Findings of the Second Shared Task for Creole Language Machine Translation, at WMT 2026

Nathaniel R. Robinson[1] **Ruth-Ann Armstrong***[,3] **Claire Bizon Monroc***[,2]
Rasul Dent*[,2] **Pranav Gupta***[,4] **Raj Dabre**[5] **Andre Coy**[6] **Kenton Murray**[7]
[1]Johns Hopkins University, USA; [2]Inria Paris, France;
[3]University of Southern California, USA; [4]Cisco, USA; [5]IIT Madras, India
[6]University of the West Indies at Mona, Jamaica; [7]George Mason University, USA
nrobin38@jhu.edu
*Listed alphabetically (equal contribution)

Abstract

We present the second annual Shared Task for Creole Language Machine Translation, the only published shared task specializing in machine translation (MT) methods for Creole languages and their communities, presented at the Conference of Machine Translation (WMT) 2026. We accepted system submissions for two tasks: MT and language identification (LID). We received submissions from eight teams: 13 MT systems from six teams across 88 language directions; and four LID systems from two teams, identifying 11 languages each. Our analysis of MT systems indicates fine-tuning with newly curated datasets and/or synthetic data as an effective strategy for producing state-of-the-art MT. For the LID task, we found that smaller, resource-efficient systems achieved state-of-the-art scores on some languages, especially when paired with data augmentation.

1 Introduction and Background

Creole languages have been historically underserved by existing technologies; hence developing better Creole language technologies has been an active area of research for some years. For individual languages, researchers have explored innovations for grammatical parsing (Wang et al., 2017; Caron et al., 2019; Mompelat et al., 2022), sentiment classification (Oyewusi et al., 2020; Muhammad et al., 2022, 2023), named entity recognition (Adelani et al., 2021), natural language inference (Armstrong et al., 2022), and speech recognition (N. Havard et al., 2025; Le Ferrand and Henri, 2026; Andrade et al., 2026). For machine translation specifically, past works have focused on Haitian (Frederking et al., 1997; Callison-Burch et al., 2011), Jamaican Patois (Robinson et al., 2022, 2023), Nigerian Pidgin / West African Pidgin (Ogueji and Ahia, 2019; Adelani et al., 2022), Mauritian Creole (Dabre and Sukhoo, 2022; Suddul et al., 2026; Boodeea et al., 2025), Sranan Tongo (Zwennicker and Stap, 2022), Singlish (Liu et al., 2022), Cape Verdean Creole (Medina, 2025; Medina et al., 2025), Guinea-Bissau Creole (Rowe et al., 2025b), Guyanese Creolese (Clarke et al., 2024), and Chavacano (Vicente et al., 2026)

Improved translation for Creole languages has led to advances in medical translation (Mompelat et al., 2026; Mompelat, 2025a), natural disaster mitigation (Robinson, 2025b), and other humanitarian aid (Lewis, 2010).

Other works have focused broadly on developing technology for many Creole languages together. Lent et al. (2022b) discussed survey findings on the needs for Creole language technologies. Lent et al. (2022a); Fekete et al. (2025) explored knowledge transfer between Creole and related languages. Lent et al. (2024) developed the first multilingual MT model specifically for Creole languages. Robinson et al. (2024) released the first public MT dataset for a large number of Creole languages and built state-of-the-art (SOTA) models for Caribbean and African Creole languages.

These efforts were steps forward for Creole language processing. However their broad scope precluded the possibility of focusing on specific languages for specific communities. This inspired the organization of the First Shared Task for Creole Language Translation at WMT 2025 (Robinson et al., 2025). This task was framed as an open challenge, to invite Creole language community members and enthusiasts to engage in the development of technologies for their own languages.

The 2025 shared task (Robinson et al., 2025) received submissions from five teams, including four submitting MT systems (Ayasi, 2025; Mompelat, 2025b; Rowe et al., 2025a; Rajcoomar, 2025) and two submitting new MT datasets (Robinson, 2025a; Rajcoomar, 2025). Rajcoomar (2025) improved upon SOTA MT in some domains for Mauritian Creole by fine-tuning a large language model (LLM) and using newly curated data. Robinson

Proceedings of the Eleventh Conference on Machine Translation, pages 1138–1158
October 28-29, 2026 ©2026 Association for Computational Linguistics

(2025a) likewise improved SOTA in some domains for MT of French Guianese Creole and Belizean Kriol, also by using new datasets. Rowe et al. (2025a) most convincingly improved over SOTA, for seven lusophone Creole languages, also by incorporating new data, but they unfortunately could not release the data.

2 Task Overview

We organized the shared task into four subtasks:

- Subtask 1A: Models for MT of Creole languages
- Subtask 1B: Models for LID of Creole languages
- Subtask 2A: Creole MT data
- Subtask 2B: Creole speech data

We only received submissions for subtasks 1A and 1B, so we focus on those in this report.

2.1 MT modeling subtask

For Subtask 1A we solicited systems to translate between any number of Creole languages and English or French. We also permitted any culturally relevant source or target languages with approval from us (such as Spanish for Haitian, which we approved for one of the participating teams).

The purpose of this subtask was to produce SOTA Creole MT models. Hence, participants were permitted to use any data or pretrained models, *except* for those expressly prohibited. The prohibited models were the public Hugging Face models `kreyol-mt`, `kreyol-mt-scratch`, and `CreoleM2M`, excluded because of training contamination with the eval sets we used. The prohibited data were the public eval sets we used, namely the test split of the public Hugging Face `kreyol-mt` dataset.

We provided the following resources for participants:

- Two baseline models
  - The `kreyol-mt-pubtrain` model available on HuggingFace
  - A newly trained baseline model based on `gemma-4-E2B` (Team, 2026)
- An official train set: the train and dev splits of the public `Kreyòl-MT` dataset[1]

[1] https://huggingface.co/datasets/jhu-clsp/kreyol-mt

2.1.1 New MT baseline model

We trained a new baseline MT model for this year's shared task. We had two reasons for this: (1) because we had to prohibit use of the original `Kreyòl-MT` flagship model on HuggingFace and (2) in order to incorporate the data submissions from last year's shared task (Robinson et al., 2025; Rajcoomar, 2025; Robinson, 2025a).

We trained a baseline model, `gemma4-creole-mt` (Gupta, 2026), by fine-tuning `gemma-4-E2B` via LoRA (Hu et al., 2022) on 50k total segments sampled from the `Kreyòl-MT` dataset (Robinson et al., 2024), and datasets submitted to last year's iteration of the shared task (Robinson et al., 2025; Rajcoomar, 2025; Robinson, 2025a). We trained on a node consisting of eight L40S Nvidia GPUs, comprising 86 language pairs spanning 50 Creole and related languages. We used ALMA-style prompts (Xu et al., 2024) with full language names, for example:

```
Translate this from Ndyuka into English:
{source sentence}
### Translation:
{target sentence}
```

We used a completion-only loss, where prompt tokens were masked out and the loss was computed only on the translation. We utilized temperature sampling following NLLB (NLLB Team et al., 2022) with $T = 5$, in order to balance high-resource pairs (e.g. `hat--eng`, 833k sentences) against lower-resourced pairs (e.g. `fng--eng`, 33 sentences). We used the following hyperparameters for LoRA: rank 64, α = 128, while targeting all linear layers. In Table 6, we report the metrics on a development set.

2.2 LID subtask

For the LID subtask, we initially proposed to consider LID for English-, French, Portuguese-, and Spanish-lexifier Creoles and Pidgins, in consideration of recent scholarship addressing these groups (Adebara et al., 2022; Vicente and Cheng, 2024; Dent et al., 2025). This was later narrowed down to French-lexifier Creoles in response to participant interest. More specifically, participants were asked to produce models capable of distinguishing between 9 Creole languages (Guadeloupean, French Guianese, Haitian, Martinican, Louisiana, Mauritian, Reunionese, Saint Lucian, and Seychellois), French, English, and an `oth` category for other languages. Among these, certain languages

have been particularly challenging for LID systems, due to either linguistic proximity (Guadeloupean, Martinican, Saint Lucian), limited amounts of contemporary data (Louisiana, French Guianese), or both.

Similarly, we initially proposed exploring the inclusion of various kinds of context metadata alongside the texts (Dent et al., 2026). However, due to several constraints, we were unable to include such metadata in this version of the shared task. Nevertheless, we did strive to include test data from a wide variety of sources. These include scholastic materials, online newspapers, and a corpus of previously anonymized French SMS messages (Panckhurst et al., 2020). The latter source allowed us to consider the effects of nonstandard transcription of the lexifier, which has been underexplored thus far.

For our baseline model, we chose GlotLID (Kargaran et al., 2023) because it is considered SOTA for coverage of long-tail languages (Suarez et al., 2026). This model covers languages in the test set, with two exceptions: it does not include Louisiana Creole or Martinican Creole.

3 Task Submissions

We received submissions from eight teams. We overview the submissions from six teams participating in subtask 1A in §3.1, and submissions from the two teams participating in subtask 1B in §3.2

3.1 MT systems

We received submissions from six teams for the MT systems task. All participants' submissions together covered 45 language pairs and 89 translation directions. Most teams submitted results on fewer than three language pairs with a specific focus: Haitian for CHV, UQAM and PIK; Pacific anglophone Creoles for UniMelb; and Réunion Creole for KYX. The remaining team, GMU, submitted systems for 80 different translation directions (translation into and out of 40 Creole languages).

Teams mostly focused on language pairs involving English, with only five submissions translating from or to French or Spanish. Submissions were however spread across language pairs, and only six directions had submissions from more than one team. Haitian was by far the most represented Creole language: it was present in four submissions.

The participating teams explored a variety of different modeling approaches. The KYX team (Öztop, 2026) used QLoRA adaptation and focused much of their paper on the pitfalls of MT evaluation for Creole languages. UniMelb (Merx et al., 2026) fine-tuned NLLB (NLLB Team et al., 2022) with a post-training schedule including back-translated and LLM-synthesized data. GMU (Meng and Anastasopoulos, 2026) built novel datasets by applying optical character recognition (OCR) to an existing database and compared fine-tuning and few-shot model adaptation. CHV (Jourdain et al., 2026) employed few-shot prompting only, with lexical and grammatical knowledge extracted and included in prompts. UQAM (Le et al., 2026) developed a dataset of 17k new, culturally relevant bitext pairs, and employed code-switching data augmentation.

We describe in more detail the submitted MT systems in the following section.

3.1.1 KYX

The KYBELIX (KYX) team submitted a system for Réunion Creole to English Translation (`rcf` → `eng`) and focused on common pitfalls that may arise when evaluating extremely resource-impoverished languages. In particular, they focused on two phenomena that can significantly impact results. First, they demonstrated how small train and development data overlap with inexact matches, which can significantly impact results on very small test sets. (The rcf-eng public test set is only 34 sentences long.) Second, they argued that character-based metrics can reward systems for mimicking spelling conventions, as opposed to actual translations when no standard orthographies exist for a language. To mitigate this, they contributed an evaluation method that is contamination- and orthography-aware. By doing so, they demonstrated that their QLoRA-based system does not have as large chrF++ (Popović, 2017) gains as initially appeared. Their conclusions are not specific merely to Réunion Creole, but are used to demonstrate broader phenomena.

KYX only submitted a `primary` system.

3.1.2 UniMelb

To tackle machine translation for low-resource Pacific Creole languages, namely, Tok Pisin, Bislama, and Solomon Pijin, the team from the University of Melbourn (UniMelb) developed a joint translation models for six directions using NLLB-3.3B. They developed two models: one for out-of-Creole MT and another for into-Creole MT. Their core method begins with joint pre-training on a large corpus dominated by religious text, followed by continued

fine-tuning on a domain-balanced mix incorporating back-translated news, spoken transcripts, and synthetic colloquial data distilled from Gemini 3.5 Flash. Key data preparation techniques include LLM-guided loanword respelling and normalization for Solomonese Pijin (`pis`), and LLM-based parallel alignment filtering to enhance training.

The resulting models outperform open multilingual baselines by over 3 chrF++ points across all directions on human-original test sets, closely matching or beating Gemini 3.5 Flash on broad-domain Bouquet and spoken transcript benchmarks. Joint multilingual training notably benefits the lowest-resource Creoles without degrading Tok Pisin performance, while domain additions transfer well across related languages. Furthermore, the study demonstrates that synthetic distillation provides rapid early quality gains that plateau after a few thousand sentence pairs, highlighting an efficient strategy for low-resource domain expansion.

The UniMelb team only submitted a `primary` system for each translation direction.

