Krey-All WMT 2025 CreoleMT System Description: Language Agnostic Strategies for Low-Resource Translation

Ananya Ayasi
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA, USA
ananya.ayasi@gmail.com

Abstract

This work is a submission to the Creole MT Task conducted as a part of the Tenth Conference on Machine Translation (WMT '25) co-located with EMNLP 2025. The focus of the work was on the Seychellois Creole- English language pair by utilizing similar Creoles to make up for the scarcity of data in Creole language systems.

1 Introduction

The work of this paper is based on Robinson et al. (2024) which presented a one of a kind, cumulative dataset for Creole language MT. This is an important topic owing to the fact that many Creole language speakers live in areas where their language is in the minority and they could benefit from better machine translation systems.

The main objective behind this paper is to see if a certain language pair translation could benefit from similar languages when building an MT system. To that end, I picked Seychellois Creole (crs)- English (eng) pair and to supplement the model training Mauritian Creole (mfe), French Guianese Creole (gcr), Louisiana Creole (lou) and Réunion Creole (rcf) were chosen. The linguistic proximity of these Creoles were studied in Papen (1978). Many Creoles also have linguistic relationships with high resource languages which gives better prospects for cross-lingual transfer (Lent et al., 2024) and hopefully can be studied in a future iteration of this paper.

In addition to the shared-task submission for crs–eng, this work evaluates the broader applicability and limits of unified tagging by comparing performance across multiple French-lexifier Creoles as well as Tok Pisin, a typologically distant creole, thereby situating the approach within a linguistically grounded framework rather than a purely engineering exercise.

2 Data

Since this work was part of the constrained track, the data and base model used were provided by Robinson et al. (2024). As mentioned above, the 5 Creole languages used were chosen based on their linguistic proximity.

Seychellois Creole and Mauritian Creole are widely regarded as the closest pair among the varieties considered here, a connection that stems largely from historical migration patterns (Baker et al., 1982). Seychelles was primarily settled from Mauritius in the late 1700s, and while there are lexical and phonological differences, several scholars have noted that they may be treated as regional variants of a single linguistic system rather than entirely separate languages (Michaelis and Rosalie, 2013) (Ramsurrun et al., 2024).

Translation Pair	Train	Val	Test
crs-eng	2,070	107	222
mfe-eng	19,100	472	811
gcr-eng	60	28	38
lou-eng	1,570	92	174
rcf-eng	191	34	28

Table 1: Number of training, validation, and test sentence pairs for each Creole–English translation dataset used in this study.

Réunion Creole, while geographically proximate, has a distinct and earlier colonization history, resulting in greater grammatical divergence from Mauritian. Nevertheless, commonalities in the pronominal system, such as the use of *zot* for both second and third-person plural, and shared plural markers like *ban(de)*, point to the effects of geographical continuity and inter-island contact. Lexical similarities, including borrowings from Malagasy and Indo-Portuguese, reinforce the role of shared areal features despite significant typological differences (Papen, 1978).

Expanding beyond the Indian Ocean, Louisiana Creole and French Guianese Creole also exhibit no-

Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 916–923
November 8–9, 2025.

table structural parallels with Mauritian, underscoring broader transoceanic affinities among French-lexifier Creoles. (Pfänder, 2013) (Klingler and Neumann-Holzschuh, 2013)

For MT research, this interconnected yet varied typological landscape offers opportunities for cross-lingual transfer learning, provided that models are designed to exploit shared structures without conflating distinct grammatical systems.(Grant and Guillemin, 2012)

As far as Creoles go, I have also chosen Tok Pisin as a linguistically dissimilar language to Seychellois Creole to demonstrate how this strategy is effective only when the varieties share substantial structural overlap.

3 Methodology

This work adopts a language-agnostic machine translation strategy for low-resource Creole–English translation, building on multilingual pretraining and fine-tuning techniques that have been shown to benefit typologically related languages in low-data regimes (Johnson et al., 2017); (Aharoni et al., 2019); (Conneau et al., 2020).

I call the approach “language-agnostic” because the model is not given any explicit linguistic rules or handcrafted features about Creole varieties; instead, all source languages are treated under a unified tag, allowing the system to learn transfer purely from shared representations without relying on language-specific annotations.

3.1 Base Model and Setup

The base model is kreyol-mt-pubtrain based on mBART (Namdarzadeh et al., 2023) as distributed in the constrained track by Robinson et al. (2024). This model is pretrained on a variety of languages, including several French-based Creoles, providing a strong initialization for transfer learning. All training experiments use the HuggingFace Transformers library (Wolf et al., 2020) with mixed precision and early stopping where indicated.

