KreolMorisienMT: A Dataset for Mauritian Creole Machine Translation

Raj Dabre
NICT, Japan
raj.dabre@nict.gp.jp

Aneerav Sukhoo
University of Mauritius, Mauritius
aneeravsukhoo@yahoo.com

Abstract

In this paper, we describe KreolMorisienMT, a dataset for benchmarking machine translation quality of Mauritian Creole. Mauritian Creole (Kreol Morisien) is a French-based creole and a lingua franca of the Republic of Mauritius. KreolMorisienMT consists of a parallel corpus between English and Kreol Morisien, French and Kreol Morisien and a monolingual corpus for Kreol Morisien. We first give an overview of Kreol Morisien and then describe the steps taken to create the corpora. Thereafter, we benchmark Kreol Morisien↔English and Kreol Morisien↔French models leveraging pre-trained models and multilingual transfer learning. Human evaluation reveals our systems' high translation quality.

1 Introduction

Creoles[1] are natural languages that develop from the simplifying and mixing of different languages into a new one within a fairly brief period of time. Most creoles are highly related to a widely spoken language, and in this paper, we focus on Mauritian Creole, which is a French based creole. Mauritian Creole, or Kreol Morisien, is widely spoken in the republic of Mauritius by approximately 1.2 million people. Kreol Morisien is an important language from the perspective of tourism because Mauritius is a country well known for its tourism industry. Therefore, enabling tourists and locals to easily communicate with each other should not only help the tourism industry, but also improve cultural understanding. Machine translation of Creoles is quite under researched, mainly due to the lack of publicly available datasets. Although research has been conducted on Kreol Morisien translation in the past (Dabre et al., 2014; Boodeea and Pudaruth, 2020), datasets were not released publicly, making it difficult to reproduce and continue research.

In this paper, we describe KreolMorisienMT, a dataset containing standardized evaluation sets for benchmarking Kreol Morisien↔English and Kreol Morisien↔French translation. We first give an overview of Kreol Morisien followed by the description of the dataset creation process. We then use the evaluation sets to benchmark strong Neural machine translation (NMT) (Bahdanau et al., 2015) baselines trained using the created parallel corpora. We mainly rely on transfer learning (Zoph et al., 2016) through multilingual (Dabre et al., 2020) fine-tuning of pre-trained models based on mBART. By leveraging transfer learning, we can obtain a translation quality of about 23-25 BLEU for Kreol Morisien–English and about 20-23 BLEU for Kreol Morisien–French. We manually evaluate translations to better understand the impact of transfer learning. Our results show that there is significant room for innovation for Kreol Morisien NMT and Kreol Morisien NLP in general. Our datasets, models and human evaluation annotations are publicly available[2].

2 Related Work

This paper mainly focuses on the creation of datasets for under resourced languages, specifically creoles, as well as leveraging multilingualism and transfer learning to improve translation quality.

Mauritius is a part of East Africa, and KreolMorisienMT falls under the broad area of research focusing on African language machine translation. The Masakhane[3] community heavily focuses on African language NLP (Nekoto et al., 2020), a heavily under resourced area. With regard to creole translation, Haitian creole was the first creole language to receive substantial attention (Lewis, 2010) and was featured in a WMT shared task[4]. Work

[1] https://en.wikipedia.org/wiki/Creole_language

[2] https://github.com/prajdabre/KreolMorisienNLG

[3] https://www.masakhane.io/

[4] https://www.statmt.org/wmt11/

Findings of the Association for Computational Linguistics: AACL-IJCNLP 2022, pages 22–29
November 20–23, 2022.

French	Kreol Morisien	English
avion	avion	airplane
bon	bon	good
gaz	gaz	gas
anormalité	anomali	abnormality
colère	koler	anger
méditation	meditasion	meditation

Table 1: Similarities (top half) and differences (bottom half) between English, French and Kreol Morisien.

on Kreol Morisien itself was focused on a bit later by Sukhoo et al. (2014), Dabre et al. (2014), and Boodeea and Pudaruth (2020) but unlike us, they did not release their datasets. Motivated by work on Cree (Teodorescu et al., 2022), we decided to focus on the creation of publicly available standardized datasets for Kreol Morisien to/from English and French translation. On a related note, Lent et al. (2021) work on language models for Nigerian Pidgin and Haitian creole.

