CANTONMT: Cantonese to English NMT Platform with Fine-Tuned Models using Real and Synthetic Back-Translation Data

Kung Yin Hong, Lifeng Han, Riza Batista-Navarro, Goran Nenadic

Department of Computer Science, The University of Manchester

Oxford Rd, Manchester M13 9PL, United Kingdom

kenrick.kung@gmail.com

{lifeng.han, riza.batista, g.nenadic}@manchester.ac.uk

Abstract

Neural Machine Translation (NMT) for low-resource languages remains a challenge for many NLP researchers. In this work, we deploy a standard data augmentation methodology by back-translation to a new language translation direction, i.e., Cantonese-to-English. We present the models we fine-tuned using the limited amount of real data and the synthetic data we generated using back-translation by three models: OpusMT, NLLB, and mBART. We carried out automatic evaluation using a range of different metrics including those that are lexical-based and embedding-based. Furthermore, we create a user-friendly interface for the models we included in this project, CANTONMT, and make it available to facilitate Cantonese-to-English MT research. Researchers can add more models to this platform via our open-source CANTONMT toolkit, available at https://github.com/kenrickkung/CantoneseTranslation.

1 Introduction

Cantonese is one of the most popular dialects of Chinese languages, after the standard language Mandarin (the current official language in China, originally from the Beijing area), originally from the capital of Guangdong province, Guangzhou (a.k.a. Canton) in China. The population of Guangdong province was 129.51 million in 2022 according to the National Bureau of Statistics of China [1]. In addition, Cantonese is also the native language in Hong Kong (HK) and Macau regions which have populations of 7,503,100 and 704,149 in 2023, according to HK Census and Statistics Department[2] and Macrotrends Global Population statistics.[3] Furthermore, because of the economic growth in Guangdong, HK and Macau, many people from other Chinese provinces also learned to speak Cantonese for job purposes and due to cultural influences. There is also a large global population outside of China speaking Cantonese, 85.5 million, according to the Cantonese Language Association (CLA) [4]. In the era of the fast development of natural language processing (NLP), many machine translation (MT) models have been proposed for the majority of languages worldwide. However, *low-resource language MT* remains a challenge for researchers. Cantonese translation using MT, specifically, is under-explored and has not been given much attention thus far.

In this work, we investigate one of the more popular MT methods, i.e. synthetic data augmentation via back-translation and model fine-tuning, as an approach to Cantonese-to-English neural MT (NMT), along the way introducing Cantonese-English as a new language pair. We select several models for evaluation including both smaller and larger language models, and compare their system performance using a range of evaluation metrics. Furthermore, we open-source our toolkit and create a web-based user-friendly platform called **CantonMT** to facilitate research on Cantonese-English translation. A public video demo is available.[5]

In the next section (Section 2), we survey related work on Cantonese-English MT, data augmentation for MT, and available demos/engines. Section 3 introduces our methodology and framework. Section 4 explains the web-based CANTONMT platform. Section 5 concludes this work with a discussion.

2 Related Work

Research work focussing on Cantonese-English MT has not gained much attention to date. Earliest efforts include the work of Wu et al. (2006)

---

[1] https://data.stats.gov.cn/english
[2] https://www.censtatd.gov.hk/en/
[3] https://www.macrotrends.net/global-metrics/countries/MAC/macao

[4] https://cantoneselanguageassociation.byu.edu/
[5] CANTONMT demo https://youtu.be/s8P5fJjS7Ls

where example-based and rule-based MT were investigated. In recent years, a project plan on Cantonese-English Translation was put forward by researchers at the University of Hong Kong (HKU) where they proposed to investigate various MT approaches, including rule-based MT (RBMT), example-based MT (EBMT), statistical MT (SMT), gated-recurrent units (GRU) and transformers (Wing, 2020). More loosely related work include research in MT for Cantonese, but without English as the target language. These include dialectal translation between Cantonese and Mandarin Chinese by Zhang (1998), Yi Mak and Lee (2022) and Liu (2022).