3.2 Tagging Strategies

The main focus of this paper was to see if other similar languages can be used to enhance the MT system for a certain language pair or rather language-agnostic machine translation as inspired by (Chen and Zhang, 2024). I mainly experimented with two techniques for the same, "All Kreyols" and "Specialized".

Following Johnson et al. (2017), all input sequences were prepended with a target language tag. In the "All Kreyols" condition which was the language-agnostic approach, all Creole data received the <2crs> tag regardless of source variety, treating all Creoles as dialectal variants for the purpose of model conditioning.

"Partial Kreyall" was basically a variant of "All Kreyoles" where even though all the languages used the same tag, only partial fine-tuning was done. In the "Specialized" condition which was language-specific, each Creole variety retained its own target tag (e.g., <2mfe>, <2lou>), enabling the model to distinguish among them. In this case, I experimented with both full fine-tuning and partial fine-tuning.

3.3 Data Augmentation via Upsampling

It is to be noted that in Table 1 mfe-eng language pairs in the train set was almost twice as that of crs-eng. Given the extreme imbalance in dataset sizes, upsampling of the target language data was employed to avoid underrepresentation of Seychellois Creole. Without intervention, batches are dominated by mfe examples, which can lead to underrepresentation of crs and limiting domain adaptation.

Inspired by Sennrich et al. (2016) and Robinson et al. (2022), initial experiments used a 5x upsampling factor for crs-eng. This ensures that crs examples are seen as frequently as those from the largest dataset, allowing the model to retain multilingual benefits while focusing on the target low-resource language. Later, a 10x factor was tested, especially in the "Specialized" setting, to further balance the distribution. However, it did not exactly lead to better results as will be discussed later.

3.4 Fine-tuning Strategies

Two broad strategies were mainly tested for fine-tuning. The first one being without any freezing, thereby fine-tuning all the parameters of the given kreyol-mt-pubtrain model.

The second one being partial freezing where only the last 4 to 6 encoder layers, all decoder layers, and optionally shared embeddings were frozen. This follows the intuition from Kirkpatrick et al. (2017) and Pfeiffer et al. (2021) that retaining most pretrained parameters can preserve generalization while adapting only the task-relevant parts.

Additional architectural regularization was explored by modifying dropout rates for feedforward, attention, and activation layers. In the notation dropout=0.2/0.05/0.05 in Table 4, the three values correspond respectively to the model's feedforward dropout, attention dropout, and activation dropout. For instance, in kreyol-mt-pubtrain's configuration, these would be set via:

```
model.config.dropout = 0.2
# feedforward layer dropout
model.config.attention_dropout = 0.05
# attention mechanism dropout
model.config.activation_dropout = 0.05
# activation function dropout
```

These rates control the probability of randomly zeroing out elements during training to prevent overfitting. Adjusting them independently allows for targeted regularization. Higher dropout in feed-forward layers can reduce co-adaptation in dense transformations, while lower dropout in attention and activation layers can help preserve sequence modeling capacity without severely impacting convergence.

3.5 Optimization

All models were optimized using AdamW, which decouples weight decay from the gradient-based updates, preventing it from accumulating in the momentum or variance terms. Weight decay values ranged from 0.1 to 0.3, and the learning rate was fixed at 1E-5 to mitigate catastrophic forgetting in the multilingual setting. Warmup schedules of 500–1000 steps were applied, followed by linear, cosine, or constant learning rate decay. In certain configurations, label smoothing of 0.1–0.2 (Szegedy et al., 2015) was incorporated to improve generalization in the low-resource MT setting. All hyperparameters were chosen based on prior work in low-resource MT and small-scale tuning on the validation set.

4 Experiments

The experiments compare baseline, All Kreyols, Partial Krey-all, and Specialized setups. Performance was evaluated using BLEU (Papineni et al., 2002) and chrF (Popović, 2015) on the held-out Seychellois Creole–English test set.

4.1 Language Embeddings

In order to truly understand and corroborate prior linguistic research pertaining to the proximity of the French-lexifier Creoles, kreyol-mt-pubtrain's encoder embeddings for them and Tok Pisin (tpi) were visualized by sampling 60 sentence pairs from each dataset and converting their source sentences into 1024-dimensional vectors using the encoder.

