AFRICAN SUBSTRATES RATHER THAN EUROPEAN LEXIFIERS TO AUGMENT AFRICAN-DIASPORA CREOLE TRANSLATION

Nathaniel R. Robinson
Carnegie Mellon University
nrrobins@cs.cmu.edu

Matthew D. Stutzman
Brigham Young University
mstutz1@byu.edu

Stephen D. Richardson
Brigham Young University
srichardson@cs.byu.edu

David R. Mortensen
Carnegie Mellon University
dmortens@cs.cmu.edu

ABSTRACT

Machine translation (MT) model training is difficult for low-resource languages. This is especially true for African-diaspora Creole languages because of data scarcity. Cross-lingual data augmentation methods with knowledge transfer from related high-resource languages are a common technique to overcome this disadvantage. For instance, practitioners may transfer knowledge from a language in the same language family as the low-resource language of interest. African-diaspora Creole languages are low-resource and simultaneously have relationships with multiple language groups. These languages, such as Haitian Creole and Jamaican Patois, are typically lexified by colonial European languages, but they are structurally similar to African languages. We explore the advantages of transferring knowledge from the European lexifier language versus the phylogenetic and typological relatives of the African substrate languages. We analysed Haitian and Jamaican MT: both controlling tightly for data properties across compared transfer languages and later allowing use of all data we collected. Our inquiry demonstrates a significant advantage in using African transfer languages in some settings.

1 INTRODUCTION

African and African-diaspora Creole languages are spoken natively by millions of people in multiple countries, including Haiti, Jamaica, the Bahamas, Trinidad and Tobago, Dominica, Cape Verde, Guinea-Bissau, the Seychelles, Mauritius, and parts of the United States, France, and the Netherlands in the Gulf of Mexico, Caribbean, and Indian Ocean (Rickford & McWhorter, 1997; Mufwene, 2006; Bartens, 2021; Velupillai, 2015). The need for language technologies, including MT, for these languages is acute due to frequent natural disasters and extensive emigration (Heinzelman & Waters, 2010; Rasmussen et al., 2015; Audebert, 2017). For example, earthquakes struck Haiti in 2021 and 2010, prompting humanitarian responses aided by language technologies (Margesson & Taft-Morales, 2010).

These Creole languages emerged among enslaved people of African descent who were forcibly brought to the Caribbean and elsewhere (primarily by French, Dutch, and English slavers). African-diaspora Creole languages are generally similar in typology but differ in their lexicons, which are largely based on European languages (Rickford & McWhorter, 1997). There is a longstanding debate regarding whether these languages are best understood as African languages with European words (Lefebvre, 2011; Rickford & McWhorter, 1997) or as new European languages that developed among enslaved African people in various regions (Faine, 1937; Hall, 1958; Mufwene, 2006). This scientific debate has a bearing on an engineering question: in training data-hungry NLP models, is the limited data from Creole languages best augmented with data from the European languages from which they inherited most of their words ("lexifier" or "superstrate" languages) or from Niger-Congo languages, from which they are often claimed to inherit their grammatical structure ("substrate" languages)?

We designed a set of controlled experiments to address both the scientific question (are African-diaspora Creoles typologically influenced by Niger-Congo languages, to the point of there existing a useful relationship for MT?) and the practical question (what languages should be used to augment African-diaspora Creole data in training MT models?). The languages for which we draw data are as follows:

hat *Haitian Creole* French-lexified creole with mixed—largely Gbe and **Kwa**—substrate [Haiti] (Lefebvre, 2011; Brousseau, 2011; Seguin, 2020)

jam *Jamaican Creole* English-lexified creole with Gbe, **Kwa**, **Igboid**, Yoruboid, and Bantu substrate [Jamaica] (Mufwene, 2002; Kouwenberg, 2008; Farquharson, 2012)

ibo *Igbo* Benue-Congo, Igboid [Nigeria] (Hammarström et al., 2023)

fat *Fante* Benue-Congo, Kwa [Ghana] (Hammarström et al., 2023)

twi *Twi* Benue-Congo, Kwa [Ghana] (Hammarström et al., 2023)

fra *French* Indo-European, Romance [France] (Hammarström et al., 2023)

eng *English* Indo-European, Germanic [England] (Hammarström et al., 2023)

Our experiments are based on bilingual machine translation: MT models trained on both source-target and pivot-target parallel data (where SOURCE is a low resource language and PIVOT is a related high-resource language). Various techniques for this type of training have been attempted. (See § 2.) We use the simplest and—for our data—the most effective: mixing the two types of parallel data and training on the combined set.

