The ARC-NKUA submission for the English-Ukrainian General Machine Translation Shared Task at WMT22

Dimitrios Roussis[1,2] and Vassilis Papavassiliou[1]
[1]Athena Research Center, Athens, Greece
[2]National and Kapodistrian University of Athens, Greece
{dimitris.roussis, vpapa}@athenarc.gr

Abstract

In what follows, we provide an overview of the ARC-NKUA ("Athena" Research Center - National and Kapodistrian University of Athens) submission to the WMT22 General Machine Translation shared task for the EN-UK (English to Ukrainian) and UK-EN (Ukrainian to English) translation directions. We describe how we constructed two Neural Machine Translation systems by training Transformer models (Vaswani et al., 2017), as well as our experiments involving: (a) ensemble decoding, (b) selected fine-tuning with a subset of the training data, (c) data augmentation with back-translated monolingual data, and (d) post-processing of the translation outputs. Furthermore, we discuss filtering techniques and the acquisition of additional data used for training the systems.

1 Introduction

Neural Machine Translation (NMT) has achieved significant improvements in translation quality in recent years, especially concerning high-resource language pairs. However, there is a lot of room for research on systems with general translation capabilities, underrepresented domains, low- or medium- resource language pairs, as well as multilingual systems. This year, the former news translation shared task widened in scope by introducing new domains, as well as the English-Ukrainian language pair among others.

We participated in the WMT22 General Machine Translation shared task for the unconstrained tracks of the EN-UK (English to Ukrainian) and UK-EN (Ukrainian to English) translation directions. The two submitted NMT systems are based on the Transformer architecture (Vaswani et al., 2017) and our experiments involve various methods and techniques such as data acquisition, filtering and selection, fine-tuning, ensemble decoding, tagged back-translation of English and Ukrainian monolingual sentences and post-processing of the translation outputs.

This paper is structured in the following way: In Section 2, we describe the parallel and monolingual corpora, as well as the acquisition, selection, filtering and pre-processing techniques that were used in our experiments. Section 3 outlines the NMT systems architecture, training parameters and the various experiments on top of our baseline systems. In Section 4, we report and discuss the experimental results of the two translation directions we participated in, while Section 5 concludes and summarizes our work.

2 Datasets

We participated in the unconstrained tracks of this year's general machine translation shared task for the English-Ukrainian and Ukrainian-English translation directions. We made use of most of the datasets given by the organizers: corpora from OPUS[1] (Tiedemann, 2012), ParaCrawl v9[2] and ELRC - EU acts in Ukrainian[3] from the ELRC-SHARE repository. Other parallel resources from this repository that were used in our systems include:

[1]https://opus.nlpl.eu/
[2]https://paracrawl.eu/news/item/17-english-ukrainian-bonus-parallel-corpus
[3]https://elrc-share.eu/repository/eu-acts-in-ukrainian/

Proceedings of the Seventh Conference on Machine Translation (WMT), pages 358–365
Abu Dhabi, December 7–8, 2022. ©2022 Association for Computational Linguistics

- Multilingual English, French, Polish to Ukrainian Parallel Corpus (processed)[4]
- Official web-portal of the Parliament of Ukraine, primary legislation[5]
- Official web-portal of the Parliament of Ukraine, Ukrainian laws in EN[6]
- Official web-portal of the Parliament of Ukraine, abstracts of UK laws[7]
- SciPar UK-EN-RU[8] (Roussis et al., 2022a)
- A Bilingual English-Ukrainian Lexicon of Named Entities Extracted from Wikipedia[9]

We also made use of three monolingual datasets given by the organizers: News crawl[10], Leipzig Corpora[11] and Legal Ukrainian Crawling[12] from the ELRC-SHARE repository. After manually inspecting the other given dataset, UberText Corpus[13], we decided not to use it for back-translation (see Section 3.2), as most punctuation is missing. Instead, we make use of monolingual corpora that we acquired (see Section 2.1), as well as the Ukrainian monolingual corpus of WikiMatrix.

2.1 Acquisition of Additional Corpora

In order to acquire additional parallel English-Ukrainian data, we used the ILSP-FC toolkit[14] (Papavassiliou et al., 2013) to crawl candidate parallel documents from websites and the LASER toolkit[15] (Artetxe and Schwenk, 2019) to mine bitexts with the use of its margin-based alignment score, after splitting each document into sentences. It is worth noting that manual inspection was also moderately applied so as to exclude machine translated websites. Additional parallel data acquisition techniques that were used are mentioned in more detail in Roussis et al. (2022a; 2022b). During parallel data acquisition, monolingual sentences in English and Ukrainian were also collected and were later used for back-translation (see Section 3.2).