Data augmentation via backtranslation has been one of the standard practices for generating a synthetic corpus for improving MT performance on low-resource language pairs. This has been popular for both statistical MT (SMT) and NMT (Sugiyama and Yoshinaga, 2019; Graça et al., 2019; Edunov et al., 2020; Nguyen et al., 2021; Pham et al., 2023). However, to the best of our knowledge, none of these efforts focused on Cantonese-to-English translation.

Existing platforms or off-the-shelf demos for Cantonese-to-English MT are very scarce. Popular MT engines from commercial IT companies, including Google Translate[6] and DeepL Translator,[7] do not include this language pair. Both of them only included simplified and traditional characters of Mandarin Chinese. Meanwhile, Microsoft Bing Translator[8] and Baidu, an IT company from China, made the Baidu Translator (Fanyi)[9] available, which includes Cantonese among several Chinese dialectal languages.[10] In the opposite direction, there are open-source tools for English-to-Cantonese MT from TransCan.[11]

3 Experimental Work

We introduce the methodology of CantonMT, experimental evaluations using the initial 38K real bilingual corpus, and extended model evaluations when we acquired 14.5K and 10K more real bilingual data from different sources.

3.1 Methodology and Framework

The methodology of this work is presented in Figure 1, which includes the following steps:

1. DataPrep: data collection and pre-processing

2. ModelFineTunePhase1: model selection for initial translator fine-tuning (ft, v1)

3. SynDataGenerate: synthetic data generation using the initial translator and cleaned data

4. ModelFineTunePhase2: second step MT fine-tuning using real and synthetic data (ft-syn)

5. ModelEval: model evaluation using both embedding-based metrics (BERTscore and COMET) and lexical metrics (SacreBLEU and hLEPOR)

For data collection, we scraped the data from the public Hong Kong forum LIHKG,[12] which was launched in 2016 and has multiple categories including sports, entertainment, hot topic, gossip, current affairs, etc. We extracted more than 1 million sentences from this website; however, the raw data comes with a lot of noise that needs to be cleaned, an example of which is shown in Figure 5 of Appendix A. We carried out data cleaning to reduce noisy strings as well as data *anonymisation* by removing user IDs from the text. We also filtered out the sentences that were too short, i.e., with less than 10 Chinese characters. In the end, we prepared 200K clean monolingual Cantonese sentences for parallel synthetic data generation purposes. We shuffled the data for model training.

In model fine-tuning phase 1, we aim to train a set of reasonable Cantonese-English MT models for synthetic data generation and model comparisons. The baseline models we selected are Opus-MT, NLLB and mBART. These were chosen to answer the following questions: (1) How much does model size impact fine-tuning performance? For this, we use Opus-MT which is a much smaller model trained on the Opus corpus using the MarianMT framework and NLLB-200, a very large language model pre-trained on 200+ languages from Mata-AI; (2) To what extent does it matter if the pre-trained translation models are exposed to Cantonese in their pre-training? For this, we add mBART (mbart-large-50-many-to-many-mmt) which is another LLM but without Cantonese in its
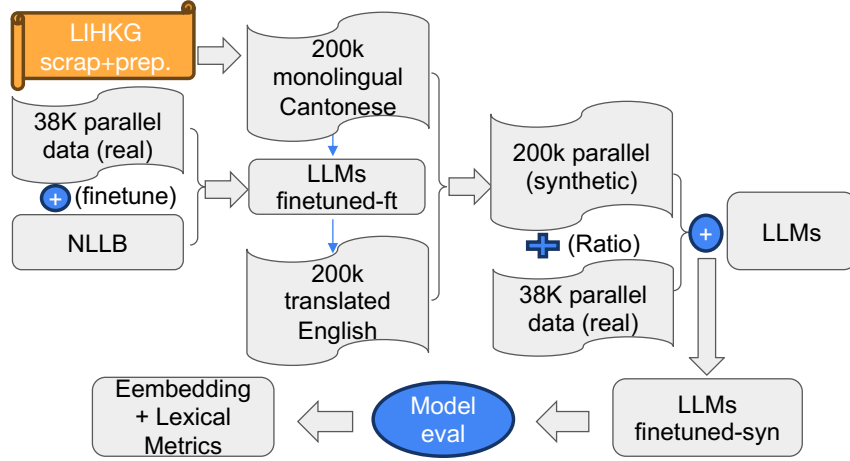