We used two source languages (`hat` and `jam`), two European pivot languages (`fra` and `eng`), and three West African pivot languages (`ibo`, `fat`, and `twi`). For `hat` MT we used `eng` as target and `fra` as a pivot language, for `jam` we used `fra` as target and `eng` as a pivot language, and we used all three African pivot languages for both. We could not include a West African language from the Gbe group, such as Fongbe, due to data scarcity.[1] In our main experiments, we trim datasets to ensure that pivot data are all translations of precisely the same target language text, meaning domain and topic are tightly controlled. The experiments consist of keeping the source and target languages constant and varying the pivot languages. We find that West African pivots outperform European pivots for `hat`→`eng` and nearly achieve parity for `jam`→`fra`. However, the African pivots that perform best do not follow precisely from the putative substrates. For example, the best pivot for `hat`→`eng` was `ibo`, even though one might expect Kwa languages `fat` or `twi`. Notably, since all of the relevant West African languages are similar in their typology, the effectiveness of augmentation with a particular language may not be a simple function of phylogenetic proximity.

We also experimented using all data available to us for all source, target, and pivot languages (rather than constricting augmentation sets to be translations of same target sentences). These secondary experiments, importantly, do not control for the dataset characteristics external to the pivot language choice, but they address more realistic settings for Creole language MT in practice.

The goal of these experiments is not to beat an existing benchmark, but to provide a systematic way of comparing the two augmentation strategies. Notably, to accurately compare with previous benchmarks would require use of previously published test sets. However, we note that our `hat`→`eng` model does achieve higher MT performance scores than any previously reported model.

The comparison of transfer learning via European lexifier languages and African substrate languages for Creole language processing has not been studied previously. In summary we contribute:

- A novel comparison of European lexifier languages to African substrate languages for cross-lingual transfer to African-diaspora Creole language MT
- Evidence that African substrate languages have potential in improving Creole language MT
- A framework for high-quality in-domain Haitian MT, with best performance resulting from African language transfer learning

[1]Seguin (2020) and Kouwenberg (2008) consider Fongbe as Haitian's substrate in their linguistic analysis.

2 Related Work

We are not the first researchers to investigate pivot and source languages for MT. Cross-lingual transfer learning is often accomplished through parent-child fine tuning, where parameters learned during the training of a high-resource language pair are used to initialize subsequent training of a low-resource pair (Zoph et al., 2016). This approach has been extended to incorporate cross-lingual embeddings (Kim et al., 2019) and additional encoders conditioned on bilingual shared latent space (Xing et al., 2022). Another method of transfer learning is via multilingual back-translation (Sennrich et al., 2016). Xia et al. (2019) generalized this approach with a two-step pivoting method to convert high-resource language data to a low-resource language.

The scope of our work focuses on a third technique: multilingual training. This is the technique of including data from one or more transfer languages as pivot languages in training (Chen et al., 2019). These pivot languages can serve as sources or targets. Their influence can improve translation for a low-resource source-target pair. For example, Gu et al. (2018) implemented the sharing of lexical- and sentence-level representations across multiple high-resource source languages with one low-resource target. The simplest realization of multilingual training is bilingual source training, where one pivot language is introduced as an additional source. Neubig & Hu (2018) explored a setting similar to ours: improving augmentation for low-resource languages by transferring knowledge from related high-resource languages, one corresponding to each low-resource language. Their results show that simple bilingual source training achieved comparable improvements to more involved methods, including larger-scale multilingual training and fine-tuning.

Robinson et al. (2022) explored multilingual back-translation and bilingual source training as cross-lingual transfer methods for low-resource Creole language MT (a setting even more pertinent to our work). They concluded that bilingual source training was preferred over back-translation. We choose bilingual source training for cross-lingual transfer to compare pivot languages. These motivating examples provide evidence that it is reliable. It is also simple with few hyperparameters, making it ideal for controlled experiments.

