Leveraging Adapters for Improved Cross-Lingual Transfer for Low-Resource Creole MT

Marcell Fekete[1] **Ernests Lavrinovics**[1] **Nathaniel R. Robinson**[2]
Heather Lent[1] **Raj Dabre**[3] **Johannes Bjerva**[1]
[1]Aalborg University, Denmark; [2]Johns Hopkins University, USA; [3] IIT Madras, India
mrfe@cs.aau.dk

1 Creoles in Machine Translation

Creole languages are low-resource languages, often genetically related to languages like English, French, and Portuguese, due to their linguistic histories with colonialism (DeGraff, 2003). As such, Creoles stand to benefit greatly from both data-efficient methods and transfer-learning from high-resource languages. At the same time, it has been observed by Lent et al. (2022b) that machine translation (MT) is a highly desired language technology by speakers of many Creoles. To this end, recent works have contributed new datasets, allowing for the development and evaluation of MT systems for Creoles (Robinson et al., 2024; Lent et al., 2024). In this work, we explore the use of the limited monolingual and parallel data for Creoles using parameter-efficient adaptation methods. Specifically, we compare the performance of different adapter architectures over the set of available benchmarks. We find adapters a promising approach for Creoles because they are parameter-efficient and have been shown to leverage transfer learning between related languages (Faisal and Anastasopoulos, 2022). While we perform experiments across multiple Creoles, we present only on Haitian Creole in this extended abstract. For future work, we aim to explore the potentials for leveraging other high-resourced languages for parameter-efficient transfer learning.

2 Methodology and Experiments

To train adapters for Haitian, we use monolingual data from NLLB-OPUS (Fan et al., 2020), and the parallel CreoleM2M training split from CREOLEVAL (Lent et al., 2024). For evaluation, we leverage two evaluation datasets from CREOLEVAL: the CreoleM2M evaluation split and the MIT-Haiti Corpus for MT; we also evaluate over FLORES-200 (Goyal et al., 2022) (see Table 1).

All experiments are conducted with Kreyòl-MT, an mBART-50 model fine-tuned on the KREYÒL-MT dataset (Robinson et al., 2024). Furthermore, we apply dNLLB-200, a 600M-parameter distillation of the original 54B-parameter NLLB-200 model as a baseline (NLLB Team et al., 2022). Both models have 12 encoder and 12 decoder layers, 16 attention heads, and 1024 dimensions, and each model have their own model vocabularies of over 250,000 sentence-piece tokens shared across all languages.

Dataset	Domain	Size (#lines)
NLLB-OPUS	Web scrape	∼15M
FLORES-200◇	Wikipedia	3,001
CreoleM2M	Religion	208,772 (train) 1,000 (eval)
MIT-Haiti◇	Education	1,559

Table 1: Datasets used in our preliminary experiments. A ◇ indicates the dataset is used only as evaluation data.

Method	Source	Config name
Bottleneck	Houlsby et al., 2019	`double_seq_bn`
+ Invertible	Houlsby et al., 2019	`._inv`
Compacter	Mahabadi et al., 2021	`compacter`
LoRA	Hu et al., 2021	`lora`
$(IA)^3$	Liu et al., 2022	`ia3`

Table 2: Adapter architectures compared in our experiments (Table adapted from Poth et al., 2023). We also experiment with prefix tuning adapters (Li and Liang, 2021) and bottleneck adapters from Pfeiffer et al. (2020), which differ only from those of Houlsby et al. (2019) in adapter placement. However some preliminary experiments found they performed worse than these five.

2.1 Experiments

Following Üstün et al. (2021), we attempt to leverage monolingual data to improve MT performance by training denoising adapters added to the encoder, the decoder, or or both components of Kreyòl-MT model. Additionally, we experiment with or without cross-attention (CA) fine-tuning between

Proceedings of the Fourth Workshop on Multilingual Representation Learning (MRL 2024), pages 212–215
November 16, 2024 ©2024 Association for Computational Linguistics

	eng→hat			hat→eng		
eval set:	*CreoleM2M*	*FLORES*	*MIT-Haiti*	*CreoleM2M*	*FLORES*	*MIT-Haiti*
compacter	**42**	**28**	**35**	76	**40**	31
double-seq-bn	40	27	34	77	38	32
double-seq-bn-inv	41	26	**35**	76	37	32
ia3	**42**	**28**	34	77	39	32
lora	**42**	**28**	**35**	**79**	38	31
Kreyòl-MT w/ CA	**42**	27	**35**	75	**40**	31
Kreyòl-MT	33	27	32	66	**40**	30
NLLB	22	26	33	34	37	**36**

Table 3: Average BLEU scores across each evaluation benchmark. Different adapter methods are on top, while baselines are on bottom.

the components. We evaluate using a number of adapter architectures (see Table 2) which to our knowledge have not yet been directly compared against each other.