Figure 1: CANTONMT Pipeline: data collection and preprocessing, synthetic data generation, model fine-tuning, model evaluation

pre-training, vs NLLB which includes Cantonese. Because the full-size NLLB is too large, we used the distilled model `nllb-200-distilled-600M`.

We fine-tuned these models using the available bilingual data from a bilingual Cantonese-English dictionary called "Yue-Dian",[13] which is in total 44K in size. We divided this data into training, development and testing sets with 38K, 3K and 3K as their respective sizes, in light of the fact that the shared tasks organised by the Workshop on Statistical Machine Translation (WMT) tend to include around 3K sentences in their test sets.

In Step 3, synthetic data generation, we used the fine-tuned LLMs (LLM-ft-v1) from Step 2 to translate the monolingual Cantonese text we collected in Step 1. In this way, we obtain 200K back-translated English sentences; these synthetic sentences together with the Cantonese sentences create the 200K synthetic parallel corpus we generated. From now on, we will refer to the synthetic parallel corpus as `200K-ParaSyn`.

In Step 4, we apply different ratios on the real parallel data we have at hand and on 200K-ParaSyn for LLM fine-tuning. We also test the influence of model switches, i.e. using different types of LLMs for LLM-ft (Phase 1) and LLM-syn (Phase 2).

In the last step, we deploy the fine-tuned LLMs in Phase 2 (LLM-syn) on the same test data and compare the results with LLM-ft (Phase 1) and baseline models without fine-tuning. We also report comparisons with commerically available translation engines such as the Baidu Translator, Bing Translator and GPT4. The implementation

of GPT-4 that we used is Cantonese Companion, which was custom-made for translation to Cantonese by a community builder.[14]

We used a range of different evaluation metrics including the lexical-based SacreBLEU (Post, 2018) and hLEPOR (Han et al., 2013a, 2021), and the embedding-based BERTscore (Zhang* et al., 2020) and COMET (Rei et al., 2020). hLEPOR has reported much higher correlation scores to the human evaluation than BLEU and other lexical-based metrics on the WMT shared task data (Han et al., 2013b). However, recent WMT metrics task findings have demonstrated the advantages of neural metrics based on embedding space similarities (Freitag et al., 2022).

3.2 Evaluations of CANTONMT

The learning curves of three base models during training using the 38K real data are shown in Figure 2 from left to right for mBART, NLLB-200 and Opus-MT. We used three epochs for mBART because it is too large for the computational resources available to us. From the learning curves, we can see that NLLB-200 has a peak score at epoch 3 then there is a dramatic drop until epoch 6, followed by an increase until epoch 10. In contrast, the Opus-MT model achieves a steady increase in its SacreBLEU score with more epochs, although there are little drops in between.