3.1.3 GMU

The George Mason University (GMU) team considered 80 translation directions overall, namely English ↔ 40 Creole languages. They cleaned the provided training data by removing pairs with sentences comprising fewer than 50% alphabetic characters, pairs with either sentence containing fewer than three words or more than 60 words, pairs with source-to-target or target-to-source character length ratio exceeding 2.2, Levenshtein ratio exceeding 0.8, and pairs with either side identified as the wrong language by the NLLB language identification model (NLLB Team et al., 2022).

They extracted additional novel bitext data using OCR and a vision language modeling (VLM). Specifically, they applied OCR-capable VLM `Qwen3.5-27B-FP8` (Qwen Team, 2026a) to the DReAM corpus (Virk et al., 2020) to produce additional data for the 28 lowest-resourced languages.

For MT modeling, they experimented with parameter-efficient finetuning of the 4B-parameter Gemma 3 model's pretrained checkpoint (Team et al., 2025) and the NLLB 1.3B-parameter model (NLLB Team et al., 2022), along with few-shot prompting of `Qwen3.6-27B` (Qwen Team, 2026b), using 4-bit quantization on a single 40 GB GPU. For model finetuning, since Haitian had substantially more training data compared to the other Creole languages, they trained a separate model for Haitian and a unified model for the remaining Creole languages using temperature sampling.

They observed that incorporating additional parallel data often improved translation performance. They also observed that, depending on the amount of data in a language, few-shot prompting could be better or worse compared to finetuning.

The team's `contrastive1` system was the 4B-paramater Gemma 3 model fine-tuned the cleaned provided train data. Their `primary` system was the same, but with the novel data incorporated. Their `contrastive2` employed the novel training data, but in a 512-shot in-context prompting paradigm with `Qwen3.6-27B`.

3.1.4 CHV

The CHV team considered four translation directions for Haitian (`hat`): English ↔ Haitian and French ↔ Haitian. Noting that existing training-based methods are limited by data scarcity, they postulated that fine-tuning on synthetic data may not be enough to widen domain coverage or vocabulary for low-resource languages. Instead, they proposed to switch compute allocation from training to test time, and relied on LLM agents' emergent translation capability: an agent can be provided additional knowledge about the translation task during inference. The authors addressed three main questions: (1) how external resources should be retrieved to improve translations, (2) what types of models they can improve, (3) and whether this inference-time compute provides enough benefit to justify its cost.

For this purpose, they augmented the open-weight 8B-parameter `Qwen3-8B` model with a knowledge injection mechanism. Three different knowledge retrieval tools were evaluated: (1) a lexicon tool that extracts glosses from a translation dictionary, (2) a memory for few-shot learning, that contains both translation examples and monolingual sentences extracted from the `Kreyòl-MT` training dataset, and (3) a grammar rules selection tool. This required tuning new hyper-parameters: i.e. the number of examples to retrieve or the similarity threshold above which an example is deemed relevant. All variants could then be combined with two injection methods: the *forced push*, which inserts all possible extracted information in the translation prompt, and the *agentic*, which lets the agent reason about the task and query each tool when needed, following the ReAct framework (Yao et al., 2023). All systems were then evaluated on the FLORES+

devtest split (He et al., 2024), and result significance was validated with a paired bootstrap. The best performing knowledge injection mechanism was then used for submission.

The *forced push* approach systematically scored better than all evaluated *agentic* methods, which tended to fall into tool-calling loops, used 6× more tokens, and yet were often unable to correctly exploit the injected knowledge for Creole language generation. Increases in scores was instead mostly provided by simple, computationally cheaper tools: for eng → hat for example, increasing the number of glosses in the prompt increased the chrF++ score by +3.83, adding parallel and monolingual data examples by +1.14, and translating French sentences into English before translating into Haitian increased +4.56. Activating reasoning on the base model with *forced push* was useful for hat → fra, on which it counters the base model's tendency to copy Haitian words into French, and remained less costly (1.7× tokens) than agentic approaches. Hence the three final submitted systems all used *forced push*: a primary submission with reasoning activated, a constrative1 submission using *forced push* without reasoning, and the base Qwen3-8B model as a contrastive2 submission.

The team compared their proposed methods with the primary baseline gemma4-creole-mt we provided, along with the base Qwen3-8B model they adapted in their submissions. They found overall that knowledge injection improved the scores over the base models for all directions, by up to 14.2 chrF++ points for eng → hat. Knowledge injection appeared to help most for into-Creole translation, where it partially corrected for the lack of Creole fluency exhibited by language models. However, the method still scored below both our secondary baseline, the Kreyòl-MT model trained on public data (Robinson et al., 2024), and the frontier LLM GPT-4.1-mini.

Lastly, the CHV team conducted preliminary tests applying their prompting method with various LLM engines from five different families, obtaining chrF++ point estimates on a single inference run on a 200-sentence subset of FLORES+. They found that gains were consistently stronger for weaker models, and diminished as base model performance improved until they were negligible for frontier models. Hyper-parameters for knowledge injection were only tuned for the weaker Qwen3-8B model; yet analyzing score distributions showed that improvements were driven by correcting the worst translations. This might explain why strong frontier models did not benefit from knowledge injection. The authors concluded that injected knowledge can improve vocabulary or coverage for weaker base models, but that it might not improve fluency beyond a certain point, where training-based methods may still be needed.

3.1.5 UQAM

The submission from Université de Québec à Montréal for Haiti (UQAM) team addresses the challenge of creating a culturally aware MT system for translation from Haitian to English or French. They attempted to accomplish this through a combination of novel cultural dataset curation, code-switching data augmentation, and multilingual model adaptation. The work is distinct for its emphasis on cultural representation, incorporating linguistic material that reflects Haitian cultural knowledge, oral traditions, and idiomatic language.

The primary contribution of the paper is the construction of a parallel corpus of culturally-relevant Haitian material, including poems, proverbs, and culturally-grounded expressions from the *Atlas Linguistique d'Haïti* (Fattier, 1998). Approximately 55 hours of spoken Haitian were transformed into a parallel translation dataset, through a pipeline involving audio segmentation, automatic transcription, machine translation, filtering, and human validation. This produced more than 17,000 additional sentence pairs, which were used to augment the Kreyòl-MT resources. This represents a substantial effort to leverage underutilized oral language resources and expand cultural diversity of available Haitian MT data.

To improve robustness to multilingual use, the authors introduce a code-switching augmentation strategy based on the use of LLMs. Using a token-level replacement approach, an LLM was prompted to replace between one and three Haitian tokens with their English or French equivalents while preserving sentence fluency and semantic coherence. These synthetic code-switching data were concatenated with the original training corpus, augmenting the total dataset available for training. The team proposed that, in addition to augmenting data, this approach increase the model's robustness to bilingual input, through exposure to multilingual training data.

The models employed were: a baseline Transformer system trained from scratch (E1); an adapted version of the NLLB-200 multilingual

model fine-tuned on the combination of `Kreyòl-MT` and code-switching data (E2), and a version of the NLLB-200 fine-tuned with `Kreyòl-MT`, code-switching and culturally-relevant data (E3). This provides an informative comparison between a conventional NMT system and a pre-trained multilingual model, assessing the contribution of cultural data across the different modeling paradigms.

Beyond standard translation quality measures, the team evaluated cultural preservation using the *CSI-Match* metric (Yao et al., 2024). The team's evaluation choice aligns well with the objectives of their paper: to assess whether cultural and idiomatic meaning is maintained during translation.

On their internal tests, the team achieved improvements over their internal baseline in both translation directions. For Haitian to English, E2 delivered substantial gains across all major metrics, increasing BLEU from 43.6 to 53.6 and chrF++ from 65.0 to 72.7. The most striking improvement was observed in cultural translation quality, with the CSI-Match score rising from 50.0 to 77.0, indicating a much stronger ability to preserve culturally grounded meanings and idiomatic content. E3 had similar performance, with only a slight decrease across evaluation metrics. For Haitian to French, both submitted systems also consistently outperformed the baseline. (See Le et al. (2026).) Cultural awareness gains were more modest than in the English direction, but remained positive.

Both adapted models (E2 and E3) showed the value of incorporating culturally enriched data and code-switching data augmentation. The team designated their E2 system as `primary` and two hyperparameter configurations of E3 as `contrastive1` and `contrastive2` for their submission. Across most metrics in internal tests, E3 performed similarly to, but slightly below, E2, indicating that the additional experimental modifications did not translate into measurable improvements. Given that the unadapted NLLB-200 model was not tested, it is difficult to say if the majority of the performance comes from using the multilingual model or from including new data.

The proposed approach delivered strong gains, but the authors pointed out significant limitations. The code-switching data may not capture the natural sociolinguistic behavior of bilingual speakers, and the use of automatic translation in processing recordings may introduce residual biases. The team also lamented limited digital resources, insufficient representation of dialectal variation, and a lack of community-driven data development.

3.1.6 PIK

Unfortunately the PIK team did not submit a system report, so we cannot report on their methods or approach.

3.2 LID systems

We overview submissions from the two teams participating in the LID systems subtask. The HT team (Ayasi, 2026) focused on organizing language labels into groups and running a hierarchy-based LID system that classified the language group as well as the language itself. The RP team (Strickland, 2026) focused on curating new data in addition to the data provided in order to mitigate class imbalances for French Guianese Creole and Réunion Creole.

3.2.1 HT

The Hytton Technologies (HT) team presents a solution based on a multiclass linear support vector machine (SVM) that uses term-frequency inverse-document-frequency (TF-IDF) weighted character n-grams representations as features. This is presented as a noise-robust solution, and thus appropriate for Creole LID. These representations are learned once and shared across models.

Furthermore, HT explores several hierarchical classification strategies. First, they define an expert grouping in which the individual language labels provided by the organizers are directly assigned distinct categories. Namely, they recognize a Lesser Antillean group (`mart`, `gcf`, and `acf`), Haitian as its own group, a third group containing all other French-based Creoles under consideration (`lou`, `gcr`, `rcf`, `mfe`, and `crs`), and a fourth grouping consisting of the non-Creole labels (`eng`, `fra`, `oth`).[2] Second they introduce three distinct data-driven methods of inferring hiearcharchy: confusion-based grouping, grouping on the basis of representation similarity, and a combined method defined by averaging the other two matrices. (The team used the representation method for their `contrastive1` submission, the confusion method for their `contrastive2`, and the combined method for their `primary`.)

Across all grouping strategies, two frameworks were explored for incorporating group labels into the final classification: hard hierarchy, in which the inferred group acts as a gate to a specialized classifier, and soft hierarchy, in which group-level

[2] These are labeled Groups 1, 3, 2, and 4 respectively.

posterior probabilities act as weights for across-the-board classification. The soft hierarchy was used for all submitted systems.

HT provided a detailed error analysis which emphasizes three points. Training data is described as the main bottleneck, as `gcr` and `rcf`, with 53 and 184 examples respectively in provided data, exhibit the highest error rates. Recall and error rate are shown to vary significantly with training data size (Spearman $\rho = 0.685$, $p = 0.014$) and (Spearman $\rho = -0.685$, $p = 0.014$), respectively, while macro-F1 does not ($\rho = 0.427$, $p = 0.167$).

The team shows that confusions are heavily concentrated between the English, Haitian, and `oth` classes, with mutual confusion between these categories accounting for more than half of all errors. However, bidirectional confusion between Martinican and Guadeloupean Creoles also accounts for 6.9% of the global error rate, and is nearly even in each direction. Remaining errors are attributed to structural factors such as named entities, lexical borrowings, noise, and short texts.

3.2.2 RP

The Rock Pidgins (RP) team combined an *n*-gram-based classifier with a linear SVM to build LID systems. They classified texts into twelve categories, which included francophone Creole languages spoken in the Caribbean and the Indian Ocean, English and French, and an `oth` classification for additional languages.

Their two LID systems differed by their training corpora. `Gwo-K_1` was trained on the `train` subset of `Kreyòl-MT` (Robinson et al., 2024). To mitigate under-representation, `Gwo-K_2` was augmented with 1000 geography-related Wikipedia sentences for French Guianese Creole and 2163 sentences from Cellier (1985) and McLellan (2024) for Réunion Creole. In both systems, features were obtained by extracting strings of three to five characters to capture subwords and orthographic conventions across the different languages. These were weighted by their TF-IDF score to create sentence-level features that were fed into a linear SVM with balanced class weighting.

Overall, `Gwo-K_1` achieved an accuracy of 97.0% on internal tests, and `Gwo-K_2` achieved 97.2%. This improved over the `GlotLID` baseline, which achieved only 74.9%, likely because it does not cover Martinician and Louisiana Creoles. RP found that lower F1 scores were attained for languages with less training data, though this was not a consistent trend. Louisiana Creole had significantly fewer training data than Guadeloupean and Martinician Creoles, but was classified more accurately. The team hypothesized that this was because Louisiana Creole evolved separately from the Creoles spoken in the Caribbean and Indian Ocean, making it more distinguishable. This indicates that more distinct languages might be more data-efficient.