In preliminary experiments we narrowed down all adapters from AdapterHub[1] to the five best performing, as shown in Table 2. We compare appendage of these adapters to three baseline models: Kreyòl-MT out-of-the-box, Kreyòl-MT with CA fine-tuning, and the 600M-parameter NLLB-200 model (NLLB Team et al., 2022).

2.2 Results

We find that some adapter architectures are more amenable to Üstün et al. (2021)'s monolingual adaptation methodology, as demonstrated by their relative increased performance over baselines (see Table 3). However, these scores consistently drop as the quality and cultural relevance of the data increases (*i.e.*, we observe much better performance on the religious-domain samples from CreoleM2M, and worse performance on MIT-Haiti, which is culturally appropriate data sourced from the community). Regarded holistically, even the best adapters do not consistently improve over CA fine-tuning between encoder and decoder, and they either improve or degrade performance by only marginal amounts. Our results also suggest that CA fine-tuning generally helps performance.

3 Conclusion and Future Work

While gains over baselines were reached *via* the monolingual adaptation, most Creoles do not have lage amounts of web-scraped data, as found in NLLB-OPUS. Thus, the ability to leverage data and transfer from closely related languages to Creoles has great potential for bolstering Creole MT. To this end, we plan experiments for parameter-efficient transfer learning, inspired by Faisal and Anastasopoulos (2022), who found success with phylogenetically-motivated adaptation. The application of phylogenetic adaptation for Creoles is not straight-forward, however. There is no consensus phylogeny of Creoles or even their broader language families (Bakker et al., 2011; Aboh, 2016). Simultaneously, previous works have shown that transfer learning to Creoles from related languages is nontrivial (Lent et al., 2022a, 2024; Robinson et al., 2022, 2023). Thus an important area of Creole MT remains selecting favorable languages for transfer learning.

In addition to phylogentic relation, we are exploring selection of transfer languages via **embedding clustering**. We cluster NLLB-200 language token embeddings with cosine and Euclidean distance, and identify Afrikaans, Igbo, and Yiddish as the nearest neighbors of Haitian. These are interesting findings, since Igbo is one of Haitian's hypothesized relatives (Seguin, 2020), and Yiddish and Afrikaans are Indo-European languages influenced by Afroasiatic and South African Khoisan languages, respectively—which appears analogous to Haitian's mixed Indo-European and Niger-Congo influences. We also explore measuring vocabulary **subword evenness**, as introduced by Pelloni et al. (2022), as a more helpful language selection method than simple typological proximity. While experiments are still underway, these explorations will help establish the languages amenable to cross-lingual transfer for Creoles, and ultimately the degree to which cross-lingual adaptation methods can benefit speakers of Creoles.

[1] https://docs.adapterhub.ml/overview.html

References

Enoch O. Aboh. 2016. Creole distinctiveness: A dead end. *Journal of Pidgin and Creole Languages*, 31(2):400–418.

Peter Bakker, Aymeric Daval-Markussen, Mikael Parkvall, and Ingo Plag. 2011. Creoles are typologically distinct from non-creoles. *Journal of Pidgin and Creole languages*, 26(1):5–42.

Michel DeGraff. 2003. Against creole exceptionalism. *Language*, 79(2):391–410.

Fahim Faisal and Antonios Anastasopoulos. 2022. Phylogeny-inspired adaptation of multilingual models to new languages. In *Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 434–452, Online only. Association for Computational Linguistics.

Angela Fan, Shruti Bhosale, Holger Schwenk, Zhiyi Ma, Ahmed El-Kishky, Siddharth Goyal, Mandeep Baines, Onur Celebi, Guillaume Wenzek, Vishrav Chaudhary, Naman Goyal, Tom Birch, Vitaliy Liptchinsky, Sergey Edunov, Edouard Grave, Michael Auli, and Armand Joulin. 2020. Beyond english-centric multilingual machine translation. *Preprint*, arXiv:2010.11125.

Naman Goyal, Cynthia Gao, Vishrav Chaudhary, Peng-Jen Chen, Guillaume Wenzek, Da Ju, Sanjana Krishnan, Marc'Aurelio Ranzato, Francisco Guzmán, and Angela Fan. 2022. The Flores-101 evaluation benchmark for low-resource and multilingual machine translation. *Transactions of the Association for Computational Linguistics*, 10:522–538.

Neil Houlsby, Andrei Giurgiu, Stanisław Jastrzebski, and Bruna Morrone. 2019. Parameter-Efficient Transfer Learning for NLP. In *Proceedings of the 36th International Conference on Machine Learning*, page 10, Long Beach, California.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2021. LoRA: Low-Rank Adaptation of Large Language Models. *arXiv preprint*. ArXiv:2106.09685 [cs].

Heather Lent, Emanuele Bugliarello, and Anders Søgaard. 2022a. Ancestor-to-creole transfer is not a walk in the park. In *Proceedings of the Third Workshop on Insights from Negative Results in NLP*, pages 68–74, Dublin, Ireland. Association for Computational Linguistics.