The automatic evaluation scores from CANTONMT models and other commercial engines are listed in Table 1. Below are some interesting findings from the evaluation outcomes.
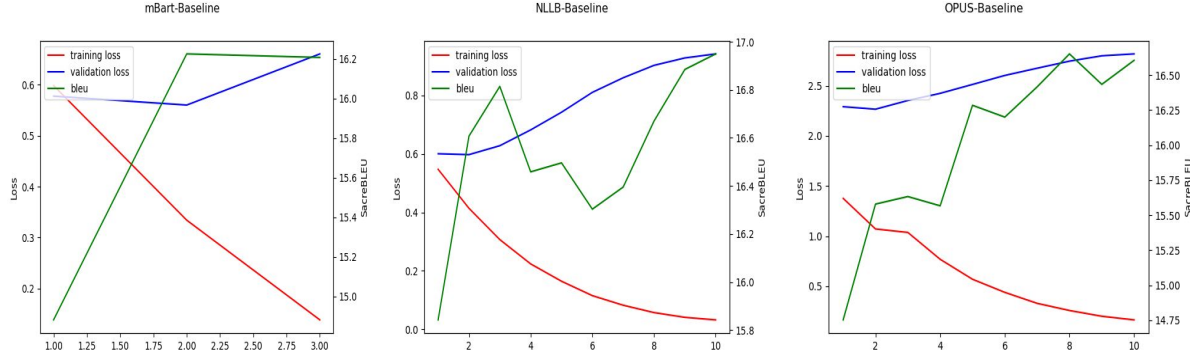

Figure 2: Learning curves during model training using real data.

- LLM-ft vs -LLM-ft-syn: (1) NLLB-syn-1:1 has slightly better scores than NLLB-bl on all metrics, but increasing the ratio of synthetic data will decrease the scores such as in the 1:3 and 1:5 configurations, with around 1 absolute SacreBLEU point. (2) Similarly, mBART-syn-1:1 also outperforms mBART-ft but increasing the ratio of synthetic data will reduce the evaluation scores such as in the 1:3 configuration. (3) Surprisingly, the synthetic model for Opus-mt does not outperform Opus-ft-bl, which indicates that the quality of the generated synthetic data matters.

- Model Switching Matters: (1) the NLLB fine-tuned model using synthetic data from mBART (second model from the top of the table) produced higher scores than using the synthetic data generated from its own (first model from the top of the table). (2) mBART fine-tuned using NLLB-generated synthetic data also outperforms mBART fine-tuning using only bilingual real data. (3) In a similar situation, Opus-MT performs differently in comparison to the other two models.

- Commercial MT models: (1) GPT4-finetuned produced the highest evaluation scores but the free version of GPTs restricts the input number of strings; the data size used for fine-tuning GPT-4 is unknown and such data is not publicly available to researchers; furthermore, it is unclear how GPT-4 performs MT; in addition, there are risks to data privacy when users choose to use engines from commercial companies. In contrast, CANTONMT is open-source, free, and researchers can continue to fine-tune it with their data or include more models, and is fully *confidential* for users.

2) Bing and Baidu translators produced similar evaluation scores to the best system from CANTONMT, though Bing produced slightly higher scores than Baidu, especially on the lexical-based metrics SacreBLEU and hLE-POR.

- Comparing to Model Deployment without Finetuning: in Cluster 3 (bottom) of Table 1, model deployment without fine-tuning has much lower scores; these scores show that fine-tuning and synthetic data augmentation lead to a large increase in scores of around 50% for all models using SecreBLEU.

3.3 Adding More Real Data

In the extension of our work, we managed to fine-tune the baseline models using more real data from another source called Wenlin[15] where we obtained another 14.5K parallel Cantonese-English dictionary. We are curious about the model performance using more real data in addition to the 38K training corpus from words.hk. We listed the comparison scores in the second cluster of Table 1 where it shows that the newly fine-tuned NLLB-200 using 52.5K data (38+14.5K=52.5K) produced higher scores on all metrics in comparison to 38K trained model; mBART fine-tuned using 52.5K obtains better scores on three metrics except for Sacre-BLEU; Opus-MT surprisingly did not get any increase across the metrics. Nevertheless, these outcomes demonstrated the possibility of improving model performance with more available real data, at least for the NLLB and mBART models. Moreover, **data quality matters**: simply adding 14.5K real data to fine-tune NLLB produced higher scores (underlined scores) than the best synthetic system