Kreol Morisien is a low-resource language where multilingualism (Dabre et al., 2020; Firat et al., 2016) and transfer learning approaches involving fine-tuning (Zoph et al., 2016) are most relevant. Self-supervised pre-trained models such as mBART (Liu et al., 2020) can be used, but they are not explicitly trained on Kreol Morisien. However, Dabre et al. (2022) showed that mBART like pre-trained models can be useful for unseen related languages, and we explore this possibility in this paper. Once strong baselines are trained, approaches such as back-translation (Sennrich et al., 2016) may be used to further improve translation quality, but we do not explore this given our limited size of monolingual corpus for Kreol Morisien.

3 Kreol Morisien

Kreol Morisien is spoken in Mauritius and Rodrigues islands, and a variant is also spoken in Seychelles. Mauritius was colonized successively by the Dutch, French and British. Although the British took over the island from the French in the early 1800, French remained as a dominant language and as such Kreol Morisien shares many features with French.

featured-translation-task.html

3.1 Kreol Morisien, English and French Similarities

Table 1 contains examples of words from French, Kreol Morisien and English. The same alphabet is used for all 3 languages, and in several cases words are either written or pronounced similarly. There are several words that are either identical, nearly identical or cognate pairs (Kanojia et al., 2020) between the 3 languages such *gaz* (gas) *avion* (airplane), *bon* (good), etc. On the other hand, despite similar pronunciations, in written French there is a heavy usage of accents which is absent in Kreol Morisien. An example is *anormalité* in French, which stands for *anomali* in Kreol Morisien meaning abnormality.

3.2 Kreol Morisien Grammar

The grammar of Kreol Morisien has been published in 2011 by Daniella Police-Michel in the book Gramer Kreol Morisien (Police-Michel et al., 2012). Kreol Morisien sentence structure follows the subject-verb-object order, the same as English and French. However, some similarities and differences with English and French can be noted as follows:

Adjective placement: Like French but unlike English, adjectives are sometimes placed after the object rather than before. *The brown bird* is translated as: *Zwazo maron-la*. Here, *maron* stands for *brown* and is moved after the object (Zwazo). The article *la* which stands for *the* is moved at the end of the sentence. On the other hand, the French translation would be *L'oiseau maron* which shows that Kreol Morisien is more grammatically similar to French in terms of adjective placement but differs in terms of article placement.

Singular-plural forms: Singular and plural forms are different between English and Kreol Morisien. *There are many birds* is translated as *Ena boukou zwazo* where the plural form *zwazo* does not take the suffix *s* as in English. Instead, the word *boukou* indicates *many* and therefore, it can be deduced that there are many birds. In French, the translated sentence is *Il y a beaucoup d'oiseaux* which has the same grammatical construction as in Kreol Morisien.

Verb dropping: Verbs are sometimes dropped in Kreol Morisien. *He is bad* is translated as *Li move* where *He* is translated to *Li* and *bad* to *move*. The verb *is* is dropped. Furthermore, in French, the translated sentence becomes *Il est méchant*, where

the verb is retained, indicating a difference from Kreol Morisien.

4 KreolMorisienMT

KreolMorisienMT is a mixed-domain dataset which was either created by manual translation of parts of Kreol Morisien and English books or by manual alignment of content in books that were already translated.