The aforementioned techniques were used to compile the first five bulleted corpora listed in section 2, as well as EU acts in Ukrainian which was given by the organizers. Nevertheless, we attempted to enrich the acquired data by also targeting approximately 300 websites to extract EN-UK parallel sentences and more than 2,000 websites to extract monolingual UK sentences. This process resulted in ~2M additional EN-UK sentence pairs and ~31.9M monolingual UK sentences.

2.2 Parallel Corpus Filtering

The following filtering methods are used on all of the parallel data (including the subset that we selected for fine-tuning, as well as the synthetic data) after punctuation normalization and tokenization with the Moses toolkit[16] (Koehn et al., 2007):

- Sentence pairs with identical source and target sides are removed (Papavassiliou et al., 2018; Pinnis, 2018).
- Duplicate sentence pairs are removed, based on either source or target side; i.e. no English or Ukrainian sentence (after being lowercased and having its digits removed) appears more than once in the training set.

[4]https://elrc-share.eu/repository/multilingual-english-french-polish-to-ukrainian-parallel-corpus-processed/
[5]https://elrc-share.eu/repository/official-web-portal-of-the-parliament-of-ukraine-primary-legislation/
[6]https://elrc-share.eu/repository/official-web-portal-of-the-parliament-of-ukraine-ukrainian-laws-in-en/
[7]https://elrc-share.eu/repository/official-web-portal-of-the-parliament-of-ukraine-abstracts-of-uk-laws/
[8]https://elrc-share.eu/repository/scipar-uk-en-ru/
[9]https://elrc-share.eu/repository/a-bilingual-english-ukrainian-lexicon-of-named-entities-extracted-from-wikipedia/
[10]http://data.statmt.org/news-crawl
[11]https://wortschatz.uni-leipzig.de/en/download/ukr/
[12]https://elrc-share.eu/repository/legal-ukrainian-crawling/
[13]https://lang.org.ua/en/corpora/#anchor5
[14]http://nlp.ilsp.gr/redmine/projects/ilsp-fc/
[15]https://github.com/facebookresearch/LASER/
[16]https://github.com/moses-smt/mosesdecoder/

- Sentence pairs in which either side consists of more than 50% non-alphabetic characters are removed (Rikters, 2018).

- Sentence pairs in which the length ratio in terms of digit characters is over 2:1 (or below 1:2) are removed.

- Sentence pairs in which either the source or target sentence contains more than 250 tokens or more than 1000 characters are removed.

- Sentence pairs in which the token ratio between the longest and the shortest sentence is higher than 2 are removed.

- Sentence pairs in which either sentence contains letters not in the range of Unicode character sets relevant to Latin and Cyrillic scripts are removed (Papavassiliou et al., 2018).

- The repeating token filter [17] from Rikters (2018) was used to remove sentence pairs originating from machine-translated content.

- Language identification with fastText [18] (Joulin et al., 2017) is used to remove sentence pairs with different languages than expected.

In Table 1, we report the number of the raw (unfiltered) English-Ukrainian sentence pairs (57.7M), the number actually used for training the baseline systems after filtering (19M), and the selected subset used for fine-tuning (10.2M). Additionally, we list the number of filtered synthetic parallel sentences generated from English monolingual sentences with the EN-UK system (60M) and from Ukrainian monolingual sentences with the UK-EN system (54.5M).

2.3 Data Selection

As we will describe in more detail in Section 3.4, we also experimented with fine-tuning the NMT systems (see Section 3.3). In particular, after training a system we continue its training with a subset of the parallel data which has been selected according to some stricter criteria. LASER-based

Type of data	**Sentence pairs**
Raw EN-UK parallel	57,727,556
Filtered EN-UK parallel	19,023,045
Filtered EN-UK parallel selected for fine-tuning	10,203,198
Back-translated from monolingual EN	60,055,592
Back-translated from monolingual UK	54,517,999

Table 1: Number of used EN-UK sentence pairs

corpus filtering has been shown to have promising results (Chaudhary et al., 2019), it has already been computed for many of the used datasets and we believe that it may prove especially useful in counteracting possible quality degradation in NMT systems trained with additional back-translated data (Tran et al., 2021).