Geography had a marked influence on classification errors: languages tended to be confused with others spoken in the same region. Interestingly, Creole languages tended not to be confused with their high-resource lexifiers, likely due larger amounts of training data for the latter.

4 Evaluation Results

Here we overview our evaluation results for both MT and LID systems subtasks.

4.1 MT evaluation

We use the randomly shuffled versions of the public `Kreyòl-MT` test sets as our primary evaluation sets. Since some teams submitted models for language pairs that are not supported by `kreyol-mt-pubtrain`, and in order to increase the diversity of the evaluation references, we also employed secondary evaluation sets. We used newly split private test sets of `Kreyòl-MT` that were not seen in training by either baseline model, as well as additional test sets from the CreoleVal and MIT-Haiti datasets (Lent et al., 2024), and the APiCS database (Michaelis et al., 2013).

In the following, we refer to the former evaluation set as *Eval Set 1*, and the latter as *Eval Set 2*. We compute the chrF++ (Popović, 2017) and BLEU (Papineni et al., 2002) scores using the default parameters of the `sacrebleu` library on each submission, as well as on the outputs of our two baseline models: `kreyol-mt-pubtrain` and `gemma4-creole-mt`, the new baseline model introduced in Section 2.1.1.

For each group of evaluation sets, we report below the scores of all language pairs for which we received more than one submission. Complete tables with results of all teams across all translation directions can be found in Appendix B. Overall, five directions were not supported by the `kreyol-mt-pubtrain` baseline model: Bislama ↔ English, Pijin ↔ English and Haitian → Spanish. For these pairs, we only used one baseline model. Evaluation results on *Eval Set 1* for translation

directions with a Creole source language are in Table 1, and those with a Creole target language in Table 2. Evaluation results obtained on *Eval Set 2* are reported in Table 3 for translation with a Creole source language, and in Table 4 for a Creole target language. Scores are reported along with the precise source for each *Eval Set 2*, where "KMT" refers to the newly split private test sets of `Kreyòl-MT`.

We first note that across submissions, scores are consistently higher on the *Eval Set 1* than on *Eval Set 2*, with important average differences ranging from approximately -15 chrF++ for UniMelb to -27 for PIK. This is also true of `kreyol-mt-pubtrain`, which has a gap of approximately -20 chrF++ points. *Eval Set 2* is dominated by the private test sets of the `Kreyòl-MT` dataset, which for each language pair often comes from few, specialized sources from a single domain. The notable exception is the `gemma4-creole-mt` baseline, whose performance stays roughly equal (+0.84) on both sets, although this masks substantial direction-specific variation.

Relative rankings between submissions also differ on evaluation sets. UniMelb and UQAM, which both designed more specialized systems focusing on a single Creole language or language group, outperform the two baselines on three out of four considered directions and evaluation sets. UniMelb achieves the best performance across all available evaluation sets for `eng`→ `tpi`, and UQAM for `hat`→ `fra`. CHV scores below the two baselines on all direction pairs and evaluation sets, despite being specific to Haitian Creole. For the large-coverage GMU submission, performance relative to the baselines depend more strongly on the evaluation set: it is highly competitive and outperforms both baselines on *Eval Set 1* for 52 directions out of 80, but only on three of 36 directions on *Eval Set 2*, where `gemma4-creole-mt` achieves higher scores on most shared directions. Despite this, GMU still provides the best-performing system across all available evaluation sets for `eng`→ `gul` and `eng`→ `icr`. The PIK team focused on three Haitian Creole → XX directions, and exhibits substantial variation in performance across target languages. On the two directions available in *Eval Set 1*, its system outperforms the `gemma4-creole-mt` baseline but not `kreyol-mt-pubtrain`. Results on *Eval Set 2* are more mixed: PIK achieves the best performance among all submissions on `hat`→ `eng`, but the lowest on `hat`→ `fra`, while on `hat`→ `spa` it only slightly outperforms `kreyol-mt-pubtrain`. Finally, the KYX submission considers a single direction `rcf` → `eng`, which only appears in *Eval Set 1*. It increases the chrF++ score over `gemma4-creole-mt` by +7.2 points, but falls short of `kreyol-mt-pubtrain` by -3.3 points.

Direction	Team	Model	chrF++	BLEU
hat→eng	*baseline*	`kreyol-mt-pubtrain`	75.9	64.5
	PIK	primary	68.9	50.4
	UQAM	contrastive2	68.3	50.9
	GMU	contrastive1	67.2	49.8
	UQAM	primary	66.2	47.6
	UQAM	contrastive1	66.0	47.8
	baseline	`gemma4-creole-mt`	55.8	37.0
	CHV	contrastive1	53.9	31.6
	CHV	primary	53.7	29.2
	CHV	contrastive2	51.3	29.0
hat→fra	UQAM	contrastive1	64.7	44.2
	UQAM	primary	63.8	42.8
	baseline	`kreyol-mt-pubtrain`	59.9	41.4
	PIK	primary	59.8	33.7
	baseline	`gemma4-creole-mt`	50.9	27.5
	CHV	primary	50.2	22.0
	CHV	contrastive1	46.8	20.3
	CHV	contrastive2	46.4	20.0
rcf→eng	GMU	contrastive2	52.5	43.3
	baseline	`kreyol-mt-pubtrain`	47.7	40.9
	KYX	primary	44.4	36.1
	GMU	contrastive1	42.8	29.7
	GMU	primary	40.8	28.8
	baseline	`gemma4-creole-mt`	40.5	30.9
tpi→eng	*baseline*	`kreyol-mt-pubtrain`	70.1	58.8
	UniMelb	primary	66.0	48.8
	GMU	primary	41.1	18.7
	baseline	`gemma4-creole-mt`	32.3	13.7
	GMU	contrastive1	29.2	12.4

Table 1: Results for Creole → XX on the `Kreyòl-MT` - *Eval Set 1*, Language pairs with submissions from more than one team.

Direction	Team	Model	chrF++	BLEU
eng→hat	*baseline*	`kreyol-mt-pubtrain`	69.8	56.9
	GMU	contrastive1	59.9	37.9
	baseline	`gemma4-creole-mt`	46.2	25.4
	CHV	primary	43.3	18.4
	CHV	contrastive1	43.0	19.1
	CHV	contrastive2	31.1	10.6
eng→tpi	UniMelb	primary	65.8	46.1
	baseline	`kreyol-mt-pubtrain`	62.8	47.2
	GMU	primary	43.1	15.9
	GMU	contrastive1	39.8	11.3
	baseline	`gemma4-creole-mt`	38.5	14.0

Table 2: Results for XX → Creole on the `Kreyòl-MT` - *Eval Set 1* Language pairs with submissions from more than one team.

4.2 LID evaluation

For evaluation of the LID task, we construct a private dataset containing examples across nine francophone Creole languages (St. Lucian Patios - `acf`,

Direction	Eval set	Team	Model	chrF++	BLEU
hat→eng	KMT	PIK	primary	50.7	27.6
	KMT	UQAM	primary	45.0	22.0
	KMT	UQAM	contrastive2	43.9	20.2
	KMT	UQAM	contrastive1	42.5	18.9
	KMT	*baseline*	`kreyol-mt-pubtrain`	41.0	18.1
	KMT	GMU	contrastive1	39.5	15.7
	KMT	*baseline*	`gemma4-creole-mt`	39.1	20.4
	KMT	CHV	contrastive1	35.6	11.8
	KMT	CHV	primary	35.6	11.1
	KMT	CHV	contrastive2	34.0	10.9
hat→fra	KMT	UQAM	primary	42.4	18.8
	KMT	UQAM	contrastive1	39.9	16.4
	KMT	*baseline*	`gemma4-creole-mt`	34.0	15.1
	KMT	*baseline*	`kreyol-mt-pubtrain`	33.7	9.2
	KMT	CHV	primary	31.1	5.4
	KMT	CHV	contrastive1	28.7	5.0
	KMT	CHV	contrastive2	28.2	4.1
	KMT	PIK	primary	23.1	5.6
tpi→eng	CreoleVal	UniMelb	primary	71.5	58.1
	CreoleVal	*baseline*	`kreyol-mt-pubtrain`	35.3	12.1
	CreoleVal	*baseline*	`gemma4-creole-mt`	33.7	15.5
	KMT	UniMelb	primary	46.3	27.5
	KMT	*baseline*	`kreyol-mt-pubtrain`	30.5	10.2
	KMT	GMU	primary	30.5	10.0
	KMT	*baseline*	`gemma4-creole-mt`	29.6	12.9
	KMT	GMU	contrastive1	23.8	6.4

Table 3: Results for Creole → XX - *Eval Set 2*, Language pairs with submissions from more than one team. KMT refers to the private test sets of the `Kreyòl-MT` dataset.

Direction	Eval set	Team	Model	chrF++	BLEU
eng→hat	KMT	*baseline*	`kreyol-mt-pubtrain`	37.4	15.6
	KMT	GMU	contrastive1	34.6	11.7
	KMT	*baseline*	`gemma4-creole-mt`	33.0	13.1
	KMT	CHV	primary	30.2	7.8
	KMT	CHV	contrastive1	30.1	8.3
	KMT	CHV	contrastive2	21.9	4.5
eng→tpi	CreoleVal	UniMelb	primary	70.4	53.8
	CreoleVal	*baseline*	`gemma4-creole-mt`	42.8	17.0
	CreoleVal	*baseline*	`kreyol-mt-pubtrain`	42.5	16.6
	KMT	UniMelb	primary	55.0	34.4
	KMT	*baseline*	`gemma4-creole-mt`	39.1	15.8
	KMT	*baseline*	`kreyol-mt-pubtrain`	37.1	12.2
	KMT	GMU	primary	37.0	13.5
	KMT	GMU	contrastive1	35.1	9.7

Table 4: Results for XX → Creole - *Eval Set 2*, Language pairs with submissions from more than one team. KMT refers to the private test sets of the `Kreyòl-MT` dataset.

Seychellois Creole - `crs`, Guadeloupean Creole - `gcf`, French Guianese Creole - `gcr`, Haitian - `hat`, Louisiana Creole - `lou`, Martinician Creole - `mart`, Mauritian Creole - mfe, and Réunion Creole - `rcf`); two high-resource related languages (English - `eng` and French - `fra`); and an `oth` category for all other languages. The sources for our dataset include blog posts, news articles, and government communications.

We compute accuracy statistics using the `scikit-learn` library and baseline scores using `GlotLID`. Table 5 contains F1 scores for the two teams that submitted results for LID, and the confusion matrices in Figures 1 and 2 indicate which pairs of languages were misclassified as each other across the different systems.

Lang.	GlotLID	RP	HT primary	HT contrastive1	HT contrastive2
acf	0.749	0.744	0.772	0.781	**0.788**
crs	**0.938**	0.881	0.855	0.850	0.839
gcf	**0.807**	0.598	0.604	0.603	0.607
gcr	**0.864**	0.721	0.035	0.058	0.067
hat	**0.924**	0.726	0.731	0.721	0.736
lou	0.000	**0.634**	0.472	0.496	0.420
mart	0.000	**0.491**	0.359	0.368	0.367
mfe	0.911	**0.939**	0.869	0.874	0.838
rcf	0.774	**0.824**	0.530	0.552	0.328
eng	**0.985**	0.771	0.899	0.882	0.877
fra	0.688	**0.872**	0.802	0.778	0.783
oth	0.507	0.606	**0.717**	0.708	0.711
all	0.679	**0.734**	0.639	0.637	0.614

Table 5: Language identification results F1 scores for the languages we evaluate. The submitted systems outperformed the baseline `GlotLID` model on `acf`, `fra`, `lou`, `mart`, `mfe`, `oth` and `rcf`.

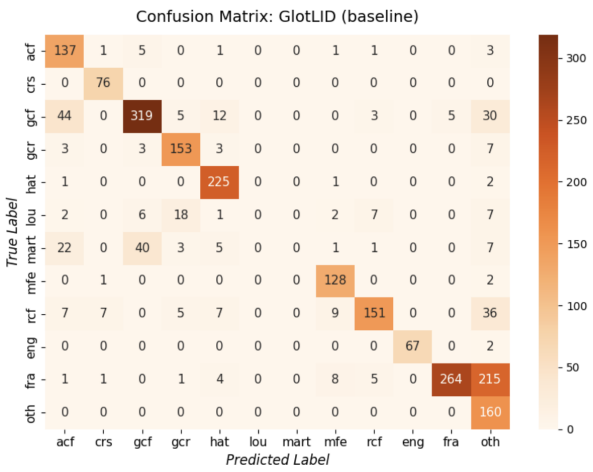

Figure 1: Confusion matrix for the GlotLID baseline model.