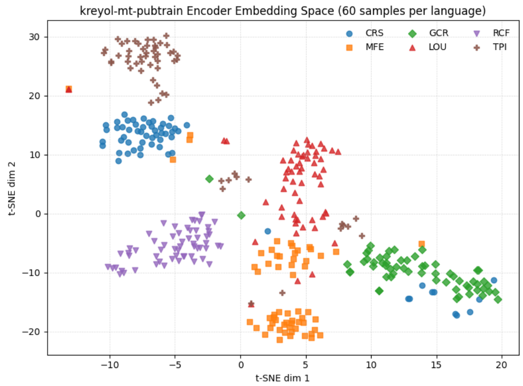

Figure 1: t-SNE visualization of kreyol-mt-pubtrain encoder embeddings for 60 sampled sentences from each language (crs = Seychellois Creole, mfe = Mauritian Creole, gcr = French Guianese Creole, lou = Louisiana Creole, rcf = Réunion Creole, tpi = Tok Pisin). The five French-lexifier Creoles form distinct yet proximate clusters, reflecting their structural similarity, whereas Tok Pisin is clearly separated in embedding space, highlighting its typological distance.

These high-dimensional vectors were then reduced to two dimensions via t-SNE (Cai and Ma, 2022), which preserves local similarity structure while making patterns viewable in a scatter plot. The resulting "t-SNE dim 1" and "t-SNE dim 2" axes have no direct linguistic meaning. Their role is to position points so that those close in the plot were also close in the original embedding space. Each language was assigned a unique color and marker, revealing how kreyol-mt-pubtrain clusters them based on learned representations.

The plot shows that the five French-lexifier Creoles form distinct but proximate clusters, with some overlap among related varieties such as gcr, mfe and lou. crs is positioned near rcf and partially overlaps with gcr, aligning with the shared French-derived vocabulary and grammar across these languages. By contrast, tpi is kept well apart from these clusters, reflecting its typological distance from the French-lexifier group. This separation reinforces the observation that unified tagging works best for closely related languages, where embeddings occupy overlapping regions and facilitate cross-lingual transfer (Ponti et al., 2021).

4.2 Baseline

The baseline fine-tuned the given kreyol-mt-pubtrain with no freezing and <2crs> tag for all data. The best run achieved BLEU = 34.61, chrF

= 60.84 after 15 epochs with a linear scheduler and 500 warmup steps. To ensure comparability, this baseline was reproduced under the same constrained-track conditions, and the scores are in line with those reported in the shared-task paper, confirming that my setup is consistent with the official baseline.

4.3 All Kreyols/ Krey-All

Here, all Creole varieties were merged under the <2crs> tag. The best configuration reached BLEU = 35.44, chrF = 61.71 using 20 epochs (stopping at epoch 18), cosine scheduling, LR = 1E-5, warmup = 1000, and label smoothing = 0.1. This shows gains over the baseline, suggesting that cross-variety pooling helps despite grammatical differences.

4.4 Partial Krey-all

For this setting, partial freezing was applied, testing different numbers of unfrozen encoder layers, decoder layers, and embedding sharing. The results were mixed with some configurations having degraded performance (e.g., BLEU = 20.14 for unfreezing last 2 encoder/decoder layers), while others recovered to near-baseline levels. The gains were limited, indicating that aggressive layer freezing may not suit closely related low-resource varieties.

4.5 Specialized

This setup retained distinct tags per variety, with and without crs upsampling. The highest performance here (BLEU = 35.19, chrF = 61.14) was achieved by disabling early stopping, allowing the model to converge fully over 20 epochs. Notably, 10x crs upsampling did not yield improvements over the baseline or the settings with 5x upsampling. The findings suggest a saturation effect, where oversampling beyond 5x no longer improves performance because the model's representation of crs is already well-established.

4.6 Other Language Pairs

To further validate the generality of the Krey-All unified tagging strategy, I extended experiments beyond the official shared-task pair (crs–eng) to additional Creole–English directions. For the French-lexifier creoles (crs, mfe, rcf, lou, gcr), I trained each pair by combining its own data with data from the other French-lexifier creoles, while preserving their respective language tags. This ensured that the unified strategy was tested symmetrically across all related varieties. The results, shown in Table 3, indicate consistent improvements across all of these creoles, with gains ranging from +0.9 to +1.3 BLEU.