To this end, we decided to select an appropriate subset of the training data with the utilization of the alignment score given by the LASER toolkit. For this reason, the LASER scores of the parallel sentences of the available corpora were examined and the following data selection strategy was adopted:

- A LASER score threshold of 1.1 was set for sentence pairs originating from the CCMatrix, CCAligned and ParaCrawl corpora. These three datasets contain a total of 42.1M raw sentence pairs and have been collected from the web.

- A LASER score threshold of 1.06 was set for sentence pairs originating from the WikiMatrix corpus as well as for those that we acquired (see Section 2.1).

2.4 Pre-Processing and Vocabulary

As mentioned in section 2.2, the Moses toolkit is used to normalize the punctuation and tokenize the datasets. Additionally, in order to handle casing, we use the “inline casing” technique (Bérard et al., 2019; Etchegoyhen and Ugarte, 2020; Molchanov, 2020) which uses specific tags to denote uppercase (<UC>), title case (<TC>) or mixed case (<MC>) words. Depending on the tags which the decoder has generated, the output sentences are re-cased

[17] https://github.com/M4t1ss/parallel-corpora-tools/blob/master/parallel/repeating-tokens.php

[18] https://fasttext.cc/docs/en/language-identification.html

during post-processing. Inline casing has been shown as the optimal approach in handling casing (Etchegoyhen and Ugarte, 2020).

After the application of the filtering pipeline, as well as the addition of tags (from inline casing or tagged back-translation) and NFC Unicode normalization, a separate BPE tokenizer with 18k merge operations is trained independently for English and Ukrainian with SubwordNMT [19] (Sennrich et al., 2016a) and BPE-dropout with probability of 0.1 is applied on the source sentences for each translation direction (Provilkov et al., 2020).

3 System Overview

Both submitted systems follow the Transformer architecture (Vaswani et al., 2017) and were trained using two RTX 2080 Ti GPUs with the utilization of the Fairseq toolkit (Ott et al., 2019). In the subsections that follow, we describe the training process of both NMT systems, as well as the techniques that we experimented with in order to improve translation quality.

3.1 Model Architecture and Training

The "big Transformer" architecture (Vaswani et al., 2017) is used as our NMT model, although we made use of 8 encoder layers instead of 6, as increasing the number of encoder layers has been shown to improve performance in many scenarios (Subramanian et al., 2021; Wang et al., 2021b). We apply dropout with probability 0.3, activation dropout with probability 0.1 and attention dropout with probability 0.1. The Adam optimizer (Kingma and Ba, 2014) is used with a peak learning rate of 0.0007 after 4,000 warmup steps which then follows inverse square root decay. The models are trained using half precision training (FP16), with 2,800 tokens per batch, while the parameters are updated every 4 batches (Ott et al., 2018). Checkpoints are saved every 20,000 updates and every 10,000 updates when fine-tuning, while the training stops if the BLEU score on the validation set does not improve for 5 checkpoints. Finally, checkpoint averaging 5 was applied to all NMT systems, i.e., we average the parameters of the 5 last checkpoints in order to obtain the final model parameters.

[19] https://github.com/rsennrich/subword-nmt

3.2 Tagged Back-Translation

Back-translation (Sennrich et al., 2016b; Edunov et al., 2018) has been proven as an effective data augmentation technique which leverages large amounts of monolingual data and is particularly useful for domain adaptation and low-resource settings (Bérard et al., 2019; Wang et al., 2021a; Wang et al., 2021b). We follow Caswell et al. (2019) in using tagged back-translation, i.e., inserting a <BT> tag in the beginning of each source sentence which has been synthetically generated; a method which is simple and robust.

For each of the two translation directions, the reverse fine-tuned models trained on parallel data are used (with beam size 5) in order to generate the synthetic outputs (see Table 1). When we enrich the training set with back-translated data, we upsample the original parallel data by a factor of 2.

3.3 Selected Fine-Tuning

Fine-tuning is usually used to adapt a NMT model to a specific domain, i.e., to improve its quality on inputs with specific characteristics. Since this year the former news translation shared task changed its focus to more general translation capabilities, there is not a specific domain which we would like our systems adapted to.

Nevertheless, fine-tuning has also been shown to have a corrective effect on systems which exhibit decreased performance after having been trained with large amounts of synthetic data (Tran et al., 2021; Wang et al., 2021a). Thus, after the training of the NMT models ends, we continue to train them using a selected subset of the training set (see Section 2.3), while also halving the dropout probability to 0.15.