On the test set, the `GlotLID` baseline and RP system were the most performant for five languages

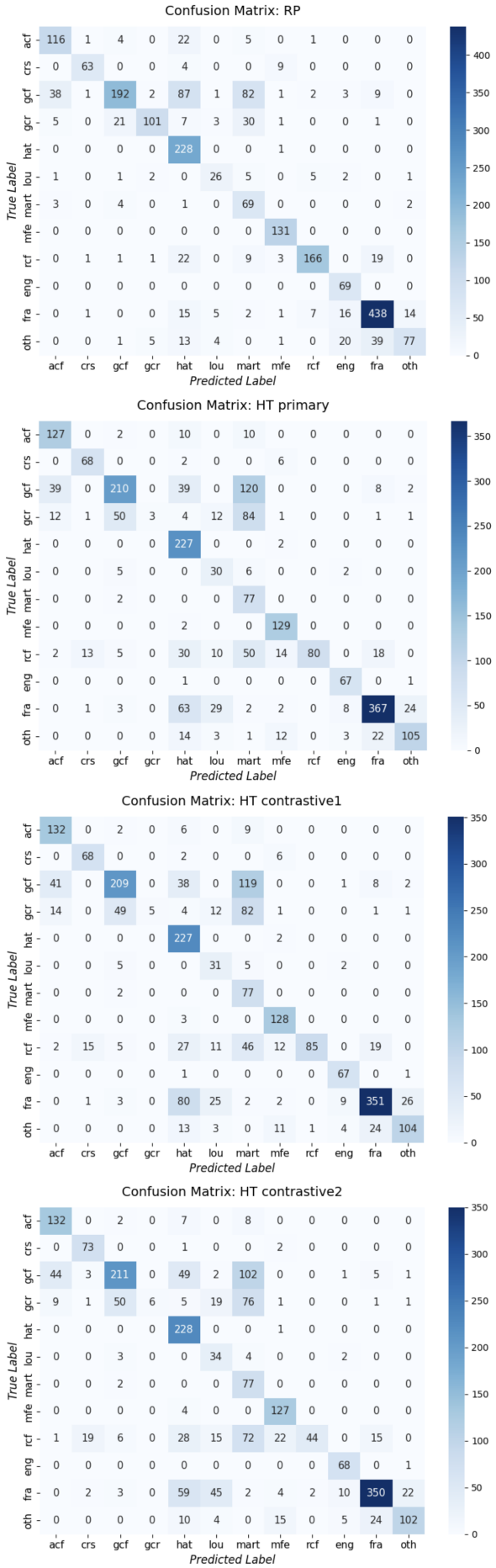

Figure 2: Confusion matrices for the four submitted models.

each. One limitation of `GlotLID` however, is its lack of support for Louisiana Creole and the fact that Martinician and Guadeloupean Creoles are both grouped under the `gcf` label. The HT-`primary` system was the most effective for classifying languages in the `oth` category.

Interestingly, all submitted systems significantly outperformed the baseline model for the identification of French sentences; nearly half of these sentences were misclassified as `oth` by `GlotLID`. In a more fine-grained analysis, we found that these misclassifications primarily consisted of text messages with nonstandard spellings and were spread across 105 languages. The most frequent misidentifications were for Ewondo (6.81% of misclassifications) and Narom (6.38% of misclassifications).

For Louisiana and Martinician Creoles, which are not covered by `GlotLID`, RP was the best performing system, and achieved F1 scores of 0.634 on the former and 0.491 on the latter. RP was also the best performing system for Mauritian Creole, Réunion Creole, and French, achieving F1 scores of 0.939, 0.824 and 0.872 on these languages respectively. RP's technique involved augmenting the Réunion Creole data used for training, which likely contributed to the strong performance.

HT's `contrastive2` method achieved the strongest performance for St. Lucian Creole, with an F1 score of 0.788. For the `oth` category, HT-`primary` achieved the strongest performance, with an F1 score of 0.717.

Across all four submitted systems, the most consistent cluster of misclassifications involved Guadeloupean Creole. This language was most frequently mislabelled as Haitian, St. Lucian Creole and Martinician Creole. This shows that the development of systems that are better able to discriminate between languages in the same geographic cluster is a strong direction for potential research.

When the languages are considered in aggregate, the systems achieved macro-F1 scores of 0.679 for the `GlotLID` baseline, 0.734 for RP, 0.639 for `HP primary`, 0.637 for `HP contrastive1`, and 0.614 for `HP contrastive2`. RP outperformed the other submitted systems by a margin of 0.055. The lack of `GlotLID` support for Louisiana and Martinician Creoles was likely a partial contributor to this performance gap. Overall, the submitted systems had consistently robust performances in spite of being trained on significantly less data than `GlotLID`.

5 Conclusions

We consider the MT systems subtask a tremendous success. Some of the systems submitted (particularly from GMU, UniMelb and UQAM) advanced SOTA for multiple language directions (dozens of directions in the case of GMU). **We found that teams that (1) built and used new datasets and/or (2) employed data augmentation performed best.** (The GMU, UniMelb, and UQAM teams had this in common.) Comparatively, the CHV team's approach using in-context learning instead of actual parameter tuning was, while creative and interesting, not as successful as other approaches. This indicates that fine-tuning with additional data to those provided by task organizers is a winning strategy. This result was corroborated by last year's shared task (Robinson et al., 2025).

RP and HT both achieved SOTA performance compared to the `GlotLID` baseline on the LID task for St. Lucian Creole, Mauritian Creole, Réunion Creole and French (Louisiana and Martinician Creoles were not covered by GlotLID). The RP team achieved a macro-F1 score of 0.7340, which was the strongest across all the submitted models and the `GlotLID` baseline.

Both teams employed a linear SVM, and TF-IDF weighted character level n-grams as features for their models. Data augmentation significantly boosted performance compared to the HT submissions for Réunion Creole (+2163 examples) and French Guianese (+1k examples), which were the least represented languages in the training set, with 189 examples and 346 examples respectively.

This illustrates that training data availability is one of the largest bottlenecks for effective language classification, and highlights the importance of collaborations between data stewards, the linguistic community, and researchers for the advancement of Creole NLP.

Overall, the submitted systems achieved winning scores on various languages that were evaluated, and achieved competitive aggregate macro-F1 scores when compared to `GlotLID`, which is a larger system trained on significantly more data. Both the RP and HT authors show that lightweight, resource-efficient models can be highly effective for LID, and that Creole researchers should continue to focus on ways to increase data-availability and on developing new data-efficient methods. SMS messages and similar highly informal forms of writing stand out as a particularly important category for future LID evaluation.

We look forward to further progress brought through future iterations of this shared task. Please contact us if you are interested in getting involved.

Limitations

A major limitation of this year's shared task was that we only trained the new baseline model on a 50k-sentence subset of our total data. Hence it was likely not as strong as it could have been. We plan to mitigate this for future shared tasks.

We also lament that no team submitted any data to the shared task this year. (I.e. Subtasks 2A and 2B were left untouched; see §2.) In the future we will try to facilitate teams submitting data so that the process is clear, especially considering that multiple participants in the systems subtasks actually developed new datasets that they simply did not submit this year.

Acknowledgments

We thank Nishat Raihan, Sanjeev Khudanpur, Niyati Bafna, Philipp Koehn, and Heather Lent for their support and advice throughout the organization of this shared task.

References

Ife Adebara, AbdelRahim Elmadany, Muhammad Abdul-Mageed, and Alcides Inciarte. 2022. AfroLID: A Neural Language Identification Tool for African Languages. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 1958–1981, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

David Ifeoluwa Adelani, Jade Abbott, Graham Neubig, Daniel D'souza, Julia Kreutzer, Constantine Lignos, Chester Palen-Michel, Happy Buzaaba, Shruti Rijhwani, Sebastian Ruder, et al. 2021. MasakhaNER: Named entity recognition for african languages. *Transactions of the Association for Computational Linguistics*, 9:1116–1131.

David Ifeoluwa Adelani, Jesujoba Oluwadara Alabi, Angela Fan, Julia Kreutzer, Xiaoyu Shen, Machel Reid, Dana Ruiter, Dietrich Klakow, Peter Nabende, Ernie Chang, Tajuddeen Gwadabe, Freshia Sackey, Bonaventure F. P. Dossou, Chris Emezue, Colin Leong, Michael Beukman, Shamsuddeen H. Muhammad, Guyo D. Jarso, Oreen Yousuf, and 26 others. 2022. A few thousand translations go a long way! leveraging pre-trained models for African news translation. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for*

Computational Linguistics: Human Language Technologies, pages 3053–3070, Seattle, United States. Association for Computational Linguistics.

Mateus N Andrade, Mouhamadou Lamine Ba, Idy Diop, and Arlindo O da Veiga. 2026. Less can be more: Towards a parameter-efficient fine-tuning of wav2vec 2.0 xlsr for low-resource cape verdean creole asr. *Resources for African Indigenous Languages (RAIL) 2026@ LREC 2026*, page 62.

Ruth-Ann Armstrong, John Hewitt, and Christopher Manning. 2022. JamPatoisNLI: A jamaican patois natural language inference dataset. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 5307–5320, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Ananya Ayasi. 2025. Krey-all WMT 2025 CreoleMT system description: Language agnostic strategies for low-resource translation. In *Proceedings of the Tenth Conference on Machine Translation*, pages 1158–1165, Suzhou, China. Association for Computational Linguistics.

Ananya Ayasi. 2026. HT WMT 2026 CreoleMT system description: Learning hierarchies for low-resource creole language identification. In *Proceedings of the Eleventh Conference on Machine Translation*, Budapest, Hungary. Association for Computational Linguistics.

Zaheenah Beebee Jameela Boodeea, Sameerchand Pudaruth, Nitish Chooramun, and Aneerav Sukhoo. 2025. Automatic translation between kreol morisien and english using the marian machine translation framework. In *Informatics*, volume 12, page 16. MDPI.

Chris Callison-Burch, Philipp Koehn, Christof Monz, and Omar Zaidan. 2011. Findings of the 2011 workshop on statistical machine translation. In *Proceedings of the sixth workshop on statistical machine translation*, pages 22–64.

Bernard Caron, Marine Courtin, Kim Gerdes, and Sylvain Kahane. 2019. A surface-syntactic UD treebank for Naija. In *Proceedings of the 18th International Workshop on Treebanks and Linguistic Theories (TLT, SyntaxFest 2019)*, pages 13–24. Association for Computational Linguistics.

Pierre Cellier. 1985. *Description syntaxique du créole réunionnais: essai de standardisation*. Ph.D. thesis, Université d'Aix-Marseille I.

Christopher Clarke, Roland Daynauth, Jason Mars, Charlene Wilkinson, and Hubert Devonish. 2024. GuyLingo: The Republic of Guyana creole corpora. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 2: Short Papers)*, pages 792–798, Mexico City, Mexico. Association for Computational Linguistics.

Raj Dabre and Aneerav Sukhoo. 2022. KreolMorisienMT: A dataset for mauritian creole machine translation. In *Findings of the Association for Computational Linguistics: AACL-IJCNLP 2022*, pages 22–29, Online only. Association for Computational Linguistics.

Rasul Dent, Pedro Ortiz Suarez, Thibault Clérice, and Benoît Sagot. 2025. Identifying Rare Languages in Common Crawl Data is a Needles-in-a-Haystack Problem. In *Findings of the Association for Computational Linguistics: EMNLP 2025*, pages 1460–1473, Suzhou, China. Association for Computational Linguistics.

Rasul Dent, Pedro Ortiz Suarez, Thibault Clérice, and Benoît Sagot. 2026. How Should We Model the Probability of a Language? In *Proceedings of the 13th Workshop on NLP for Similar Languages, Varieties and Dialects*, pages 223–233, Rabat, Morocco. Association for Computational Linguistics.

Dominique Fattier. 1998. *Contribution à l'étude de la genèse d'un créole : l'Atlas linguistique d'Haïti, cartes et commentaires*. Ph.D. thesis, Aix-Marseille 1. Thèse de doctorat dirigée par Chaudenson, Robert Lettres Aix-Marseille 1 1998.

Marcell Fekete, Nathaniel Romney Robinson, Ernests Lavrinovics, Djeride Jean-Baptiste, Raj Dabre, Johannes Bjerva, and Heather Lent. 2025. Limited-resource adapters are regularizers, not linguists. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 222–237, Vienna, Austria. Association for Computational Linguistics.