The research is scarce, but we are also not the first to explore MT for African-diaspora Creole languages (Lent et al., 2022b). Frederking et al. (1998) developed statistical Haitian MT systems. Following Haiti's 2010 earthquake, the 2011 Workshop on Statistical Machine Translation (WMT) featured a shared task to translate Haitian SMS messages for disaster relief (Callison-Burch et al., 2011). In addition to the participants of this challenge, later researchers have used WMT SMS data for Haitian MT research (Stymne, 2012; Sennrich, 2012; Dholakia & Sarkar, 2014). Though the majority of Creole MT work has focused on Haitian, there has been some investigation into Mauritian Creole and West African Pidgin MT (Dabre & Sukhoo, 2022; Dabre et al., 2014; Ogueji & Ahia, 2019). Robinson et al. (2022) explored MT for both Haitian and Jamaican. Their work is particularly motivating for us because of their insights that European languages may not be good candidate pivot languages for the Creoles they lexify, due to significant syntactic and orthographic differences. Some researchers have also explored ancestral cross-lingual transfer for Creole languages (Lent et al., 2022a), but our work differs in that we compare African substrates to European lexifiers. No prior work has considered African substrates for cross-lingual transfer in Creole MT.

3 Methodology and Experiments

As noted in § 1, we conducted two sets of experiments with Haitian-to-English MT (using French as a lexifier pivot language) and Jamaican-to-French MT (using English as pivot), with the same set of Benue-Congo substrate pivot languages for both. The first experiment set focused on comparing choice of transfer language as an isolated variable. The second set used full data sets available in a way that compares more factors at once than language choice, but is more realisticly applicable.

3.1 Data and Training Specifications

Our Jamaican bitext was collected from publicly available Jamaican[2] and English[3] New Testament translations. All our other data come from entries in the translation memories created by translators

[2] https://chop.bible.com/bible/
[3] https://ebible.org/download.php

for the Church of Jesus Christ of Latter-day Saints over the past several years. Data availability motivated our choices of languages to compare. Translations of the same English source sentences into multiple languages may occur frequently in the Church's data, which includes discourses, scripture, instructional materials, magazine articles, and various other publications. While a majority of the content is religious; humanitarian, welfare, genealogical, administrative, and other contents are included. Since the data is focused on the Church's domains of interest, MT models produced by our experiments may not generalize to a wide range of MT tasks. Efforts are underway to make the Church's data available in over 90 languages to the larger research community, but its general use is currently restricted.

We cleaned our data according to the Best Practices in Translation Memory Management[4] established by the GILT Leaders Forum. We replaced and normalized spurious strings, standardized quotation marks across languages, normalized white space, and removed control and escape characters. We removed entire segments that were empty; contained unbalanced brackets, quote marks, or parentheses; contained identical source and target utterances; contained fewer than 3 words on either source or target side; or contained a target more than 1.3x the length of the source or vice versa. We removed special characters such as copyright, trademark, and registered symbols and bullet points. We will release our data cleaning software upon publication.

We used Google's open-source SentencePiece encoder (Kudo & Richardson, 2018) to byte-pair encode (BPE) both source and target for all datasets. In each experiment we trained one BPE model for each language. BPE specifications for each experiment are described in §3.2 and §3.3. For all experiments we used a BPE vocabulary size of 12000 for French, English, Igbo, Fante, and Twi. Because Haitian is isolating, we opted for a smaller BPE vocabulary size of 8000. Our Jamaican dataset was too small to accommodate a vocabulary size larger than 6466, which we selected. In all experiments we decoded outputs prior to evaluation.

We used the same test sets for Haitian→English MT and Jamaican→French MT for all experiments presented in this paper. We ensured that the target side of each augmentation bitext excluded any sentences that appeared in the target side of the test set or the validation set, and we removed any test segments that appeared in training sets even after cleaning, for all experiments.

We used OpenNMT's (Klein et al., 2017) open-source training architecture[5] to train Transformer models for MT (Vaswani et al., 2017). The models for our main experiments were trained on a single 2080Ti GPU, and those for full-data experiments on a single A6000 GPU. Our hyperparameters were determined by recommended settings from OpenNMT[6] with some adjustments: we set `world_size` to 1 and `gpu_ranks` to [0] and used 1500 validation steps and 3000 save steps. For convenience we reiterate some specifications here: we used the Adam optimizer (Kingma & Ba, 2017) with 8000 warm-up steps and the NOAM decay method (Popel & Bojar, 2018). Our models had 6 encoding layers and 6 decoding layers with embedding dimension 512.