4.1 Data Sources

Our major sources are the holy Bible and story books. We used the online Bible from here[5]. Kreol Morisien sentences were manually aligned to their English and French counterparts to ensure high quality. Similarly, we had at our disposal 5 story books which were available in Kreol Morisien and English. However since we did not have PDF equivalents for most of the books, we ended up transcribing them. One such book which is available online is *The Flame Tree*[6] but manual alignment was done to ensure quality. We also created dictionaries, basic sentences and useful expressions manually from scratch for all 3 languages which account for most of the data. We expect dictionaries[7] to aid language learners. We included approximately 1500 basic expressions covering the following cases:, greetings, getting medical help, obtaining food from restaurants or supermarkets, simple conversations (weather, talking about oneself or others), money, accommodation.

The basic expressions should be useful for language learning as well as for use in a tourism setting. Due to the lack of human capital, not all content is translated into 3 languages, and there is more Kreol Morisien–English data than Kreol Morisien–French data. There is also a small amount of Kreol Morisien monolingual corpus, which we extracted mainly from untranslated books and online[8] articles. In the end, we obtained 23,310 and 16,739 pairs for English–Kreol Morisien and French–Kreol Morisien, respectively, as well as 45,364 Kreol Morisien monolingual sentences. The monolingual sentences are not in the Kreol Morisien side of the parallel corpus.

English–Kreol Morisien					
split	**L**	**AL-s**	**AL-t**	**U-s**	**U-t**
train	21,810	6.5	5.8	28,004	28,232
dev	500	16.9	16.2	2,330	2,164
test	1,000	17.0	16.0	3,700	3,323
French–Kreol Morisien					
split	**L**	**AL-s**	**AL-t**	**U-s**	**U-t**
train	15,239	2.6	2.0	16,171	16,754
dev	500	18.0	16.2	2,817	2,164
test	1,000	18.0	16.0	4,566	3,323
Kreol Morisien Monolingual					
split	**L**	**AL**	-		
-	45,364	15.8	-	52,425	-

Table 2: Corpora statistics for KreolMorisienMT. L, AL, U and -s/-t indicate #lines, average sentence length, #unique words and source/target language, respectively.

4.2 Dataset Statistics and Evaluation Splits

Of the 23,310 pairs for English–Kreol Morisien, 12,467 were dictionary entries. Similarly, for French–Kreol Morisien, of 16,739 pairs 12,424 were dictionary entries. Since the main goal is to develop translation systems that can translate full sentences, we decided to choose the longest sentences for the development and test sets. Furthermore, we decided to have trilingual evaluation sets following Guzmán et al. (2019) and Goyal et al. (2021). To this end, we first extracted a trilingual corpus of 13,861 sentences, sorted the corpora according to the number of words on the Kreol Morisien side and chose the top 1,500 ones representing the longest sentences. We then randomly chose 500 for the development set and 1,000 for the test set, both of which are trilingual. We remove the pairs from the English–Kreol Morisien, French–Kreol Morisien and Kreol Morisien corpora that overlap with the development and test set, resulting in 21,810, 15,239 sentence pairs and 45,364 sentences, respectively.

Table 2 contains an overview of the corpora. It is evident that there is a big mismatch between the length distributions of training and evaluation sets, but we prioritize the evaluation of medium to longer length sentences, so we have little choice.

5 Experiments

We describe the experimental settings including datasets used, training details, and models.

[5] https://www2.bible.com/en-GB/bible/344/MAT.1.NTKM2009

[6] https://shawkutis.weebly.com/uploads/1/9/7/4/19747661/flame_tree_lane_final.pdf

[7] Google translate is often used as a dictionary and we expect our dictionaries to enable out MT systems to act as dictionaries too.

[8] https://www.lalitmauritius.org/

5.1 Datasets

In addition to the parallel corpora from KreolMorisienMT, we use 5M randomly sampled sentence pairs from the UN corpus for French–English (Ziemski et al., 2016) which we use for pre-training a French↔English bidirectional NMT model which we contrast with the mBART-50 pre-trained denoising/MT models (Tang et al., 2021).