Robert Frederking, Ralf D Brown, and Christopher Hogan. 1997. The diplomat rapid development speech mt system. In *Proceedings of Machine Translation Summit VI: Systems*, pages 261–262.

Pranav Gupta. 2026. wmt-creole-mt-baselines: Multilingual mt baselines on the kreyol-mt dataset. `https://github.com/prannerta100/wmt-creole-mt-baselines`. Baselines (Gemma-4 E2B LoRA fine-tune; mBART-large-50) for 86 language pairs across 50 Creole and related languages, built on the Kreyol-MT dataset (Robinson et al., NAACL 2024).

Zhiwei He, Tian Liang, Wenxiang Jiao, Zhuosheng Zhang, Yujiu Yang, Rui Wang, Zhaopeng Tu, Shuming Shi, and Xing Wang. 2024. Exploring human-like translation strategy with large language models. *Transactions of the Association for Computational Linguistics*, 12:229–246.

Edward J Hu, yelong shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2022. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations*.

Louis Jourdain, Aurélie Pradelles, and Nasredine Semmar. 2026. ChapsVision (CHV) at WMT 2026 CreoleMT: Shifting creole MT from train-time to test-time compute for haitian creole. In *Proceedings of the Eleventh Conference on Machine Translation*, Budapest, Hungary. Association for Computational Linguistics.

Amir Hossein Kargaran, Ayyoob Imani, François Yvon, and Hinrich Schuetze. 2023. GlotLID: Language Identification for Low-Resource Languages. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 6155–6218, Singapore. Association for Computational Linguistics.

Tan Ngoc Le, Tabitha Megane Bellune, and Fatiha Sadat. 2026. UQAM for the haitian team at WMT 2026 CreoleMT system description: Enhancing cultural awareness for haitian creole machine translation. In *Proceedings of the Eleventh Conference on Machine Translation*, Budapest, Hungary. Association for Computational Linguistics.

Éric Le Ferrand and Fabiola Henri. 2026. Child support: Leveraging lexifiers resources to support creoles ASR. In *Proceedings of the Ninth Workshop on the Use of Computational Methods in the Study of Endangered Languages (ComputEL-9)*, pages 111–117, San Diego, California, USA. Association for Computational Linguistics.

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, and 2 others. 2024. CreoleVal: Multilingual multitask benchmarks for creoles. *Transactions of the Association for Computational Linguistics*, 12:950–978.

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*.

Zhengyuan Liu, Shikang Ni, Aiti Aw, and Nancy Chen. 2022. Singlish message paraphrasing: A joint task of creole translation and text normalization. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 3924–3936.

Alina McLellan. 2024. Presentational nana constructions in reunion creole: From bi-clausal cleft to monoclausal construction. *Isogloss. Open Journal of Romance Linguistics*, 10(1):1–27.

Roberto C. Medina. 2025. *Creating NLP Resources for Cape Verdean Creole: Challenges and Solutions for Underrepresented Languages*. Ph.D. thesis. Copyright - Database copyright ProQuest LLC; ProQuest does not claim copyright in the individual underlying works; Last updated - 2026-08-17.

Roberto Carlos Medina, Fernando Batista, and Estanislau Lima. 2025. Crioleset: A new corpus for cape verdean creole, towards robust machine translation. In *EPIA Conference on Artificial Intelligence*, pages 326–338. Springer.

Chutong Meng and Antonios Anastasopoulos. 2026. GMU WMT 2026 CreoleMT system description. In *Proceedings of the Eleventh Conference on Machine Translation*, Budapest, Hungary. Association for Computational Linguistics.

Raphaël Merx, Nick Thieberger, and Ekaterina Vylomova. 2026. The university of melbourne WMT 2026 CreoleMT submission: A domain-balanced approach to low-resource pacific creole machine translation. In *Proceedings of the Eleventh Conference on Machine Translation*, Budapest, Hungary. Association for Computational Linguistics.

Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath, and Magnus Huber, editors. 2013. *APiCS Online*. Max Planck Institute for Evolutionary Anthropology, Leipzig.

Ludovic Mompelat. 2025a. Recommendations for overcoming linguistic barriers in healthcare: Challenges and innovations in NLP for Haitian Creole. In *Proceedings of the Third Workshop on Resources and Representations for Under-Resourced Languages and Domains (RESOURCEFUL-2025)*, pages 20–31, Tallinn, Estonia. University of Tartu Library, Estonia.

Ludovic Mompelat. 2025b. WMT 2025 CreoleMT systems description : Martinican creole and French. In *Proceedings of the Tenth Conference on Machine Translation*, pages 1198–1200, Suzhou, China. Association for Computational Linguistics.

Ludovic Mompelat, Daniel Dakota, and Sandra Kübler. 2022. How to parse a creole: When martinican creole meets French. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 4397–4406, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

Ludovic Mompelat, David Tézil, and Rose Flaure Accilien. 2026. Annotating clinical risk and variation in Haitian Creole medical translation. In *Proceedings of the 20th Linguistic Annotation Workshop (LAW XX)*,

pages 1–11, San Diego, California, USA. Association for Computational Linguistics.

Shamsuddeen Hassan Muhammad, Idris Abdulmumin, Seid Muhie Yimam, David Ifeoluwa Adelani, Ibrahim Said Ahmad, Nedjma Ousidhoum, Abinew Ali Ayele, Saif Mohammad, Meriem Beloucif, and Sebastian Ruder. 2023. SemEval-2023 task 12: Sentiment analysis for African languages (AfriSenti-SemEval). In *Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023)*, pages 2319–2337, Toronto, Canada. Association for Computational Linguistics.

Shamsuddeen Hassan Muhammad, David Ifeoluwa Adelani, Sebastian Ruder, Ibrahim Sa'id Ahmad, Idris Abdulmumin, Bello Shehu Bello, Monojit Choudhury, Chris Chinenye Emezue, Saheed Salahudeen Abdullahi, Anuoluwapo Aremu, Alípio Jorge, and Pavel Brazdil. 2022. NaijaSenti: A Nigerian Twitter sentiment corpus for multilingual sentiment analysis. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 590–602, Marseille, France. European Language Resources Association.

William N. Havard, Renauld Govain, Benjamin Lecouteux, and Emmanuel Schang. 2025. Modèles auto-supervisés de traitement de la parole pour le créole haitien. In *Actes des 32ème Conférence sur le Traitement Automatique des Langues Naturelles (TALN), volume 1 : articles scientifiques originaux*, pages 542–554, Marseille, France. ATALA & ARIA.

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, and 20 others. 2022. No language left behind: Scaling human-centered machine translation. *Preprint*, arXiv:2207.04672.

Kelechi Ogueji and Orevaoghene Ahia. 2019. Pidginunmt: Unsupervised neural machine translation from west african pidgin to english. *arXiv preprint arXiv:1912.03444*.

Wuraola Fisayo Oyewusi, Olubayo Adekanmbi, and Olalekan Akinsande. 2020. Semantic enrichment of nigerian pidgin english for contextual sentiment classification. *arXiv preprint arXiv:2003.12450*.

Rachel Panckhurst, Cédric Lopez, and Mathieu Roche. 2020. A French text-message corpus: 88milSMS. Synthesis and usage. *Corpus*, (20).

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.

Maja Popović. 2017. chrF++: words helping character n-grams. In *Proceedings of the Second Conference on Machine Translation*, pages 612–618, Copenhagen, Denmark. Association for Computational Linguistics.

Qwen Team. 2026a. Qwen3.5: Towards native multimodal agents.

Qwen Team. 2026b. Qwen3.6-27B: Flagship-level coding in a 27B dense model.

Yush Rajcoomar. 2025. KozKreolMRU WMT 2025 CreoleMT system description: Koz kreol: Multistage training for English–mauritian creole MT. In *Proceedings of the Tenth Conference on Machine Translation*, pages 1183–1190, Suzhou, China. Association for Computational Linguistics.

Nathaniel Robinson. 2025a. JHU WMT 2025 CreoleMT system description: Data for belizean Kriol and French guianese creole MT. In *Proceedings of the Tenth Conference on Machine Translation*, pages 1191–1197, Suzhou, China. Association for Computational Linguistics.

Nathaniel Robinson, Claire Bizon Monroc, Rasul Dent, Stefan Watson, Kenton Murray, Raj Dabre, Andre Coy, and Heather Lent. 2025. Findings of the first shared task for creole language machine translation at WMT25. In *Proceedings of the Tenth Conference on Machine Translation*, pages 520–531, Suzhou, China. 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 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. 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 Romney Robinson. 2025b. Take shelter, zanmi: Digitally alerting cyclone victims in their languages. In *Proceedings of the Fourth Workshop on NLP for Positive Impact (NLP4PI)*, pages 70–76, Vienna, Austria. 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*.

Jacqueline Rowe, Ona De Gibert, Mateusz Klimaszewski, Coleman Haley, Alexandra Birch, and Yves Scherrer. 2025a. EdinHelsOW WMT 2025 CreoleMT system description: Improving lusophone creole translation through data augmentation, model merging and LLM post-editing. In *Proceedings of the Tenth Conference on Machine Translation*, pages 1166–1182, Suzhou, China. Association for Computational Linguistics.

Jacqueline Rowe, Edward Gow-Smith, and Mark Hepple. 2025b. Limitations of religious data and the importance of the target domain: Towards machine translation for Guinea-Bissau creole. In *Proceedings of the Eighth Workshop on Technologies for Machine Translation of Low-Resource Languages (LoResMT 2025)*, pages 183–200, Albuquerque, New Mexico, U.S.A. Association for Computational Linguistics.

Emmett Strickland. 2026. Gwo-K kreyol classifier: Rock pidgins (RP) submission to the WMT 2026 CreoleMT language identification shared task. In *Proceedings of the Eleventh Conference on Machine Translation*, Budapest, Hungary. Association for Computational Linguistics.

Pedro Ortiz Suarez, Laurie Burchell, Catherine Arnett, Rafael Mosquera, Sara Hincapié Monsalve, Thom Vaughan, Damian Stewart, Malte Ostendorff, Idris Abdulmumin, Vukosi Marivate, Shamsuddeen Hassan Muhammad, Atnafu Lambebo Tonja, Hend Al-Khalifa, Nadia Ghezaiel Hammouda, Verrah Akinyi Otiende, Tack Hwa Wong, Jakhongir Saydaliev, Melika Nobakhtian, Muhammad Ravi Shulthan Habibi, and 78 others. 2026. CommonLID: Re-evaluating State-of-the-Art Language Identification Performance on Web Data. In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 33063–33080, San Diego, California, United States. Association for Computational Linguistics.

Geerish Suddul, Sarvesh Gianant Singh Gopee, Mithil Bhootooa, and Ravi Foogooa. 2026. Transformer-based machine translation from english to mauritian creole in low-resource settings. In *2026 12th International Conference on Communication and Signal Processing (ICCSP)*, pages 1214–1219.

Gemma Team. 2026. Gemma 4 technical report. *Preprint*, arXiv:2607.02770.

Gemma Team, Aishwarya Kamath, Johan Ferret, Shreya Pathak, Nino Vieillard, Ramona Merhej, Sarah Perrin, Tatiana Matejovicova, Alexandre Ramé, Morgane Rivière, Louis Rouillard, Thomas Mesnard, Geoffrey Cideron, Jean bastien Grill, Sabela Ramos, Edouard Yvinec, Michelle Casbon, Etienne Pot, Ivo Penchev, and 197 others. 2025. Gemma 3 technical report. *Preprint*, arXiv:2503.19786.

Aileen Joan Vicente, Theresse Faith Amamampang, Dunn Dexter Lahaylahay, and Charibeth Cheng. 2026. Chavacanomt: a corpus and evaluation of neural machine translation for philippine creole spanish. *Language Resources and Evaluation*, 60(1):18.

Aileen Joan Vicente and Charibeth Cheng. 2024. Language Identification of Philippine Creole Spanish: Discriminating Chavacano From Related Languages. In *Proceedings of the Eleventh Workshop on NLP for Similar Languages, Varieties, and Dialects (VarDial 2024)*, pages 188–196, Mexico City, Mexico. Association for Computational Linguistics.

Shafqat Mumtaz Virk, Harald Hammarström, Markus Forsberg, and Søren Wichmann. 2020. The DReaM corpus: A multilingual annotated corpus of grammars for the world's languages. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 878–884, Marseille, France. European Language Resources Association.

Hongmin Wang, Yue Zhang, GuangYong Leonard Chan, Jie Yang, and Hai Leong Chieu. 2017. Universal dependencies parsing for colloquial singaporean english. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1732–1744.