3.4 Ensemble Decoding

Ensemble decoding has been shown to have mostly minor effects on performance, although it can improve performance on specific translation directions (Oravecz et al., 2020; Tran et al., 2021; Subramanian et al., 2021; Wang et al., 2021a; Wang et al., 2021b). During inference, the probability distributions over the next token are averaged

#	System	EN - UK		UK - EN	
		FLORES101	WMT22	FLORES101	WMT22
(1)	Baseline	30.7	24.2	36.4	40.9
(2)	(1) + Selected Fine-Tuning	**31.0**	**24.4**	36.8	41.5
(3)	(1) + Back-Translation	30.5	23.7	37.4	40.9
(4)	(3) + Selected Fine-Tuning	30.8	24.0	37.7	41.7
(5)	Ensemble	30.7	24.0	**37.8**	**41.9**
WMT22	Best + Post-Processing	-	**25.2**	-	**41.9**

Table 2: BLEU scores on FLORES101 and WMT22 test sets for English to Ukrainian (EN-UK) and Ukrainian to English (UK-EN) systems.

according to the systems used in ensemble decoding.

It is generally better to use ensemble decoding with NMT systems trained with different seeds or different subsets of the training set (Oravecz et al., 2020; Subramanian et al., 2021). Unfortunately, hardware and time constraints did not allow us to follow this approach and thus, we experimented with ensembling 2 or 3 models from the resulting systems mentioned in the paper.

3.5 Post-Processing

In the WMT 2022 test data provided by the organizers, we observed specific peculiarities which were handled by post-processing scripts. In particular, the Ukrainian data used in the evaluation of the Ukrainian-English systems contained emojis which our systems were not able to handle. We used a simple post-processing script on the English outputs to copy emojis from the beginning or the end of the original Ukrainian input sentences. As regards the Ukrainian outputs of the English-Ukrainian systems, we used a script to replace double quotes ("…") with angled quotation marks («…»), as well as to fix anonymous placeholders according to their original style in the English inputs.

4 Results

We perform the evaluation of our systems using the FLORES101 test set (Goyal et al., 2022) and the WMT22 General Machine Translation test set given by the organizers. Scores are reported in terms of the detokenized case-sensitive BLEU score (Papineni et al., 2002) and have been computed with the SacreBLEU toolkit[20] (Post, 2018). In Table 2, we can see the resulting scores from our experiments, as well as the scores of the submitted models.

4.1 English to Ukrainian

The submitted NMT system for English to Ukrainian has been trained only with parallel data, fine-tuned with a subset of them (see Section 3.3) and its outputs have been post-processed (see Section 3.5). In Table 2, we can see that the effect of back-translation is negative for the EN-UK system. Selected fine-tuning exhibited a corrective effective which, nevertheless, was not enough to offset the initial degradation caused by the addition of synthetic data. However, we also obtain a small improvement (+0.2 BLEU on the WMT22 test set) when fine-tuning the baseline system trained only with parallel data. The largest increase in BLEU scores (+0.8) on the WMT22 test set, is observed after the application of post-processing on the outputs of the final system, which has been trained only with parallel data and fine-tuned on a selected subset of them. This increase does not concern the FLORES101 test set, since there are significant differences in the use of quotation marks between the two test sets. Finally, ensemble decoding did not provide any advantage in our experiments.

4.2 Ukrainian to English

As we can see in Table 2, back-translation initially degrades translation quality but, contrary to the results discussed in Section 4.1, ultimately leads to increased performance after fine-tuning with a selected set of the training data. Ensemble decoding usually has a marginal effect on NMT systems and we see a small increase by its use here as well. For this translation direction, we do not observe any significant difference after the application of post-processing, although we

[20] https://github.com/mjpost/sacreBLEU

decided to use it in the final system, since we do not believe it has any negative effects (less than 50 sentences were affected). Thus, the submitted NMT system for Ukrainian to English is based on all the techniques that we experimented with: back-translation (see Section 3.2), selected fine-tuning (see Section 3.3), ensemble decoding (see Section 3.4) and post-processing (see Section 3.5).

5 Conclusion

In this paper, we have presented the ARC-NKUA submission to the WMT22 General Machine Translation shared task for the English to Ukrainian and Ukrainian to English translation directions. The submitted systems follow the Transformer architecture and were determined after experimentation with back-translation, selected fine-tuning, and ensemble decoding. We showed that the corrective effect of fine-tuning with a subset of the training set can ultimately increase the translation quality of a system which has exhibited degradation due to having been exposed to a large number of synthetic data, while it also proved useful for systems trained only with parallel data.