---

[15] https://wenlin.com/

Model Name	SacreBLEU	hLEPOR	BERTscore	COMET
nllb-forward-bl	16.5117	0.5651	0.9248	0.7376
nllb-forward-syn-h:h	15.7751	0.5616	0.9235	0.7342
nllb-forward-syn-1:1	**16.5901**	0.5686	**0.925**	**0.7409**
nllb-forward-syn-1:1-10E	16.5203	**0.5689**	0.9247	0.738
nllb-forward-syn-1:3	15.9175	0.5626	0.924	0.7376
nllb-forward-syn-1:5	15.8074	0.562	0.9237	0.7386
nllb-forward-syn-1:1-mbart	**16.8077**	**0.571**	**0.9256**	**0.7425**
nllb-forward-syn-1:3-mbart	15.8621	0.5617	0.9246	0.7384
nllb-forward-syn-1:1-opus	16.5537	0.5704	0.9254	0.7416
nllb-forward-syn-1:3-opus	15.9348	0.5651	0.9242	0.7374
mbart-forward-bl	15.7513	0.5623	0.9227	0.7314
mbart-forward-syn-1:1-nllb	**16.0358**	**0.5681**	**0.9241**	**0.738**
mbart-forward-syn-1:3-nllb	15.326	0.5584	0.9225	0.7319
opus-forward-bl-10E	**15.0602**	**0.5581**	**0.9219**	**0.7193**
opus-forward-syn-1:1-10E-nllb	13.0623	0.5409	0.9164	0.6897
opus-forward-syn-1:3-10E-nllb	13.3666	0.5442	0.9167	0.6957
baidu	16.5669	0.5654	0.9243	0.7401
bing	17.1098	0.5735	0.9258	0.7474
gpt4-ft(CantoneseCompanion)	**19.1622**	**0.5917**	**0.936**	**0.805**
nllb-forward-bl-plus-wenlin14.5k	*16.6662*	*0.5828*	*0.926*	*0.7496*
mbart-forward-bl-plus-wenlin14.5k	15.2404	0.5734	0.9238	0.7411
opus-forward-bl-plus-wenlin14.5k	13.0172	0.5473	0.9157	0.6882
nllb-200-deploy-no-finetune	11.1827	0.4925	0.9129	0.6863
opus-deploy-no-finetune	10.4035	0.4773	0.9082	0.6584
mbart-deploy-no-finetune	8.3157	0.4387	0.9005	0.6273
nllb-forward-all3corpus	*16.9986*	*0.583*	*0.927*	*0.7549*
nllb-forward-all3corpus-10E	16.1749	0.5728	0.9254	0.7508
mbart-forward-all3corpus	16.3204	0.5766	0.9253	0.7482
opus-forward-all3corpus-10E	14.4699	0.5621	0.9191	0.7074

Table 1: Automatic Evaluation Scores from Different Models in CANTONMT. bl: bilingual real data; syn: synthetic data; h:h - half and half; 1:1/3/5 - 100% real + 100/300/500% synthetic; 10E: 10 epochs (default: 3); top-down second slot: model switch: model type using NLLB but synthetic data from other models (mBART and OpusMT); top-down third slot: including model switch for mBART fine-tuning using synthetic data generated from NLLB; similarly top-down forth slot: including model switch for OpusMT fine-tuning using synthetic data from NLLB. Bottom slot of Cluster 1: Bing/Baidu Translator and GPT4-finetuned Cantonese Companion; **bold** case is the best score of the same slot among the same model categories. Cluster 2: bilingual fine-tuned models using 38K words.hk data plus 14.5k Wenlin data; *italic* indicates the number outperforms the same model fine-tuned with less data 38K. Cluster 3: Deployed Model without fine-tuning Cluster 4: Finetuned with the previous 2 corpora and an additional 10K data from OPUS Corpora we managed to find in the end - it shows the evaluation improvement continues.

that used 38x2=76K data.