Haoran Xu, Young Jin Kim, Amr Sharaf, and Hany Hassan Awadalla. 2024. A paradigm shift in machine translation: Boosting translation performance of large language models. In *The Twelfth International Conference on Learning Representations*.

Binwei Yao, Ming Jiang, Tara Bobinac, Diyi Yang, and Junjie Hu. 2024. Benchmarking machine translation with cultural awareness. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 13078–13096, Miami, Florida, USA. Association for Computational Linguistics.

Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. 2023. React: Synergizing reasoning and acting in language models. *Preprint*, arXiv:2210.03629.

Just Zwennicker and David Stap. 2022. Towards a general purpose machine translation system for sranantongo. *arXiv preprint arXiv:2212.06383*.

Yusuf Öztop. 2026. KYX WMT 2026 CreoleMT system description: A contamination- and orthography-aware evaluation for réunion creole–english. In *Proceedings of the Eleventh Conference on Machine Translation*, Budapest, Hungary. Association for Computational Linguistics.

A Baseline Model Development Details

See Table 6 for scores of `gemma4-creole-mt` on our internal dev sets.

Metric	Base Gemma-4	+ LoRA adapter	Δ
Macro BLEU	7.19	18.48	+11.29
Macro chrF	20.90	36.94	+16.04
Corpus BLEU	5.57	18.48	+12.91
Corpus chrF	21.82	38.48	+16.66

Table 6: `gemma-4-E2B` baseline model's performance metrics.

B Full MT Results

The remaining tables in the paper show the full evaluation results for all MT systems. See Tables 7, 8, 9, and 10.

Direction	Team	Model	chrF++	BLEU
acf→eng	*baseline*	`kreyol-mt-pubtrain`	62.8	46.1
	GMU	primary	57.5	39.3
	GMU	contrastive1	56.9	38.2
	baseline	`gemma4-creole-mt`	34.8	16.7
aoa→eng	GMU	contrastive2	41.5	35.2
	GMU	primary	35.1	23.6
	GMU	contrastive1	28.4	23.2
	baseline	`gemma4-creole-mt`	20.3	16.9
	baseline	`kreyol-mt-pubtrain`	16.7	15.2
bah→eng	GMU	contrastive2	46.8	19.3
	GMU	primary	42.9	24.6
	baseline	`kreyol-mt-pubtrain`	35.8	29.1
	GMU	contrastive1	35.3	28.6
	baseline	`gemma4-creole-mt`	33.2	21.6
brc→eng	GMU	contrastive2	64.0	50.5
	GMU	primary	51.0	34.8
	baseline	`kreyol-mt-pubtrain`	25.1	10.7
	GMU	contrastive1	21.7	7.7
	baseline	`gemma4-creole-mt`	18.9	6.7
bzj→eng	GMU	primary	47.0	27.4
	GMU	contrastive1	42.8	23.3
	baseline	`kreyol-mt-pubtrain`	40.9	21.5
	baseline	`gemma4-creole-mt`	33.8	17.4
bzk→eng	GMU	primary	64.6	49.3
	GMU	contrastive2	63.9	48.4
	baseline	`kreyol-mt-pubtrain`	56.9	40.3
	GMU	contrastive1	56.6	38.8
	baseline	`gemma4-creole-mt`	55.0	37.3
cri→eng	GMU	primary	39.2	24.3
	GMU	contrastive2	34.8	22.0
	baseline	`kreyol-mt-pubtrain`	32.1	19.7
	GMU	contrastive1	26.6	14.6
	baseline	`gemma4-creole-mt`	20.8	8.8
crs→eng	GMU	contrastive1	58.5	34.1
	baseline	`kreyol-mt-pubtrain`	57.6	33.8
	GMU	primary	57.5	33.8
	baseline	`gemma4-creole-mt`	46.2	21.0
dcr→eng	GMU	contrastive2	61.7	48.6
	GMU	primary	54.0	41.6
	GMU	contrastive1	35.5	24.5
	baseline	`gemma4-creole-mt`	34.0	20.3
	baseline	`kreyol-mt-pubtrain`	21.2	12.2
djk→eng	*baseline*	`kreyol-mt-pubtrain`	57.4	37.8
	GMU	primary	36.5	14.2
	GMU	contrastive1	36.0	14.9
	baseline	`gemma4-creole-mt`	27.0	10.8
fab→eng	GMU	contrastive2	36.9	21.4
	GMU	primary	36.6	20.2
	GMU	contrastive1	24.2	10.6
	baseline	`gemma4-creole-mt`	15.8	2.6
	baseline	`kreyol-mt-pubtrain`	12.9	0.8
fng→eng	GMU	contrastive1	40.6	30.9
	GMU	primary	36.8	25.7
	GMU	contrastive2	34.0	16.8
	baseline	`gemma4-creole-mt`	14.2	5.3
	baseline	`kreyol-mt-pubtrain`	10.2	2.7
fpe→eng	GMU	primary	59.4	45.7
	baseline	`kreyol-mt-pubtrain`	40.4	26.7
	GMU	contrastive1	38.5	23.4
	baseline	`gemma4-creole-mt`	34.5	20.6
gcf→eng	GMU	contrastive1	68.8	53.8
	baseline	`kreyol-mt-pubtrain`	67.5	52.8
	GMU	contrastive2	66.9	50.9
	GMU	primary	66.6	51.4
	baseline	`gemma4-creole-mt`	54.2	36.3
jam→eng	GMU	primary	67.4	52.9
	GMU	contrastive1	64.7	53.0
	baseline	`kreyol-mt-pubtrain`	53.4	34.9
	baseline	`gemma4-creole-mt`	52.0	32.3
kea→eng	GMU	contrastive1	84.9	77.3
	GMU	primary	80.2	69.8
	baseline	`kreyol-mt-pubtrain`	80.1	71.6
	baseline	`gemma4-creole-mt`	55.6	32.9
kri→eng	GMU	contrastive1	46.5	31.4
	GMU	primary	45.2	26.8
	baseline	`gemma4-creole-mt`	38.5	21.0
	baseline	`kreyol-mt-pubtrain`	27.6	9.4
ktu→eng	GMU	primary	34.4	16.0
	GMU	contrastive1	23.2	11.4
	baseline	`gemma4-creole-mt`	20.8	4.9
	baseline	`kreyol-mt-pubtrain`	10.3	0.8
lou→eng	*baseline*	`kreyol-mt-pubtrain`	51.4	32.7
	GMU	contrastive1	48.6	28.4
	GMU	primary	48.0	28.9
	baseline	`gemma4-creole-mt`	39.6	19.2
mart1259→eng	*baseline*	`kreyol-mt-pubtrain`	90.7	78.4
	GMU	primary	82.6	70.7
	GMU	contrastive1	80.3	68.9
	GMU	contrastive2	79.0	65.7
	baseline	`gemma4-creole-mt`	72.6	56.8
mfe→eng	*baseline*	`kreyol-mt-pubtrain`	46.3	25.0
	GMU	contrastive1	44.8	22.9
	GMU	primary	44.4	22.4
	baseline	`gemma4-creole-mt`	34.1	13.1
mue→eng	GMU	contrastive2	42.3	27.1
	GMU	primary	31.1	15.9
	GMU	contrastive1	18.2	3.7
	baseline	`gemma4-creole-mt`	13.7	1.3
	baseline	`kreyol-mt-pubtrain`	8.6	0.5
pap→eng	*baseline*	`kreyol-mt-pubtrain`	79.5	70.1
	GMU	primary	57.2	37.7
	GMU	contrastive1	54.1	35.7
	baseline	`gemma4-creole-mt`	53.1	35.0
pap→fra	*baseline*	`kreyol-mt-pubtrain`	57.4	29.6
	baseline	`gemma4-creole-mt`	37.3	15.9
pcm→eng	GMU	contrastive1	67.2	47.2
	baseline	`kreyol-mt-pubtrain`	65.4	45.0
	GMU	primary	65.4	43.9
	baseline	`gemma4-creole-mt`	57.1	35.4
pov→eng	*baseline*	`kreyol-mt-pubtrain`	63.2	49.7
	GMU	contrastive2	62.1	45.3
	GMU	contrastive1	56.5	39.3
	GMU	primary	56.0	38.7
	baseline	`gemma4-creole-mt`	34.8	21.8
pre→eng	GMU	primary	68.9	52.7
	GMU	contrastive2	63.8	50.2
	GMU	contrastive1	47.7	34.2
	baseline	`gemma4-creole-mt`	22.5	4.0
	baseline	`kreyol-mt-pubtrain`	15.9	2.0
rcf→eng	GMU	contrastive2	52.5	43.3
	baseline	`kreyol-mt-pubtrain`	47.7	40.9
	KYX	primary	44.4	36.1
	GMU	contrastive1	42.8	29.7
	GMU	primary	40.8	28.8
	baseline	`gemma4-creole-mt`	40.5	30.9

Direction	Team	Model	chrF++	BLEU
gcr→eng	GMU	contrastive2	50.5	29.9
	GMU	primary	49.1	30.8
	GMU	contrastive1	49.0	31.1
	baseline	gemma4-creole-mt	41.3	20.8
	baseline	kreyol-mt-pubtrain	39.3	19.8
gpe→eng	GMU	primary	57.4	37.0
	GMU	contrastive2	55.9	33.0
	GMU	contrastive1	45.1	21.7
	baseline	gemma4-creole-mt	44.8	21.5
	baseline	kreyol-mt-pubtrain	24.8	6.2
gul→eng	GMU	primary	62.4	44.0
	GMU	contrastive1	56.0	32.7
	baseline	gemma4-creole-mt	41.3	21.6
	baseline	kreyol-mt-pubtrain	34.8	13.4
gyn→eng	GMU	contrastive2	48.4	31.5
	GMU	primary	46.1	28.2
	baseline	kreyol-mt-pubtrain	44.1	28.1
	GMU	contrastive1	40.0	23.2
	baseline	gemma4-creole-mt	30.5	15.6
hat→eng	*baseline*	kreyol-mt-pubtrain	75.9	64.5
	PIK	primary	68.9	50.4
	UQAM	contrastive2	68.3	50.9
	GMU	contrastive1	67.2	49.8
	UQAM	primary	66.2	47.6
	UQAM	contrastive1	66.0	47.8
	baseline	gemma4-creole-mt	55.8	37.0
	CHV	contrastive1	53.9	31.6
	CHV	primary	53.7	29.2
	CHV	contrastive2	51.3	29.0
hat→fra	UQAM	contrastive1	64.7	44.2
	UQAM	primary	63.8	42.8
	baseline	kreyol-mt-pubtrain	59.9	41.4
	PIK	primary	59.8	33.7
	baseline	gemma4-creole-mt	50.9	27.5
	CHV	primary	50.2	22.0
	CHV	contrastive1	46.8	20.3
	CHV	contrastive2	46.4	20.0
icr→eng	GMU	primary	73.1	58.9
	baseline	kreyol-mt-pubtrain	67.7	53.5
	GMU	contrastive1	62.2	45.8
	baseline	gemma4-creole-mt	44.0	29.6
sag→eng	*baseline*	kreyol-mt-pubtrain	41.4	25.9
	GMU	contrastive1	30.4	14.1
	GMU	primary	29.3	11.9
	baseline	gemma4-creole-mt	23.7	6.3
srm→eng	GMU	primary	29.5	18.8
	baseline	kreyol-mt-pubtrain	28.7	19.0
	GMU	contrastive1	23.1	6.2
	baseline	gemma4-creole-mt	15.4	2.5
srn→eng	GMU	primary	48.2	29.3
	baseline	kreyol-mt-pubtrain	44.8	27.4
	GMU	contrastive1	40.5	20.8
	baseline	gemma4-creole-mt	27.9	11.7
svc→eng	GMU	contrastive2	67.9	56.2
	baseline	kreyol-mt-pubtrain	60.3	46.4
	GMU	primary	57.9	42.1
	GMU	contrastive1	51.9	23.6
	baseline	gemma4-creole-mt	42.0	25.7
tpi→eng	*baseline*	kreyol-mt-pubtrain	70.1	58.8
	UNIMELB	primary	66.0	48.8
	GMU	primary	41.1	18.7
	baseline	gemma4-creole-mt	32.3	13.7
	GMU	contrastive1	29.2	12.4
trf→eng	GMU	contrastive2	82.2	66.9
	baseline	kreyol-mt-pubtrain	78.0	61.3
	baseline	gemma4-creole-mt	77.2	60.3
	GMU	contrastive1	76.3	58.9
	GMU	primary	76.1	58.1
wes→eng	GMU	primary	67.0	56.3
	GMU	contrastive2	60.3	44.0
	GMU	contrastive1	59.4	46.0
	baseline	gemma4-creole-mt	39.8	24.1
	baseline	kreyol-mt-pubtrain	25.5	7.2

Table 7: Evaluation on *Eval Set 1* - Creole → XX.