Our systems underperformed in comparison with other submitted systems, according to automatic scores calculated by the organizers[21], although human judgements will be used for official ranking. In the future, we aim at better investigating the effects of acquiring additional parallel and monolingual data, following different filtering, selection and pre-processing strategies, as well as implementing several techniques which have been generally shown to increase translation quality, but hardware and time constraints did not allow us to experiment upon. Possible techniques that could be investigated include reranking, larger NMT model architecture, iterative back-translation, ensembling models trained on different subsets of the training set and exploiting higher-resource similar languages.

6 Acknowledgements

This work has been supported by the European Language Resource Coordination (ELRC), a service contract under the EC's Connecting Europe Facility SMART 2019/1083 programme.

[21]https://github.com/wmt-conference/wmt22-news-systems

References

Mikel Artetxe, and Holger Schwenk. 2019. Margin-based Parallel Corpus Mining with Multilingual Sentence Embeddings. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3197-3203, Florence, Italy. Association for Computational Linguistics.

Alexandre Bérard, Ioan Calapodescu, and Claude Roux. 2019. Naver Labs Europe's Systems for the WMT19 Machine Translation Robustness Task. In *Proceedings of the Fourth Conference on Machine Translation (Volume 2: Shared Task Papers, Day 1)*, pages 526-532, Florence, Italy. Association for Computational Linguistics.

Isaac Caswell, Ciprian Chelba, and David Grangier. 2019. Tagged Back-Translation. In *Proceedings of the Fourth Conference on Machine Translation (Volume 1: Research Papers)*, pages 53-63, Florence, Italy. Association for Computational Linguistics.

Vishrav Chaudhary, Yuqing Tang, Francisco Guzmán, Holger Schwenk, and Philipp Koehn. 2019. Low-Resource Corpus Filtering using Multilingual Sentence Embeddings. In *Proceedings of the Fourth Conference on Machine Translation (Volume 3: Shared Task Papers, Day 2)*, pages 261-266, Florence, Italy. Association for Computational Linguistics.

Sergey Edunov, Myle Ott, Michael Auli, and David Grangier. 2018. Understanding Back-Translation at Scale. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 489-500, Brussels, Belgium. Association for Computational Linguistics.

Thierry Etchegoyhen, and Harritxu G. Ugarte. 2020. To Case or not to case: Evaluating Casing Methods for Neural Machine Translation. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 3752-3760, Marseille, France. European Language Resources Association.

Naman Goyal, Cynthia Gao, Vishrav Chaudhary, Peng-Jen Chen, Guillaume Wenzek, Da Ju, Sanjana Krishnan, Marc'Aurelio Ranzato, Franscisco Guzmán, and Angela Fan. 2022. The Flores-101 Evaluation Benchmark for Low-Resource and Multilingual Machine Translation. *Transaction of the Association for Computational Linguistics*, 10:522-538.

Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. 2017. Bag of Tricks for Efficient Text Classification. In *Proceedings of the 15th*

Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers, pages 427-431, Valencia, Spain. Association for Computational Linguistics.

Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondřej Bojar, Alexandra Constantin, and Evan Herbst. 2007. Moses: Open Source Toolkit for Statistical Machine Translation. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions*, pages 177–180, Prague, Czech Republic. Association for Computational Linguistics.

Diederik P. Kingma, and Jimmy Ba. 2014. Adam: A Method for Stochastic Optimization. *arXiv preprint arXiv:1412.6980*.

Alexander Molchanov. 2020. PROMT Systems for WMT 2020 Shared News Translation Task. In *Proceedings of the Fifth Conference on Machine Translation*, pages 248-253, Online. Association for Computational Linguistics.

Csaba Oravecz, Katina Bontcheva, László Tihanyi, David Kolovratnik, Bhavani Bhaskar, Adrien Lardilleux, Szymon Klocek, and Andreas Eisele. 2020. eTranslation's Submissions to the WMT 2020 News Translation Task. In *Proceedings of the Fifth Conference on Machine Translation*, pages 254-261, Online. Association for Computational Linguistics.

Myle Ott, Sergey Edunov, David Grangier, and Michael Auli. 2018. Scaling Neural Machine Translation. In *Proceedings of the Third Conference on Machine Translation: Research Papers*, pages 1-9, Brussels, Belgium. Association for Computational Linguistics.

Myle Ott, Sergey Edunov, Alexei Baevski, Angela Fan, Sam Gross, Nathan Ng, David Grangier, and Michael Auli. 2019. fairseq: A Fast, Extensible Toolkit for Sequence Modeling. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations)*, pages 48-53, Minneapolis, Minnesota. Association for Computational Linguistics.