To further stress-test the approach, I then applied the same unified tagging setup to Tok Pisin (tpi) by pooling it together with all the French-lexifier creoles. Unlike the others, tpi is typologically distant, and the results confirm that the strategy does not generalize well in this case (–1.28 BLEU). Importantly, to maintain the sanctity of the experiments, tpi data was not used when training the French-lexifier systems. This separation highlights the core finding: unified tagging provides consistent benefits when applied to closely related creoles, but degrades performance when applied to a structurally different language.

All experiments were conducted on an NVIDIA A100 GPU. Also note that for the shared task submission, crs–eng Krey-All was submitted as the primary system, while crs–eng Partial Krey-All and crs–eng Specialized Krey-All were submitted as contrastive systems 1 and 2, respectively.

5 Results and Analysis

The language-agnostic tagging strategy which was the main goal of this paper proved most effective, as shown by two clear outcomes. First, tagging crs across all Creoles consistently yielded higher scores compared to runs where tags were language-specific. Second, even when crs data was oversampled by 10x, scores did not improve over the 5x runs, despite mfe data being more abundant, indicating that mfe served as a strong substitute for crs in the shared tag setup. This aligns with prior findings that shared tags encourage parameter sharing across related low-resource languages (Johnson et al., 2017); (Sachan and Neubig, 2018).

No freezing continued to outperform partial freezing, suggesting that kreyol-mt-pubtrain's multilingual pretraining already aligns Creole varieties in the embedding space. Scheduler and label smoothing choices impacted results more than weight decay, with cosine or linear schedules and smoothing = 0.1 performing best overall.

6 Conclusion

This study examined a unified tagging strategy using kreyol-mt-pubtrain to improve translation quality for the low-resource Seychellois Cre-

Table 2: Best results in each category for Creole–English MT.

Technique	Details	BLEU	chrF
Baseline	No freezing	34.61	60.84
Partial Krey-all	Unfreeze last 4 encoder / all decoder layers+ shared embeddings, dropout=0.2/0.05/0.05	34.72	61.11
Specialized	No freezing and no early stopping	35.19	61.14
All Kreyols	No freezing	**35.44**	**61.71**

Creole–Eng	Baseline	Krey-All
crs–eng	34.61	35.44
rcf–eng	41.00	42.12
lou–eng	32.72	33.65
gcr–eng	49.89	51.21
mfe–eng	25.23	26.29
tpi–eng	53.25	51.97

Table 3: **BLEU scores** for baseline vs. Krey-All unified tagging across Creole–English pairs. French-lexifier creoles consistently improve under unified tagging, while Tok Pisin (tpi), a typologically distant language, shows degraded performance.

ole–English pair. By tagging crs across all Creole varieties, the model achieved higher scores than language-specific tagging. The absence of gains from 10x crs upsampling, despite the availability of larger mfe data, further indicated that Mauritian Creole served as an effective proxy for crs in the shared-tag setup. These findings reinforce that related Creole languages can transfer knowledge efficiently when trained under a unified representation.

While absolute gains are modest (<1 BLEU/chrF), they are consistent across runs and show that leveraging related Creoles can stabilize low-resource training. Moreover, since the constrained track provides the same pretraining data for both base and fine-tuning, the improvements are naturally limited compared to unconstrained settings.

Beyond crs–eng, applying the same strategy to other Creole–English directions (rcf, lou, gcr,mfe) produced similar improvements, confirming that the benefits of unified tagging generalize across French-lexifier Creoles. In contrast, Tok Pisin (tpi), a typologically distant creole, did not benefit, which corroborates the claim that this approach is most effective when applied to closely related languages.

No-freezing fine-tuning proved most effective, suggesting that kreyol-mt-pubtrain's multilingual pretraining already aligns Creole varieties well in embedding space. The best system, trained on all Creoles without freezing, reached a BLEU of 35.44 and chrF of 61.71 for crs-eng, outperforming the baseline. Overall, cross-lingual transfer with linguistically related low-resource languages emerges as a promising strategy in constrained MT settings, particularly when paired with balanced exposure and a unified tagging scheme.

If there is one takeaway from this paper, it is that prior linguistic research remains vital even in the age of large language models. More often than not, leveraging linguistic insight can be just as critical as scaling data volume. (Gu, 2025)

Limitations

While the results are promising, there are several limitations to this work:

- Data imbalance: The extreme disparity in dataset sizes (e.g., 2K pairs for crs-eng vs. 19K for mfe-eng) may cause the model to underfit the smallest dataset even after upsampling. It is also to be noted that in the original paper, the scores for mfe-eng were much lower than crs-eng despite the larger training set. This should be further investigated in upcoming iterations.