Direction	Team	Model	chrF++	BLEU
eng→acf	*baseline*	kreyol-mt-pubtrain	64.7	46.1
	GMU	contrastive1	58.7	39.9
	GMU	primary	58.3	38.9
	baseline	gemma4-creole-mt	8.5	0.1
eng→aoa	GMU	contrastive2	45.0	33.0
	GMU	primary	30.6	18.8
	GMU	contrastive1	27.2	18.4
	baseline	gemma4-creole-mt	20.4	18.2
	baseline	kreyol-mt-pubtrain	13.1	15.8
eng→bah	*baseline*	kreyol-mt-pubtrain	29.3	13.1
	GMU	contrastive1	26.8	8.6
	GMU	primary	26.5	11.5
	baseline	gemma4-creole-mt	22.8	10.3
	GMU	contrastive2	18.8	8.0
eng→brc	GMU	contrastive2	56.5	34.4
	GMU	primary	49.0	21.9
	baseline	kreyol-mt-pubtrain	21.5	9.4
	GMU	contrastive1	19.5	5.0
	baseline	gemma4-creole-mt	16.6	4.1
eng→kea	GMU	contrastive1	53.4	27.6
	GMU	primary	51.8	26.6
	baseline	kreyol-mt-pubtrain	51.0	26.2
	baseline	gemma4-creole-mt	38.2	12.5
eng→kri	*baseline*	gemma4-creole-mt	38.0	16.9
	GMU	primary	37.6	23.9
	GMU	contrastive1	34.7	17.8
	baseline	kreyol-mt-pubtrain	10.0	2.3
eng→ktu	GMU	primary	40.7	16.4
	GMU	contrastive1	33.0	12.2
	baseline	gemma4-creole-mt	26.9	1.9
	baseline	kreyol-mt-pubtrain	8.4	1.5
eng→lou	*baseline*	kreyol-mt-pubtrain	49.8	28.2
	GMU	contrastive1	46.0	22.5
	GMU	primary	44.6	22.4
	baseline	gemma4-creole-mt	32.4	11.6

Direction	Team	Model	chrF++	BLEU
eng→bzj	*baseline*	kreyol-mt-pubtrain	34.6	14.6
	GMU	contrastive1	30.0	8.9
	GMU	primary	29.4	5.1
	baseline	gemma4-creole-mt	21.5	2.0
eng→bzk	GMU	contrastive2	53.4	33.8
	GMU	contrastive1	45.9	24.6
	baseline	kreyol-mt-pubtrain	45.3	26.2
	GMU	primary	45.0	24.1
	baseline	gemma4-creole-mt	12.7	6.1
eng→cri	GMU	contrastive2	51.2	30.7
	GMU	contrastive1	35.1	15.9
	baseline	kreyol-mt-pubtrain	33.0	16.0
	GMU	primary	28.9	8.6
	baseline	gemma4-creole-mt	18.6	5.0
eng→crs	*baseline*	kreyol-mt-pubtrain	56.2	29.6
	GMU	primary	55.0	27.9
	GMU	contrastive1	54.7	27.0
	baseline	gemma4-creole-mt	36.6	11.2
eng→dcr	GMU	contrastive2	45.3	33.8
	GMU	primary	35.7	19.4
	baseline	gemma4-creole-mt	32.1	25.3
	GMU	contrastive1	28.9	18.9
	baseline	kreyol-mt-pubtrain	10.3	8.7
eng→djk	*baseline*	kreyol-mt-pubtrain	74.6	68.2
	GMU	primary	35.8	16.9
	GMU	contrastive1	35.4	15.3
	baseline	gemma4-creole-mt	12.9	0.3
eng→fab	GMU	contrastive2	33.4	18.8
	GMU	primary	27.8	10.9
	GMU	contrastive1	20.4	6.5
	baseline	gemma4-creole-mt	11.1	0.7
	baseline	kreyol-mt-pubtrain	8.8	0.7
eng→fng	GMU	contrastive2	55.1	34.8
	GMU	primary	51.6	32.1
	GMU	contrastive1	41.8	22.4
	baseline	gemma4-creole-mt	26.7	9.7
	baseline	kreyol-mt-pubtrain	9.4	4.6
eng→fpe	GMU	primary	37.0	17.9
	baseline	kreyol-mt-pubtrain	30.8	20.1
	GMU	contrastive1	27.8	11.3
	baseline	gemma4-creole-mt	11.7	9.3
eng→gcf	*baseline*	kreyol-mt-pubtrain	73.7	60.6
	GMU	contrastive2	70.4	56.0
	GMU	contrastive1	70.2	55.7
	GMU	primary	67.8	52.8
	baseline	gemma4-creole-mt	50.6	29.0
eng→gcr	GMU	contrastive2	48.9	22.3
	GMU	primary	36.4	15.5
	GMU	contrastive1	35.4	19.6
	baseline	gemma4-creole-mt	28.8	11.5
	baseline	kreyol-mt-pubtrain	25.9	9.6
eng→gpe	GMU	primary	40.1	21.9
	GMU	contrastive1	34.6	14.8
	GMU	contrastive2	28.4	13.7
	baseline	gemma4-creole-mt	25.0	8.0
	baseline	kreyol-mt-pubtrain	7.6	1.9
eng→gul	GMU	contrastive1	39.7	11.0
	GMU	primary	31.2	11.8
	baseline	gemma4-creole-mt	26.6	7.0
	baseline	kreyol-mt-pubtrain	22.3	6.6
eng→gyn	GMU	contrastive2	52.2	30.8
	GMU	primary	48.5	24.8
	baseline	kreyol-mt-pubtrain	38.0	16.0
	GMU	contrastive1	36.2	13.4
	baseline	gemma4-creole-mt	27.9	6.0

Direction	Team	Model	chrF++	BLEU
eng→mart1259	*baseline*	kreyol-mt-pubtrain	73.1	53.6
	GMU	contrastive2	61.9	42.4
	GMU	contrastive1	59.6	36.6
	GMU	primary	56.8	27.5
	baseline	gemma4-creole-mt	52.2	30.2
eng→mfe	*baseline*	kreyol-mt-pubtrain	49.6	28.7
	GMU	contrastive1	44.0	19.2
	GMU	primary	43.7	19.5
	baseline	gemma4-creole-mt	29.3	7.2
eng→mue	GMU	contrastive2	32.0	10.3
	GMU	primary	21.4	6.0
	GMU	contrastive1	16.4	4.3
	baseline	gemma4-creole-mt	8.8	0.6
	baseline	kreyol-mt-pubtrain	7.0	0.4
eng→pap	*baseline*	kreyol-mt-pubtrain	69.2	53.9
	GMU	primary	49.8	25.9
	GMU	contrastive1	49.7	25.0
	baseline	gemma4-creole-mt	43.7	22.4
eng→pcm	GMU	contrastive1	61.4	35.8
	baseline	kreyol-mt-pubtrain	61.0	36.0
	GMU	primary	59.9	33.8
	baseline	gemma4-creole-mt	44.8	20.3
eng→pov	*baseline*	kreyol-mt-pubtrain	52.4	36.7
	GMU	contrastive2	51.1	31.7
	GMU	primary	48.8	29.2
	GMU	contrastive1	46.0	25.4
	baseline	gemma4-creole-mt	25.3	7.6
eng→pre	GMU	contrastive2	51.1	27.6
	GMU	primary	50.8	31.3
	GMU	contrastive1	32.7	11.5
	baseline	gemma4-creole-mt	25.3	6.8
	baseline	kreyol-mt-pubtrain	10.7	1.8
eng→rcf	*baseline*	kreyol-mt-pubtrain	39.5	31.0
	GMU	contrastive2	38.6	23.1
	GMU	contrastive1	32.1	17.0
	GMU	primary	31.0	13.7
	baseline	gemma4-creole-mt	28.0	19.1
eng→sag	*baseline*	kreyol-mt-pubtrain	36.3	13.2
	GMU	primary	34.9	10.4
	GMU	contrastive1	32.4	8.4
	baseline	gemma4-creole-mt	29.1	4.8
eng→srm	*baseline*	kreyol-mt-pubtrain	24.1	13.3
	GMU	contrastive1	19.7	5.0
	baseline	gemma4-creole-mt	17.9	2.7
	GMU	primary	17.5	1.6
eng→srn	GMU	primary	37.4	20.1
	baseline	kreyol-mt-pubtrain	35.6	14.0
	baseline	gemma4-creole-mt	32.2	13.3
	GMU	contrastive1	31.7	12.9
eng→svc	GMU	contrastive2	56.5	37.5
	baseline	kreyol-mt-pubtrain	54.6	36.8
	GMU	contrastive1	50.6	29.3
	GMU	primary	48.7	26.9
	baseline	gemma4-creole-mt	38.9	18.0
eng→tpi	UNIMELB	primary	65.8	46.1
	baseline	kreyol-mt-pubtrain	62.8	47.2
	GMU	primary	43.1	15.9
	GMU	contrastive1	39.8	11.3
	baseline	gemma4-creole-mt	38.5	14.0
eng→trf	*baseline*	kreyol-mt-pubtrain	89.0	82.5
	baseline	gemma4-creole-mt	67.6	43.9
	GMU	contrastive2	63.0	34.0
	GMU	primary	58.2	30.3
	GMU	contrastive1	56.7	28.4

Direction	Team	Model	chrF++	BLEU	Direction	Team	Model	chrF++	BLEU
eng→hat	*baseline*	`kreyol-mt-pubtrain`	69.8	56.9	**eng→wes**	GMU	contrastive2	49.5	26.2
	GMU	contrastive1	59.9	37.9		GMU	primary	45.8	23.6
	baseline	`gemma4-creole-mt`	46.2	25.4		*baseline*	`gemma4-creole-mt`	33.1	16.3
	CHV	primary	43.3	18.4		GMU	contrastive1	32.8	10.9
	CHV	contrastive1	43.0	19.1		*baseline*	`kreyol-mt-pubtrain`	12.7	1.4
	CHV	contrastive2	31.1	10.6					
eng→icr	GMU	primary	70.2	57.1	**fra→hat**	*baseline*	`kreyol-mt-pubtrain`	64.0	51.0
	baseline	`kreyol-mt-pubtrain`	57.2	41.5		*baseline*	`gemma4-creole-mt`	43.4	22.2
	GMU	contrastive1	57.1	38.8		CHV	contrastive1	40.4	15.2
	baseline	`gemma4-creole-mt`	11.3	5.7		CHV	primary	39.9	14.0
						CHV	contrastive2	33.0	9.6
eng→jam	GMU	contrastive1	62.5	44.7	**fra→pap**	*baseline*	`kreyol-mt-pubtrain`	46.4	30.2
	GMU	primary	58.7	36.4		*baseline*	`gemma4-creole-mt`	34.2	15.3
	baseline	`kreyol-mt-pubtrain`	43.1	24.5					
	baseline	`gemma4-creole-mt`	24.4	9.7					

Table 8: Evaluation on *Eval Set 1* - XX → Creole.