Vassilis Papavassiliou, Prokopis Prokopidis, and Gregor Thurmair. 2013. A modular open-source focused crawler for mining monolingual and bilingual corpora from the web. In *Proceedings of the sixth workshop on building and using comparable corpora*, pages 43-51, Sofia, Bulgaria. Association for Computational Linguistics.

Vassilis Papavassiliou, Sokratis Sofianopoulos, Prokopis Prokopidis, and Stelios Piperidis. 2018. The ILSP/ARC submission to the WMT 2018 Parallel Corpus Filtering Shared Task. In *Proceedings of the Third Conference of Machine translation: Shared Task Papers*, pages 928-933, Belgium, Brussels. Association for Computational Linguistics.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a Method for Automatic Evaluation of Machine Translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318, Philadelphia, Pennsylvania, USA. Association for Computational Linguistics.

Mārcis Pinnis. 2018. Tilde's parallel corpus filtering methods for WMT 2018. In *Proceedings of the Third Conference on Machine Translation: Shared Task Papers*, pages 939-945, Belgium, Brussels. Association for Computational Linguistics.

Matt Post. 2018. A Call for Clarity in Reporting BLEU Scores. In *Proceedings of the Third Conference on Machine Translation: Research Papers*, pages 186-191, Brussels, Belgium. Association for Computational Linguistics.

Ivan Provilkov, Dmitrii Emelianenko, and Elena Voita. 2020. BPE-Dropout: Simple and Effective Subword Regularization. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1882-1892, Online. Association for Computational Linguistics.

Matīss Rikters. 2018. Impact of Corpora Quality on Neural Machine Translation. In *Human Language Technologies–The Baltic Perspective*, pages 126-133. IOS Press.

Dimitris Roussis, Vassilis Papavassiliou, Prokopis Prokopidis, Stelios Piperidis, and Vassilis Katsouros. 2022a. SciPar: A Collection of Parallel Corpora from Scientific Abstracts. In *Proceedings of the 13th Language Resources and Evaluation Conference*, pages. 2652–2657, Marseille, France. European Language Resources Association (ELRA).

Dimitris Roussis, Vassilis Papavassiliou, Sokratis Sofianopoulos, Prokopis Prokopidis, and Stelios Piperidis. 2022b. Constructing Parallel Corpora from COVID-19 News using MediSys Metadata. In *Proceedings of the 13th Language Resources and Evaluation Conference*, pages 1068-1072, Marseille, France. European Language Resources Association (ELRA).

Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016a. Neural Machine Translation of Rare Words with Subword Units. In *Proceedings of the 54th Annual Meeting of the Association for*

Computational Linguistics (Volume 1: Long Papers), pages 1715-1725, Berlin, Germany. Association for Computational Linguistics.

Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016b. Improving Neural Machine Translation Models with Monolingual Data. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers),* pages 86-96, Berlin, Germany. Association for Computational Linguistics.

Sandeep Subramanian, Oleksii Hrinchuk, Virginia Adams, and Oleksii Kuchaiev. 2021. NVIDIA NeMo's Neural Machine Translation Systems for English-German and English-Russian News and Biomedical Tasks at WMT21. In *Proceedings of the Sixth Conference on Machine Translation (WMT)*, pages 197-204, Online. Association for Computational Linguistics.

Jörg Tiedemann. 2012. Parallel Data, Tools and Interfaces in OPUS. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 2214-2218, Istanbul, Turkey. European Language Resources Association (ELRA).

Chau Tran, Shruti Bhosale, James Cross, Philipp Koehn, Sergey Edunov, and Angela Fan. 2021. Facebook AI's WMT21 News Translation Task Submission. In *Proceedings of the Sixth Conference on Machine Translation (WMT)*, pages 205-215, Online. Association for Computational Linguistics.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is All you Need. *arXiv preprint arXiv: 1706.03762.*

Weixuan Wang, Wei Peng, Xupeng Meng, and Qun Liu. 2021a. Huawei AARC's Submissions to the WMT21 Biomedical Translation Task: Domain Adaptation from a Practical Perspective. In *Proceedings of the Sixth Conference on Machine Translation*, pages 868-873, Online. Association for Computational Linguistics.

Xing Wang, Tu Zhaopeng, and Shurning Shi. 2021b. Tencent AI Lab Machine Translation Systems for the WMT21 Biomedical Translation Task. In *Proceedings of the Sixth Conference on Machine Translation*, pages 874-878, Online. Association for Computational Linguistics.