5.2 Training details

We train transformer (Vaswani et al., 2017) models using the YANMTT toolkit (Dabre and Sumita, 2021) which is based on the HuggingFace transformers library (Wolf et al., 2020). We use the training sets of KreolMorisienMT to create a joint English, French, Kreol Morisien 16,000 sub-words tokenizer using sentencepiece (Kudo and Richardson, 2018) for all our experiments except for fine-tuning mBART-50 based models. We do not extend the mBART-50 vocabulary. We tune hyperparameters as applicable (See Appendix A). Multilingual models are trained using the language indicator token proposed by Johnson et al. (2017). All models are trained to convergence on the relevant development sets, where convergence is said to take place if the development set BLEU score does not increase for 20 consecutive evaluations. BLEU scores are calculated using sacreBLEU with default parameters (Post, 2018). For decoding, we choose the model checkpoint with the highest validation set BLEU score and use a default beam size of 4 and length penalty of 0.8.

5.3 Models trained

We train and evaluate models for Kreol Morisien to English, English to Kreol Morisien, French to Kreol Morisien and Kreol Morisien to French. For each direction, we train:

Scratch: Unidirectional models.

Fine-tuned: Unidirectional and multilingual multiway models. We use 3 types of pre-trained models: our own English↔French models, denoising mBART-50 and its many-to-many fine-tuned version for MT from Tang et al. (2021).

6 Results

Table 6 compares unidirectional and multiway models trained from scratch and via fine-tuning.

Baselines: Owing to the tiny training set, most of which is a dictionary, unidirectional baseline models without any pre-training show poor performance of <10 BLEU. This is especially the case for translation involving French and Kreol Morisien. However, multiway models improve by up to 3.5 BLEU indicating the value of multilingualism.

Type	PT	Direction			
		cr-en	en-cr	cr-fr	fr-cr
Uni	-	9.1	9.9	4.6	5.6
Multi	-	11.1	11.5	7.9	9.3
Uni	Fr↔En	22.9	22.6	17.9	19.2
Multi	Fr↔En	22.7	22.5	19.9	22.4
Uni	MB-D	21.5	20.1	15.4	16.4
Multi	MB-D	22.3	20.8	18.3	21.0
Uni	MB-T	24.3	22.0	19.0	19.8
Multi	MB-T	**24.9**	**22.8**	**20.4**	**22.8**

Table 3: Unidirectional (Uni) and Multiway (Multi) model sacreBLEU scores with and without pre-training (PT) for translation involving Kreol Morisien (cr), English (en) and French (fr). Pre-trained models are: our own (Fr↔En), mBART-50 denoising (MB-D), and the many-to-many fine-tuned version of mBART-50 (MB-T) from Tang et al. (2021).

Fine-tuning: Both unidirectional and multilingual fine-tuning of the French↔English model trained on the UN corpus as well as the mBART-50 models leads to large improvements of >10 BLEU compared to their baseline counterparts. Especially, the performance of fine-tuning the mBART-50 models is impressive. mBART-50's vocabulary does not explicitly cover Kreol Morisien, but models fine-tuned on them still are comparable to or even outperform the French↔English model, which does. This shows the impressive power of massively multilingual models.

Denoising vs Translation Pre-training: Comparing the results of fine-tuning on the mBART-50 denoising model (MB-D) and its many-to-many translation version (MB-T) as well as the French↔English model (Fr↔En), we can see that in the absence of Kreol Morisien monolingual corpora for denoising pre-training, it is better to consider translation models for fine-tuning. However, denoising models perform reasonably well.

6.1 Human Evaluation

We randomly sample 50 examples from the test set for each translation direction and ask a native speaker of Kreol Morisien, French and English to rate the adequacy and fluency (Snover et al., 2009) of translations on a scale of 1 to 5. Additionally,

Input	Ena mem ki tom lor bann serviter, maltret zot e touy zot.
Reference	Others grabbed the servants, then beat them up and killed them.
Baseline	Some have been agreed on those servants, and they are murdered.
Fine-Tuned	Some people even fall on servants, maltreat them and kill them.
Input	"E natirelman mo prezant mo bon kamarad, Mourgat", Madam Ourit finn kontinie.
Reference	Mrs Octopus continued, "And naturally, I present my good friend Mr Squid".
Baseline	"Hey, I've got a good friends, Mr Octopus."
Fine-Tuned	"Hey obviously I present my good friend, Squid", Mrs Octopus went on.