Direction	Eval set	Team	Model	chrF++	BLEU	Direction	Eval set	Team	Model	chrF++	BLEU
acf→eng	KMT	*baseline*	`gemma4-creole-mt`	35.0	20.1	**kri→eng**	KMT	*baseline*	`gemma4-creole-mt`	35.4	18.6
	KMT	*baseline*	`kreyol-mt-pubtrain`	34.1	9.9		KMT	GMU	primary	32.5	10.2
	KMT	GMU	primary	32.5	7.9		KMT	GMU	contrastive1	27.7	7.2
	KMT	GMU	contrastive1	32.2	9.6		KMT	*baseline*	`kreyol-mt-pubtrain`	21.6	3.4
bis→eng	APiCS	UNIMELB	primary	43.6	22.0	**ktu→eng**	KMT	*baseline*	`gemma4-creole-mt`	36.2	21.2
	APiCS	*baseline*	`gemma4-creole-mt`	23.6	6.6		KMT	GMU	primary	21.7	3.6
	CreoleVal	UNIMELB	primary	63.4	47.1		KMT	GMU	contrastive1	15.9	0.9
	CreoleVal	*baseline*	`gemma4-creole-mt`	26.2	9.2		KMT	*baseline*	`kreyol-mt-pubtrain`	12.4	0.1
bzj→eng	KMT	*baseline*	`gemma4-creole-mt`	40.9	25.0	**mfe→eng**	KMT	*baseline*	`gemma4-creole-mt`	42.0	23.8
	KMT	GMU	primary	36.2	11.5		KMT	GMU	primary	41.6	19.1
	KMT	GMU	contrastive1	30.8	8.6		KMT	GMU	contrastive1	41.4	19.4
	KMT	*baseline*	`kreyol-mt-pubtrain`	27.2	5.5		KMT	*baseline*	`kreyol-mt-pubtrain`	41.0	18.2
crs→eng	KMT	*baseline*	`kreyol-mt-pubtrain`	42.8	21.0	**pap→eng**	KMT	*baseline*	`kreyol-mt-pubtrain`	50.3	28.6
	KMT	GMU	contrastive1	40.2	19.0		KMT	*baseline*	`gemma4-creole-mt`	46.9	28.2
	KMT	*baseline*	`gemma4-creole-mt`	40.1	22.3		KMT	GMU	primary	46.5	24.8
	KMT	GMU	primary	39.4	17.8		KMT	GMU	contrastive1	45.6	24.4
djk→eng	KMT	*baseline*	`kreyol-mt-pubtrain`	31.5	13.2	**pap→fra**	KMT	*baseline*	`gemma4-creole-mt`	41.4	24.8
	KMT	*baseline*	`gemma4-creole-mt`	30.2	15.5		KMT	*baseline*	`kreyol-mt-pubtrain`	37.3	12.9
	KMT	GMU	primary	28.7	9.8						
	KMT	GMU	contrastive1	26.0	8.2						
gcr→eng	KMT	*baseline*	`gemma4-creole-mt`	50.1	36.9	**pcm→eng**	KMT	*baseline*	`gemma4-creole-mt`	36.1	20.5
	KMT	GMU	contrastive2	37.1	16.6		KMT	GMU	contrastive1	32.4	11.5
	KMT	*baseline*	`kreyol-mt-pubtrain`	30.1	9.3		KMT	GMU	primary	31.9	10.6
	KMT	GMU	primary	27.5	8.0		KMT	*baseline*	`kreyol-mt-pubtrain`	30.4	9.4
	KMT	GMU	contrastive1	27.5	9.3						
gul→eng	KMT	*baseline*	`gemma4-creole-mt`	39.7	23.1	**pis→eng**	CreoleVal	UNIMELB	primary	67.0	52.5
	KMT	GMU	primary	32.4	7.2		CreoleVal	*baseline*	`gemma4-creole-mt`	28.7	11.4
	KMT	GMU	contrastive1	32.1	7.8						
	KMT	*baseline*	`kreyol-mt-pubtrain`	21.1	2.4						
hat→eng	KMT	PIK	primary	50.7	27.6	**sag→eng**	KMT	*baseline*	`gemma4-creole-mt`	28.7	12.9
	KMT	UQAM	primary	45.0	22.0		KMT	*baseline*	`kreyol-mt-pubtrain`	25.0	7.0
	KMT	UQAM	contrastive2	43.9	20.2		KMT	GMU	primary	22.4	5.5
	KMT	UQAM	contrastive1	42.5	18.9		KMT	GMU	contrastive1	21.2	4.8
	KMT	*baseline*	`kreyol-mt-pubtrain`	41.0	18.1						
	KMT	GMU	contrastive1	39.5	15.7						
	KMT	*baseline*	`gemma4-creole-mt`	39.1	20.4						
	KMT	CHV	contrastive1	35.6	11.8						
	KMT	CHV	primary	35.6	11.1						
	KMT	CHV	contrastive2	34.0	10.9						
hat→fra	KMT	UQAM	primary	42.4	18.8	**srm→eng**	KMT	*baseline*	`gemma4-creole-mt`	29.3	14.1
	KMT	UQAM	contrastive1	39.9	16.4		KMT	GMU	primary	25.5	4.4
	KMT	*baseline*	`gemma4-creole-mt`	34.0	15.1		KMT	*baseline*	`kreyol-mt-pubtrain`	22.4	4.3
	KMT	*baseline*	`kreyol-mt-pubtrain`	33.7	9.2		KMT	GMU	contrastive1	18.6	2.4
	KMT	CHV	primary	31.1	5.4						
	KMT	CHV	contrastive1	28.7	5.0						
	KMT	CHV	contrastive2	28.2	4.1						
	KMT	PIK	primary	23.1	5.6						
hat→spa	MIT-Haiti	*baseline*	`gemma4-creole-mt`	37.7	13.7	**srn→eng**	KMT	*baseline*	`gemma4-creole-mt`	36.9	19.2
	MIT-Haiti	PIK	primary	14.2	3.7		KMT	GMU	primary	32.6	10.8
	MIT-Haiti	*baseline*	`kreyol-mt-pubtrain`	12.1	0.7		KMT	*baseline*	`kreyol-mt-pubtrain`	29.8	9.7
							KMT	GMU	contrastive1	27.6	8.2
icr→eng	KMT	*baseline*	`gemma4-creole-mt`	38.0	23.0	**tpi→eng**	CreoleVal	UNIMELB	primary	71.5	58.1
	KMT	GMU	primary	35.3	9.4		CreoleVal	*baseline*	`kreyol-mt-pubtrain`	35.3	12.1
	KMT	GMU	contrastive1	33.8	10.2		CreoleVal	*baseline*	`gemma4-creole-mt`	33.7	15.5
	KMT	*baseline*	`kreyol-mt-pubtrain`	31.2	7.3		KMT	UNIMELB	primary	46.3	27.5
							KMT	*baseline*	`kreyol-mt-pubtrain`	30.5	10.2
							KMT	GMU	primary	30.5	10.0
							KMT	*baseline*	`gemma4-creole-mt`	29.6	12.9
							KMT	GMU	contrastive1	23.8	6.4

Direction	Eval set	Team	Model	chrF++	BLEU	Direction	Eval set	Team	Model	chrF++	BLEU
jam→eng	KMT	*baseline*	gemma4-creole-mt	39.6	24.5						
	KMT	GMU	primary	32.3	8.3						
	KMT	GMU	contrastive1	31.1	9.1						
	KMT	*baseline*	kreyol-mt-pubtrain	29.2	6.9						

Table 9: Evaluation on *Eval Set 2*, Creole → XX.

Direction	Eval set	Team	Model	chrF++	BLEU	Direction	Eval set	Team	Model	chrF++	BLEU
eng→acf	KMT	GMU	primary	25.6	5.7	**eng→ktu**	KMT	*baseline*	gemma4-creole-mt	50.0	25.9
	KMT	GMU	contrastive1	25.3	6.5		KMT	GMU	primary	26.5	2.2
	KMT	*baseline*	kreyol-mt-pubtrain	24.8	5.8		KMT	GMU	contrastive1	14.6	0.1
	KMT	*baseline*	gemma4-creole-mt	12.7	0.3		KMT	*baseline*	kreyol-mt-pubtrain	11.9	0.0
eng→bis	APiCS	UNIMELB	primary	55.0	24.6	**eng→mfe**	KMT	*baseline*	kreyol-mt-pubtrain	30.3	10.5
	APiCS	*baseline*	gemma4-creole-mt	11.3	2.1		KMT	*baseline*	gemma4-creole-mt	28.9	10.0
	CreoleVal	UNIMELB	primary	46.8	23.9		KMT	GMU	primary	28.2	9.8
	CreoleVal	*baseline*	gemma4-creole-mt	12.0	0.1		KMT	GMU	contrastive1	28.1	8.6
eng→bzj	KMT	*baseline*	gemma4-creole-mt	36.2	16.6	**eng→pap**	KMT	*baseline*	kreyol-mt-pubtrain	48.4	26.3
	KMT	GMU	primary	28.8	8.9		KMT	GMU	contrastive1	40.7	17.5
	KMT	GMU	contrastive1	14.1	1.2		KMT	GMU	primary	39.7	16.6
	KMT	*baseline*	kreyol-mt-pubtrain	13.1	0.9		KMT	*baseline*	gemma4-creole-mt	39.1	18.1
eng→crs	KMT	*baseline*	gemma4-creole-mt	41.2	21.4	**eng→pcm**	KMT	*baseline*	gemma4-creole-mt	34.3	15.1
	KMT	*baseline*	kreyol-mt-pubtrain	39.4	17.1		KMT	GMU	contrastive1	26.5	6.8
	KMT	GMU	contrastive1	36.1	13.2		KMT	GMU	primary	26.2	6.3
	KMT	GMU	primary	35.6	13.6		KMT	*baseline*	kreyol-mt-pubtrain	25.0	5.8
eng→djk	KMT	*baseline*	kreyol-mt-pubtrain	30.4	15.8	**eng→pis**	CreoleVal	UNIMELB	primary	49.0	22.6
	KMT	GMU	primary	30.1	10.5		CreoleVal	*baseline*	gemma4-creole-mt	11.3	0.1
	KMT	GMU	contrastive1	30.0	9.7						
	KMT	*baseline*	gemma4-creole-mt	13.9	1.9						
eng→gcr	KMT	*baseline*	gemma4-creole-mt	48.4	36.8	**eng→sag**	KMT	*baseline*	gemma4-creole-mt	36.0	15.0
	KMT	GMU	contrastive2	22.7	6.0		KMT	*baseline*	kreyol-mt-pubtrain	27.4	6.4
	KMT	GMU	primary	20.1	3.6		KMT	GMU	primary	26.6	6.5
	KMT	*baseline*	kreyol-mt-pubtrain	19.3	2.2		KMT	GMU	contrastive1	24.2	4.6
	KMT	GMU	contrastive1	18.9	2.9						
eng→gul	KMT	GMU	contrastive1	20.0	2.4	**eng→srm**	KMT	*baseline*	gemma4-creole-mt	31.6	12.7
	KMT	*baseline*	gemma4-creole-mt	17.6	1.8		KMT	*baseline*	kreyol-mt-pubtrain	14.2	1.6
	KMT	GMU	primary	16.1	1.1		KMT	GMU	contrastive1	13.8	1.8
	KMT	*baseline*	kreyol-mt-pubtrain	12.0	0.1		KMT	GMU	primary	12.1	1.7
eng→hat	KMT	*baseline*	kreyol-mt-pubtrain	37.4	15.6	**eng→srn**	KMT	*baseline*	gemma4-creole-mt	36.2	15.9
	KMT	GMU	contrastive1	34.6	11.7		KMT	*baseline*	kreyol-mt-pubtrain	21.8	5.0
	KMT	*baseline*	gemma4-creole-mt	33.0	13.1		KMT	GMU	contrastive1	21.6	5.2
	KMT	CHV	primary	30.2	7.8		KMT	GMU	primary	20.9	5.0
	KMT	CHV	contrastive1	30.1	8.3						
	KMT	CHV	contrastive2	21.9	4.5						
eng→icr	KMT	GMU	primary	20.4	2.9	**eng→tpi**	CreoleVal	UNIMELB	primary	70.4	53.8
	KMT	GMU	contrastive1	19.1	2.4		CreoleVal	*baseline*	gemma4-creole-mt	42.8	17.0
	KMT	*baseline*	kreyol-mt-pubtrain	17.2	1.5		CreoleVal	*baseline*	kreyol-mt-pubtrain	42.5	16.6
	KMT	*baseline*	gemma4-creole-mt	11.0	0.2		KMT	UNIMELB	primary	55.0	34.4
							KMT	*baseline*	gemma4-creole-mt	39.1	15.8
							KMT	*baseline*	kreyol-mt-pubtrain	37.1	12.2
							KMT	GMU	primary	37.0	13.5
							KMT	GMU	contrastive1	35.1	9.7
eng→jam	KMT	*baseline*	gemma4-creole-mt	36.1	18.4	**fra→hat**	KMT	*baseline*	gemma4-creole-mt	31.3	10.1
	KMT	GMU	primary	21.6	4.5		KMT	*baseline*	kreyol-mt-pubtrain	31.1	10.0
	KMT	GMU	contrastive1	19.9	3.9		KMT	CHV	contrastive1	24.8	4.3
	KMT	*baseline*	kreyol-mt-pubtrain	17.1	2.4		KMT	CHV	primary	24.7	4.1
							KMT	CHV	contrastive2	19.4	1.8
eng→kri	KMT	*baseline*	gemma4-creole-mt	34.6	17.4	**fra→pap**	KMT	*baseline*	gemma4-creole-mt	35.2	14.0
	KMT	GMU	contrastive1	14.6	2.2		KMT	*baseline*	kreyol-mt-pubtrain	34.8	12.4
	KMT	GMU	primary	12.9	2.1						
	KMT	*baseline*	kreyol-mt-pubtrain	11.4	1.4						

Table 10: Evaluation on *Eval Set 2*, XX → Creole.