4 CANTONMT Platform

To further facilitate Cantonese-English MT research and for users to easily access freely available fine-tuned models, we developed a user-friendly interface for the CantonMT platform. Users can choose different models and translation directions (Cantonese⇔English) via the interface (Figure 4 in the Appendix). The web application contains two main parts, the Interface and the Server.

4.1 User Interface

To test the user interface and different models for translation, users can choose from different model types and source languages, which dynamically capture the available models in the server, and allow users to select different training methods for
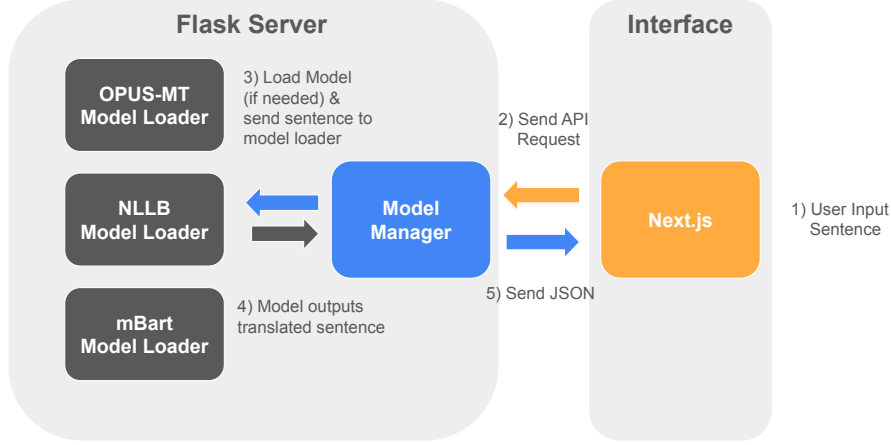
Figure 3: CANTONMT Server and Interface Flowchart diagram.

the model. One can then type the source sentence in the input box and click the "Translate" button to obtain the translation output from the model. The application layout is quite modular in case different model types or languages are added to the system, which could potentially be used as a base framework for different translation systems. It is possible to simply add more languages to the input and output if one wishes to expand the implementations. The look-and-feel of this web application is based on a template (Wrigley, 2023) for an AI Code translator, which was customised and developed in TypeScript with the Next.js framework. The reason for choosing this framework is that it provides a very modern and minimalistic approach.

4.2 Server

A diagram outlining the modules can be seen in Figure 3 to understand the general structure of the server. Users can easily run the server on their local machines by following the instructions provided in a README file. The server has two main functionalities, where the first one will output the list of model paths given the model type and source languages. The second one provides the translation, where one could provide the details of the model and also the sentence in the language specified, and the server would respond with the translated sentence using the model output.

During our implementation, due to memory constraints, the server crashed multiple times on our local machine. To mitigate the risk of server crashes, a *model manager* was produced, which implements a Least Recently Used (LRU) cache for the different model loaders, where the least recently used model will be deleted from memory if it exceeds the limit of the number of models. The server is built entirely based upon the *Python Flask* library. The reason for choosing this framework is that the models can be run on top of the Python Transformers library, which provides seamless implementation without much additional effort.

5 Discussion and Conclusion

In this work, we investigated the back-translation methodology for bilingual synthetic data generation for the sake of data augmentation for NMT, on a new language translation direction, Cantonese-to-English. We tested both smaller-sized OpusMT and extra-large LLMs NLLB and mBART both using available bilingual real data and larger synthetic data. Our experiments show that all the fine-tuned models outperformed the baseline deployment models with large margins. Furthermore, the synthetic model nllb-syn-1:1-mbart produced higher scores using the model switch method compared to those without the model switch. Lastly, the best performing fine-tuned models have similar (or even higher) evaluation scores than the current commercially available translators of Baidu and Microsoft-Bing.