- Domain mismatch: Training corpora for different Creoles may differ in domain, register, or orthography, potentially introducing noise (Karakasidis et al., 2023).

- Model constraints: We were limited to the kreyol-mt-pubtrain model in the constrained track, restricting architectural or pre-training modifications.

- Evaluation scope: BLEU and chrF scores, while informative, do not capture deeper semantic adequacy or fluency. No human evaluation was performed.

Ethics Statement

This work uses only publicly released datasets from Robinson et al. (2024) and adheres to the constraints of the WMT 2025 Creole MT Task. No personally identifiable information (PII) is present in the training or evaluation data.

While the goal is to improve accessibility for speakers of low-resource Creole languages, I acknowledge that MT systems may propagate biases present in the source data, especially given the historical and sociolinguistic contexts of Creole-speaking communities. Deployment of such systems should therefore be accompanied by careful evaluation in real-world contexts, with community feedback guiding improvements.

Additionally, these models are intended to augment human communication rather than replace professional translators, especially in sensitive domains such as legal, medical, or governmental communication.

Acknowledgements

The author thanks Nate (author of Robinson et al. (2024)) for his invaluable guidance and mentorship throughout this work. Gratitude is also extended to family (Seema, Ajai, Anwaya) and friends (Ishwrya, Balaji, Maanasa, Sruthi) for their unwavering support during the course of this research. This work is dedicated to Simi, Eldin, Ewan, and Ryan.

References

Roee Aharoni, Melvin Johnson, and Orhan Firat. 2019. Massively multilingual neural machine translation. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 3874–3884, Minneapolis, Minnesota. Association for Computational Linguistics.

Baker, Philip; Corne, and Chris. 1982. *Isle de France Creole: Affinities and origins*. Karoma, Ann Arbor, MI.

T. Tony Cai and Rong Ma. 2022. Theoretical foundations of t-sne for visualizing high-dimensional clustered data.

Xiao Chen and Chirui Zhang. 2024. Language-agnostic zero-shot machine translation with language-specific modeling. In *2024 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451, Online. Association for Computational Linguistics.

Anthony Grant and Diana Guillemin. 2012. The complex of creole typological features: The case of mauritian creole. *Journal of Pidgin and Creole Languages*, 27(1):48–104.

Wenshi Gu. 2025. Linguistically informed chatgpt prompts to enhance japanese-chinese machine translation: A case study on attributive clauses. *PloS one*, 20(1):e0313264.

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.

Georgios Karakasidis, Nathaniel Robinson, Yaroslav Getman, Atieno Ogayo, Ragheb Al-Ghezi, Ananya Ayasi, Shinji Watanabe, David R Mortensen, and Mikko Kurimo. 2023. Multilingual tts accent impressions for accented asr. In *International Conference on Text, Speech, and Dialogue*, pages 317–327. Springer.

James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A. Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, Demis Hassabis, Claudia Clopath, Dharshan Kumaran, and Raia Hadsell. 2017. Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences*, 114(13):3521–3526.

Thomas A. Klingler and Ingrid Neumann-Holzschuh. 2013. Louisiana creole. In Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath, and Magnus Huber, editors, *The survey of pidgin and creole languages. In "The survey of pidgin and creole languages". Volume 2: Portuguese-based, Spanish-based, and French-based Languages*. Oxford University Press, Oxford.

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.

Susanne Maria Michaelis and Marcel Rosalie. 2013. Seychelles creole. In Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath, and Magnus

Huber, editors, *The survey of pidgin and creole languages. In "The survey of pidgin and creole languages". Volume 2: Portuguese-based, Spanish-based, and French-based Languages*. Oxford University Press, Oxford.

Behnoosh Namdarzadeh, Sadaf Mohseni, Lichao Zhu, Guillaume Wisniewski, and Nicolas Ballier. 2023. Fine-tuning MBART-50 with French and Farsi data to improve the translation of Farsi dislocations into English and French. In *Proceedings of Machine Translation Summit XIX, Vol. 2: Users Track*, pages 152–161, Macau SAR, China. Asia-Pacific Association for Machine Translation.

Robert Papen. 1978. *The French-based Creoles of the Indian Ocean: An analysis and comparison*. University of California, San Diego. Dissertation.

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.