Table 4: Examples for Kreol Morisien to English translation.

Direction	Adequacy	Fluency	#Perfect
cr-en	3.44	4.44	26
en-cr	3.73	4.35	40
cr-fr	2.64	3.70	12
fr-cr	3.30	4.24	26

Table 5: Adequacy, fluency and number of perfect translations out of 50 examples rated by a native speaker.

we ask the speaker to mark perfect translations. Due to lack of human power, we only evaluate the best system from Table 3. Annotations are in our public repository. Table 5 contains the results. Comparing Tables 3 and 5, the human evaluation scores appear to be correlated with BLEU. Kreol Morisien to French translation was rated to be of poorer quality compared to other directions. This can be attributed to the smaller training data size, the higher linguistic complexity of French than Kreol Morisien. Additionally, more than half of the translations were rated perfect with room for improvement. This shows that BLEU might underestimate the quality of translations.

6.2 Translation Examples

Table 4 contains examples generated by our MT systems for Kreol Morisien to English translation.

In the first example, taken from the holy Bible, the baseline system mistakes the act of *grabbing the servants* for *agreeing with the servants* and misses the part where the *servants are beaten up*. On the other hand, the fine-tuned model manages to capture both phenomenon properly. Both systems make the mistake of translating *others* as *some*, but this is understandable because a translation of the word *ena* in Kreol Morisien in English is *some*. The fine-tuned system also uses the word *maltreat* instead of *beat* and while this does reduce the adequacy of the translation, the general meaning is conveyed properly.

In the second example, taken from a story book, and the baseline system completely mistranslates the Kreol Morisien sentence. However, the fine-tuned model, except for the placement of the phrase *Mrs Octopus went on* to the end of the sentence and the imprecise translation of *natirelman* to *obviously*, translates almost perfectly. In the reference, *Mrs Octopus continued* is at the beginning of the sentence, and in the translation, *Mrs Octopus went on* is at the end of the sentence. The equivalent of *Mrs Octopus went on* in Kreol Morisien, *Madam Ourit finn kontinie*, is also at the end of the sentence and this explains the positioning in the translation. Multiple references and metrics may help in better evaluation by not penalizing such translations.

7 Conclusion

We have presented KreolMorisienMT, a dataset for machine translation between Mauritian Creole (Kreol Morisien) to/from English and French. Our datasets contain dictionary and sentence pairs belonging to a mix of domains and their sizes range from roughly 17,000 to 23,000 pairs. We also provide a monolingual corpus for Kreol Morisien containing about 45,000 sentences. We conduct translation experiments using KreolMorisienMT in conjunction with large English–French corpora and mBART-50 pre-trained models, leading to improvements of up to 15 BLEU, despite most of the training data being dictionary pairs. Adequacy and Fluency based human evaluation indicates high translation quality, despite BLEU scores being in the range of 20 to 25, indicating the need for better metrics. In the future, we plan to expand KreolMorisienMT with additional data as well as on additional generation tasks for Kreol Morisien. The Kreol Morisien monolingual corpus will be used in the future to extend pre-trained denoising models via light-weight adapter pre-training (Üstün et al., 2021).

References

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural machine translation by jointly learning to align and translate. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.

Zaheenah Boodeea and Sameerchand Pudaruth. 2020. Kreol morisien to english and english to kreol morisien translation system using attention and transformer model. *International Journal of Computing and Digital Systems*, 09(6):1143–1153.

Raj Dabre, Chenhui Chu, and Anoop Kunchukuttan. 2020. A survey of multilingual neural machine translation. *ACM Comput. Surv.*, 53(5).

Raj Dabre, Himani Shrotriya, Anoop Kunchukuttan, Ratish Puduppully, Mitesh Khapra, and Pratyush Kumar. 2022. IndicBART: A pre-trained model for indic natural language generation. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 1849–1863, Dublin, Ireland. Association for Computational Linguistics.