In terms of concerns of **data privacy** such as handling of sensitive data (e.g., in clinical applications related to health analytics of patient data (Han et al., 2024)), CANTONMT can be fully controlled by users without interference from any third parties. We open-source our platform so that researchers can continue to integrate new models into the toolkit to promote Cantonese-English MT. We also plan to carry out human evaluations on the outputs from different systems to get more insights into the system errors.

Limitations

The synthetic data generated in this work is based on the fine-tuned model using 38K words.hk bilingual dictionary corpus, the first corpus we managed to find. This restricted the synthetic data quality. In the following-up work, we plan to use the further fine-tuned model on all three corpora, words.hk, wenlin, and opus-10K, to generate better back-translated synthetic data. We expect this will improve the synthetic data fine-tuned models.

The whole procedure of how difficult it was to collect real Cantonese-English bilingual data shows that Cantonese-English MT is still at its beginning stage with many obstacles and challenges to public research.

Acknowledgements

References

Sergey Edunov, Myle Ott, Marc'Aurelio Ranzato, and Michael Auli. 2020. On the evaluation of machine translation systems trained with back-translation.

Markus Freitag, Ricardo Rei, Nitika Mathur, Chi-kiu Lo, Craig Stewart, Eleftherios Avramidis, Tom Kocmi, George Foster, Alon Lavie, and André F. T. Martins. 2022. Results of WMT22 metrics shared task: Stop using BLEU – neural metrics are better and more robust. In *Proceedings of the Seventh Conference on Machine Translation (WMT)*, pages 46–68, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics.

Miguel Graça, Yunsu Kim, Julian Schamper, Shahram Khadivi, and Hermann Ney. 2019. Generalizing back-translation in neural machine translation.

Aaron Li-Feng Han, Derek F. Wong, Lidia S. Chao, Liangye He, Yi Lu, Junwen Xing, and Xiaodong Zeng. 2013a. Language-independent model for machine translation evaluation with reinforced factors. In *Proceedings of Machine Translation Summit XIV: Posters*, Nice, France.

Aaron Li-Feng Han, Derek F. Wong, Lidia S. Chao, Yi Lu, Liangye He, Yiming Wang, and Jiaji Zhou. 2013b. A description of tunable machine translation evaluation systems in WMT13 metrics task. In *Proceedings of the Eighth Workshop on Statistical Machine Translation*, pages 414–421, Sofia, Bulgaria. Association for Computational Linguistics.

Lifeng Han, Serge Gladkoff, Gleb Erofeev, Irina Sorokina, Betty Galiano, and Goran Nenadic. 2024. Neural machine translation of clinical text: An empirical investigation into multilingual pre-trained language models and transfer-learning. *Frontiers in Digital Health*, 6:1211564.

Lifeng Han, Irina Sorokina, Gleb Erofeev, and Serge Gladkoff. 2021. cushLEPOR: customising hLEPOR metric using optuna for higher agreement with human judgments or pre-trained language model LaBSE. In *Proceedings of the Sixth Conference on Machine Translation*, pages 1014–1023, Online. Association for Computational Linguistics.

Evelyn Kai-Yan Liu. 2022. Low-resource neural machine translation: A case study of Cantonese. In *Proceedings of the Ninth Workshop on NLP for Similar Languages, Varieties and Dialects*, pages 28–40, Gyeongju, Republic of Korea. Association for Computational Linguistics.

Xuan-Phi Nguyen, Shafiq Joty, Thanh-Tung Nguyen, Kui Wu, and Ai Ti Aw. 2021. Cross-model back-translated distillation for unsupervised machine translation. In *International Conference on Machine Learning*, pages 8073–8083. PMLR.

Nghia Luan Pham, Van Vinh Nguyen, and Thang Viet Pham. 2023. A data augmentation method for english-vietnamese neural machine translation. *IEEE Access*, 11:28034–28044.

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.

Ricardo Rei, Craig Stewart, Ana C Farinha, and Alon Lavie. 2020. COMET: A neural framework for MT evaluation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2685–2702, Online. Association for Computational Linguistics.