Heather Lent, Kelechi Ogueji, Miryam de Lhoneux, Orevaoghene Ahia, and Anders Søgaard. 2022b. What a creole wants, what a creole needs. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 6439–6449, Marseille, France. European Language Resources Association.

Heather Lent, Kushal Tatariya, Raj Dabre, Yiyi Chen, Marcell Fekete, Esther Ploeger, Li Zhou, Ruth-Ann Armstrong, Abee Eijansantos, Catriona Malau, Hans Erik Heje, Ernests Lavrinovics, Diptesh Kanojia, Paul Belony, Marcel Bollmann, Loïc Grobol, Miryam de Lhoneux, Daniel Hershcovich, Michel DeGraff, Anders Søgaard, and Johannes Bjerva. 2024. Creoleval: Multilingual multitask benchmarks for creoles. *Preprint*, arXiv:2310.19567.

Xiang Lisa Li and Percy Liang. 2021. Prefix-tuning: Optimizing continuous prompts for generation. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 4582–4597, Online. Association for Computational Linguistics.

Haokun Liu, Derek Tam, Mohammed Muqeeth, Jay Mohta, Tenghao Huang, Mohit Bansal, and Colin Raffel. 2022. Few-Shot Parameter-Efficient Fine-Tuning is Better and Cheaper than In-Context Learning. *arXiv preprint*. ArXiv:2205.05638 [cs].

Rabeeh Karimi Mahabadi, James Henderson, and Sebastian Ruder. 2021. Compacter: Efficient Low-Rank Hypercomplex Adapter Layers. In *Advances in Neural Information Processing Systems*, volume 34, pages 1022–1035. Curran Associates, Inc.

NLLB Team, Marta R. Costa-jussà, James Cross, Onur Çelebi, Maha Elbayad, Kenneth Heafield, Kevin Heffernan, Elahe Kalbassi, Janice Lam, Daniel Licht, Jean Maillard, Anna Sun, Skyler Wang, Guillaume Wenzek, Al Youngblood, Bapi Akula, Loic Barrault, Gabriel Mejia-Gonzalez, Prangthip Hansanti, John Hoffman, Semarley Jarrett, Kaushik Ram Sadagopan, Dirk Rowe, Shannon Spruit, Chau Tran, Pierre Andrews, Necip Fazil Ayan, Shruti Bhosale, Sergey Edunov, Angela Fan, Cynthia Gao, Vedanuj Goswami, Francisco Guzmán, Philipp Koehn, Alexandre Mourachko, Christophe Ropers, Safiyyah Saleem, Holger Schwenk, and Jeff Wang. 2022. No language left behind: Scaling human-centered machine translation.

Olga Pelloni, Anastassia Shaitarova, and Tanja Samardzic. 2022. Subword evenness (SuE) as a predictor of cross-lingual transfer to low-resource languages. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 7428–7445, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Jonas Pfeiffer, Ivan Vulić, Iryna Gurevych, and Sebastian Ruder. 2020. MAD-X: An Adapter-Based Framework for Multi-Task Cross-Lingual Transfer. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 7654–7673, Online. Association for Computational Linguistics.

Clifton Poth, Hannah Sterz, Indraneil Paul, Sukannya Purkayastha, Leon Engländer, Timo Imhof, Ivan

Vulić, Sebastian Ruder, Iryna Gurevych, and Jonas Pfeiffer. 2023. Adapters: A unified library for parameter-efficient and modular transfer learning. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 149–160, Singapore. Association for Computational Linguistics.

Nathaniel Robinson, Raj Dabre, Ammon Shurtz, Rasul Dent, Onenamiyi Onesi, Claire Monroc, Loïc Grobol, Hasan Muhammad, Ashi Garg, Naome Etori, Vijay Murari Tiyyala, Olanrewaju Samuel, Matthew Stutzman, Bismarck Odoom, Sanjeev Khudanpur, Stephen Richardson, and Kenton Murray. 2024. Kreyòl-MT: Building MT for Latin American, Caribbean and colonial African creole languages. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 3083–3110, Mexico City, Mexico. Association for Computational Linguistics.

Nathaniel Robinson, Cameron Hogan, Nancy Fulda, and David R. Mortensen. 2022. Data-adaptive transfer learning for translation: A case study in Haitian and jamaican. In *Proceedings of the Fifth Workshop on Technologies for Machine Translation of Low-Resource Languages (LoResMT 2022)*, pages 35–42, Gyeongju, Republic of Korea. Association for Computational Linguistics.

Nathaniel Romney Robinson, Matthew Dean Stutzman, Stephen D Richardson, and David R Mortensen. 2023. African substrates rather than european lexifiers to augment african-diaspora creole translation. In *4th Workshop on African Natural Language Processing*.

Luisa Seguin. 2020. Transparency and language contact: The case of haitian creole, french, and fongbe. *Journal of Pidgin and Creole Languages*, 35(2):218–252.

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.