Raj Dabre, Aneerav Sukhoo, and Pushpak Bhattacharyya. 2014. Anou tradir: Experiences in building statistical machine translation systems for mauritian languages – creole, English, French. In *Proceedings of the 11th International Conference on Natural Language Processing*, pages 82–88, Goa, India. NLP Association of India.

Raj Dabre and Eiichiro Sumita. 2021. YANMTT: yet another neural machine translation toolkit. *CoRR*, abs/2108.11126.

Orhan Firat, Kyunghyun Cho, and Yoshua Bengio. 2016. Multi-way, multilingual neural machine translation with a shared attention mechanism. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 866–875, San Diego, California. Association for Computational Linguistics.

Naman Goyal, Cynthia Gao, Vishrav Chaudhary, Peng-Jen Chen, Guillaume Wenzek, Da Ju, Sanjana Krishnan, Marc'Aurelio Ranzato, Francisco Guzman, and Angela Fan. 2021. The flores-101 evaluation benchmark for low-resource and multilingual machine translation. Cite arxiv:2106.03193.

Francisco Guzmán, Peng-Jen Chen, Myle Ott, Juan Pino, Guillaume Lample, Philipp Koehn, Vishrav Chaudhary, and Marc'Aurelio Ranzato. 2019. The FLORES evaluation datasets for low-resource machine translation: Nepali–English and Sinhala–English. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 6098–6111, Hong Kong, China. Association for Computational Linguistics.

Melvin Johnson, Mike Schuster, Quoc V. Le, Maxim Krikun, Yonghui Wu, Zhifeng Chen, Nikhil Thorat, Fernanda Viégas, Martin Wattenberg, Greg Corrado, Macduff Hughes, and Jeffrey Dean. 2017. Google's multilingual neural machine translation system: Enabling zero-shot translation. *Transactions of the Association for Computational Linguistics*, 5:339–351.

Diptesh Kanojia, Malhar Kulkarni, Pushpak Bhattacharyya, and Gholamreza Haffari. 2020. Challenge dataset of cognates and false friend pairs from Indian languages. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 3096–3102, Marseille, France. European Language Resources Association.

Taku Kudo and John Richardson. 2018. SentencePiece: A simple and language independent subword tokenizer and detokenizer for neural text processing. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 66–71, Brussels, Belgium. Association for Computational Linguistics.

Heather Lent, Emanuele Bugliarello, Miryam de Lhoneux, Chen Qiu, and Anders Søgaard. 2021. On language models for creoles. In *Proceedings of the 25th Conference on Computational Natural Language Learning*, pages 58–71, Online. Association for Computational Linguistics.

William Lewis. 2010. Haitian Creole: How to build and ship an MT engine from scratch in 4 days, 17 hours, & 30 minutes. In *Proceedings of the 14th Annual conference of the European Association for Machine Translation*, Saint Raphaël, France. European Association for Machine Translation.

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.

Wilhelmina Nekoto, Vukosi Marivate, Tshinondiwa Matsila, Timi Fasubaa, Taiwo Fagbohungbe, Solomon Oluwole Akinola, Shamsuddeen Muhammad, Salomon Kabongo Kabenamualu, Salomey Osei, Freshia Sackey, Rubungo Andre Niyongabo, Ricky Macharm, Perez Ogayo, Orevaoghene Ahia, Musie Meressa Berhe, Mofetoluwa Adeyemi, Masabata Mokgesi-Selinga, Lawrence Okegbemi, Laura Martinus, Kolawole Tajudeen, Kevin Degila, Kelechi Ogueji, Kathleen Siminyu, Julia Kreutzer, Jason Webster, Jamiil Toure Ali, Jade Abbott, Iroro Orife, Ignatius Ezeani, Idris Abdulkadir Dangana, Herman Kamper, Hady Elsahar, Goodness Duru, Ghollah Kioko, Murhabazi Espoir, Elan van Biljon, Daniel Whitenack, Christopher Onyefuluchi, Chris Chinenye Emezue, Bonaventure F. P. Dossou, Blessing Sibanda, Blessing Bassey, Ayodele Olabiyi, Arshath Ramkilowan, Alp Öktem, Adewale Akinfaderin, and Abdallah Bashir. 2020. Participatory research for low-resourced machine translation: A case

study in African languages. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 2144–2160, Online. Association for Computational Linguistics.

