: A lexical database for English. In *Human Language Technology: Proceedings of a Workshop held at Plainsboro, New Jersey, March 8-11, 1994*.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Annual Meeting of the Association for Computational Linguistics*.

Wei Peng, Chongxuan Huang, Tianhao Li, Yun Chen, and Qun Liu. 2020. Dictionary-based data augmentation for cross-domain neural machine translation.

Jörg Tiedemann and Santhosh Thottingal. 2020. OPUS-MT — Building open translation services for the World. In *Proceedings of the 22nd Annual Conferenec of the European Association for Machine Translation (EAMT)*, Lisbon, Portugal.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need.

Jiajun Zhang and Chengqing Zong. 2016. Bridging neural machine translation and bilingual dictionaries.

Acknowledgements

The team would like to extend our heartfelt thanks to the CS4248 teaching team, and our project mentor (Luo Yang) in particular, for their invaluable support and guidance throughout this project. We are also extremely grateful to Prof. Min-Yen for his insightful advice and constructive feedback during our consultation session, which were instrumental in the successful completion of our project.

Statement of Independent Work

1A. Declaration of Original Work. By entering our Student IDs below, we certify that we completed our assignment independently of all others (except where sanctioned during in-class sessions), obeying the class policy outlined in the introductory lecture. In particular, we are allowed to discuss the problems and solutions in this assignment, but have waited at least 30 minutes by doing other activities unrelated to class before attempting to complete or modify our answers as per the class policy.

This particular document did not use any AI Tools to proofcheck and was constructed and edited purely by manual work.

Signed, A0244482A, A0235256E, A0217689I, A0217387W, A0220750U {e0865694, e0727256, e0543725, e0543423, e0555942}@u.nus.edu