Amane Sugiyama and Naoki Yoshinaga. 2019. Data augmentation using back-translation for context-aware neural machine translation. In *Proceedings of the Fourth Workshop on Discourse in Machine Translation (DiscoMT 2019)*, pages 35–44, Hong Kong, China. Association for Computational Linguistics.

Liu Hey Wing. 2020. Machine translation models for cantonese-english translation project plan.
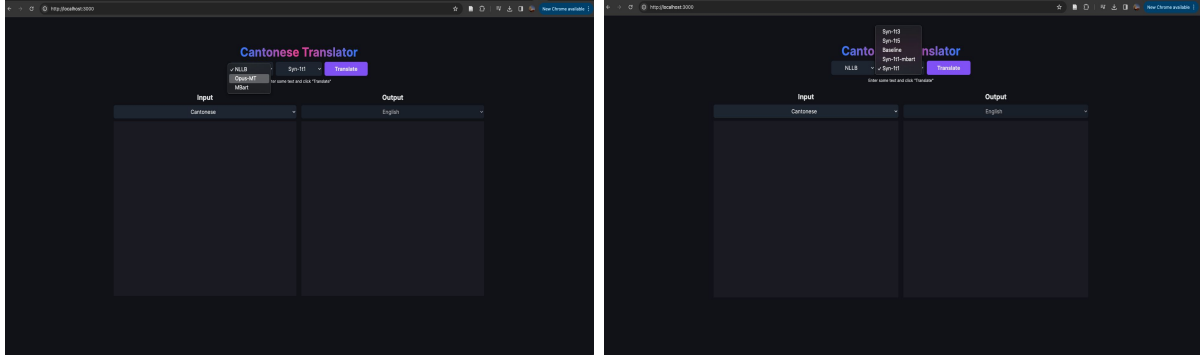
Mckay Wrigley. 2023. ai-code-translator.

Figure 4: CANTONMT Platform with options of model types, training categories, and translating directions. Frontend: TypeScript with Next.js. Backend: Python - Flask

Figure 5: Example text extracted from LIHKG website with lots noise before cleaning and anonymisation

Yan Wu, Xiukun Li, and Caesar Lun. 2006. A structural-based approach to Cantonese-English machine translation. In *International Journal of Computational Linguistics & Chinese Language Processing, Volume 11, Number 2, June 2006*, pages 137–158.

Hei Yi Mak and Tan Lee. 2022. Low-resource nmt: A case study on the written and spoken languages in hong kong. In *Proceedings of the 2021 5th International Conference on Natural Language Processing and Information Retrieval*, NLPIR '21, page 81–87, New York, NY, USA. Association for Computing Machinery.

Tianyi Zhang*, Varsha Kishore*, Felix Wu*, Kilian Q. Weinberger, and Yoav Artzi. 2020. Bertscore: Evaluating text generation with bert. In *International Conference on Learning Representations*.

Xiaoheng Zhang. 1998. Dialect MT: A case study between Cantonese and Mandarin. In *COLING 1998 Volume 2: The 17th International Conference on Computational Linguistics*.

A Appendix

Example raw text extracted from LIHKG website can be seen in Figure 5 before cleaning. CantonMT

user-friendly interface is shown in Figure 4 (Frontend: TypeScript with Next.js. Backend: Python - Flask).

The model parameters from OpusMT, extra-large NLLB and mBART are shown in Table 2, which shows that NLLB and mBART have doubled the number of transformer layers and have almost 10 times more parameters than OpusMT.

	Opus	NLLB	mBart
Layers	12	24	24
Hidden Unit	512	1024	1024
Model Parameters	77.9M	615M	610.9M
Language Pair	No	Yes	No
Release Year	2020	2022	2020

Table 2: Parameters from deployed models. Language pair: if the model contains Cantonese-English as a language pair