Jonas Pfeiffer, Aishwarya Kamath, Andreas Rücklé, Kyunghyun Cho, and Iryna Gurevych. 2021. Adapterfusion: Non-destructive task composition for transfer learning.

Stefan Pfänder. 2013. Guyanais. In Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath, and Magnus Huber, editors, *The survey of pidgin and creole languages. In "The survey of pidgin and creole languages". Volume 2: Portuguese-based, Spanish-based, and French-based Languages*. Oxford University Press, Oxford.

Edoardo Maria Ponti, Julia Kreutzer, Ivan Vulić, and Siva Reddy. 2021. Modelling latent translations for cross-lingual transfer.

Maja Popović. 2015. chrF: character n-gram F-score for automatic MT evaluation. In *Proceedings of the Tenth Workshop on Statistical Machine Translation*, pages 392–395, Lisbon, Portugal. Association for Computational Linguistics.

Neha Ramsurrun, Rolando Coto-Solano, and Michael Gonzalez. 2024. Parsing for mauritian creole using universal dependencies. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 12622–12632.

Nathaniel Robinson, Perez Ogayo, Swetha Gangu, David R. Mortensen, and Shinji Watanabe. 2022. When is tts augmentation through a pivot language useful?

Nathaniel R. Robinson, Raj Dabre, Ammon Shurtz, Rasul Dent, Onenamiyi Onesi, Claire Bizon Monroc, Loïc Grobol, Hasan Muhammad, Ashi Garg, Naome A. Etori, Vijay Murari Tiyyala, Olanrewaju Samuel, Matthew Dean Stutzman, Bismarck Bamfo Odoom, Sanjeev Khudanpur, Stephen D. Richardson, and Kenton Murray. 2024. Kreyòl-mt: Building mt for latin american, caribbean and colonial african creole languages.

Devendra Sachan and Graham Neubig. 2018. Parameter sharing methods for multilingual self-attentional translation models. In *Proceedings of the Third Conference on Machine Translation: Research Papers*, pages 261–271, Brussels, Belgium. 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.

Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and Zbigniew Wojna. 2015. Rethinking the inception architecture for computer vision.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi 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 M. Rush. 2020. Huggingface's transformers: State-of-the-art natural language processing.

A Additional Results

Table 4: Full experimental results (BLEU/chrF on crs-ENG).

Technique	Details	Epochs	Stopped at	LR Sched.	Warmup	BLEU	chrF
Baseline	No freezing	15	15	linear	500	**34.61**	**60.84**
All Kreyols	No freezing	5	5	linear	750	34.88	61.26
All Kreyols	No freezing	20	11	linear	500	34.43	61.18
All Kreyols	No freezing	20	18	cosine	1000	**35.44**	**61.71**
All Kreyols	No freezing	30	14	linear	750	35.06	61.56
All Kreyols	No freezing	25	10	cosine	1000	35.25	61.49
All Kreyols	No freezing	20	13	constant	1000	35.11	61.45
All Kreyols	No freezing	30	4	cosine	1000	34.70	60.95
Partial Krey-all	Unfreeze last 2 encoder/decoder layers	5	4	cosine	1000	20.14	48.55
Partial Krey-all	Unfreeze last 2 encoder/decoder layers + shared embeddings	5	5	cosine	1000	33.90	60.24
Partial Krey-all	Unfreeze last 4 encoder / all decoder layers + shared embeddings	20	7	cosine	1000	33.91	60.71
Partial Krey-all	Unfreeze last 4 encoder / all decoder layers + shared embeddings; dropout=0.3/0.1/0.1	20	7	cosine	1000	33.91	60.71
Partial Krey-all	Unfreeze last 4 encoder / all decoder layers + shared embeddings, dropout=0.2/0.05/0.05	20	15	cosine	1000	**34.72**	61.11
Partial Krey-all	Unfreeze last 6 encoder / all decoder layers + shared embeddings, no dropout	20	10	cosine	1000	34.20	**61.15**
Specialized	Unfreeze last 4 encoder / all decoder layers + shared embeddings, dropout=0.2/0.05/0.05	20	7	cosine	1000	34.51	60.27
Specialized	Unfreeze last 6 encoder / all decoder layers + shared embeddings, no dropout	20	13	cosine	1000	34.09	60.26
Specialized	No freezing	20	5	cosine	1000	34.81	60.93
Specialized	No freezing + crs upsampled 10×	5	5	cosine	1000	34.66	60.56
Specialized	No freezing and no early stopping	20	20	cosine	1000	**35.19**	**61.14**