D. Police-Michel, A. Carpooran, and G. Florigny. 2012. *Gramer kreol morisien: volim I. Dokiman referans*. Number v. 1 in Gramer kreol morisien. Akademi Kreol Morisien, Ministry of Education and Human Resources.

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, Belgium, Brussels. Association for Computational Linguistics.

Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016. 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.

Matthew Snover, Nitin Madnani, Bonnie Dorr, and Richard Schwartz. 2009. Fluency, adequacy, or HTER? Exploring different human judgments with a tunable MT metric. In *Proceedings of the Fourth Workshop on Statistical Machine Translation*, pages 259–268, Athens, Greece. Association for Computational Linguistics.

Aneerav Sukhoo, Pushpak Bhattacharyya, and Mahen Soobron. 2014. Translation between english and mauritian creole: A statistical machine translation approach. *2014 IST-Africa Conference Proceedings*, pages 1–10.

Xu Tan, Jiale Chen, Di He, Yingce Xia, Tao Qin, and Tie-Yan Liu. 2019. Multilingual neural machine translation with language clustering. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 963–973, Hong Kong, China. Association for Computational Linguistics.

Yuqing Tang, Chau Tran, Xian Li, Peng-Jen Chen, Naman Goyal, Vishrav Chaudhary, Jiatao Gu, and Angela Fan. 2021. Multilingual translation from denoising pre-training. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 3450–3466, Online. Association for Computational Linguistics.

Daniela Teodorescu, Josie Matalski, Delaney Lothian, Denilson Barbosa, and Carrie Demmans Epp. 2022. Cree corpus: A collection of nêhiyawêwin resources. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6354–6364, Dublin, Ireland. Association for Computational Linguistics.

Ahmet Üstün, Alexandre Berard, Laurent Besacier, and Matthias Gallé. 2021. Multilingual unsupervised neural machine translation with denoising adapters. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 6650–6662, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Proceedings of the Advances in Neural Information Processing Systems 30*, pages 5998–6008. Curran Associates, Inc.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander Rush. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online. Association for Computational Linguistics.

Michał Ziemski, Marcin Junczys-Dowmunt, and Bruno Pouliquen. 2016. The United Nations parallel corpus v1.0. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 3530–3534, Portorož, Slovenia. European Language Resources Association (ELRA).

Barret Zoph, Deniz Yuret, Jonathan May, and Kevin Knight. 2016. Transfer learning for low-resource neural machine translation. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1568–1575, Austin, Texas. Association for Computational Linguistics.

A Training and Hyperparameter Tuning Details

Models trained from scratch use the transformer-base architecture (Vaswani et al., 2017) whereas the French↔English model uses the transformer-big architecture. For models trained from scratch and those fine-tuned on our French↔English models, we varied the dropout, label smoothing and ADAM optimizer learning rates. Dropout values we considered were 0.1, 0.2 and 0.3. Label smoothing values considered were 0.1, 0.2 and 0.3. Learning rate values we considered were 10^{-3}, $3*10^{-3}$, 10^{-4} and $3*10^{-4}$. We found that the optimal dropout, label smoothing and learning rate values were 0.2, 0.2 and 10^{-4}, respectively. For fine-tuning mBART-50 and the many-to-many fine-tuned version of mBART-50 from Tan et al. (2019), we found that learning rate of $3*10^{-5}$, label smoothing of 0.1 and dropouts of 0.3 worked best. For pre-training our French↔English model, we use a learning rate of 10^{-3}, dropout of 0.1 and label smoothing of 0.1.