3.2 Comparison of Transfer Languages

For our strictly controlled experiment on Haitian MT we used equivalent augmentation sentences for French, Igbo, Twi and Fante. Each of these sets contained translations of a shared set of 18821 English sentences. (I.e. the target halves of the bitexts were identical.) The data setup of these experiments is illustrated in Figure 1. We followed an equivalent process for Jamaican MT. Precise data amounts are under "constrained data" in Table 1.

The identical-target data sets we used for these experiments were subsets of the full data available. (See "full data" in Table 1.) In Haitian MT experiments we also reduced the size of the authentic data. It would likely not be practical or enlightening to augment a bitext of 178485 segments with a small augmentation set of 18821 sentences. Haitian is low-resource, but it is exceptionally high-resource for a Caribbean Creole language. We reduced the authentic train set to 5500 sentences, selected randomly from the total set of 176485, to (1) increase the practicality of augmentation and (2) be more representative of data availability for more prototypical African-diaspora Creole languages. (We selected 5500 because it is roughly equal to the size of our full Jamaican-French bi-

[4]https://github.com/GILT-Forum/TM-Mgmt-Best-Practices/blob/master/best-practices.md
[5]https://github.com/OpenNMT/OpenNMT-py
[6]See "How do I train the Transformer model?"

text.) In experiments with the 5500-segment Haitian MT training set, we also reduced the validation set size to 500, to reflect the realistic condition that fewer training data imply fewer validation data in realistic settings.

For both Haitian MT and Jamaican MT experiments, we trained each augmentation language BPE model on the ∼19K sentences from its respective augmentation set. We also trained the source language BPE on the corresponding side of the train set, to reflect realistic low-resource settings. We relied on more robust BPE models for target languages, however. Relying on the assumption that most MT practioners would seek a large monolingual English or French corpus for BPE training, even in low-resource scenarios, we used the English BPE model from our full-data experiments for Haitian→English translation and the French BPE model from our full-data experiments for Jamaican→French. (We reiterate that we only gave English or French this encoding advantage as target languages, giving them no advantage over their African counterparts when used as pivot languages.)

To accurately compare transfer language choice, we trained all models for 27000 steps after observing empirically that the validation accuracy tended to plateau for all models after this point.

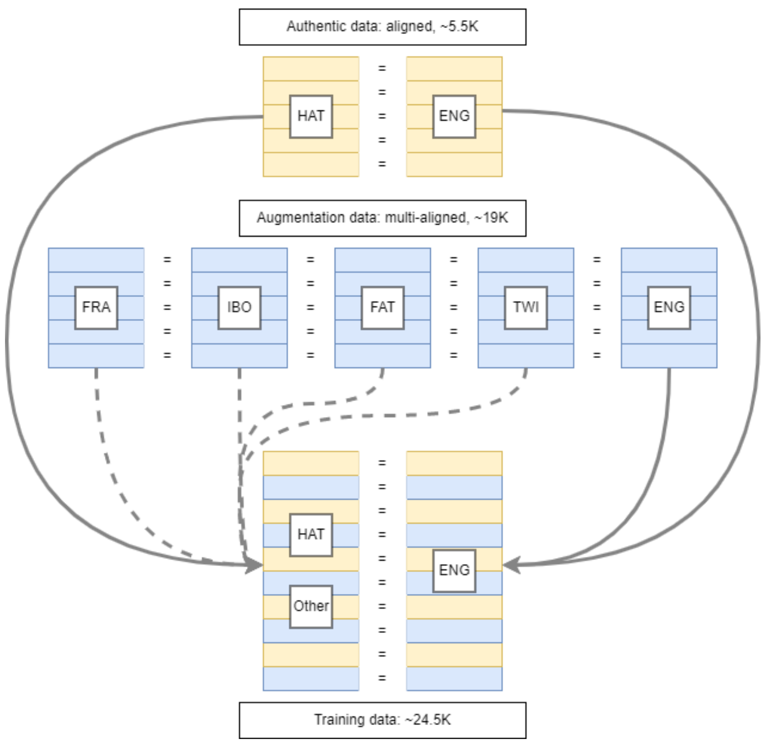

Figure 1: Illustration of our data configuration for scientific transfer language comparison for HAT→ENG translation. Solid arrows indicate subsets that are always included in the train set; dashed arrows indicate that one or another of the different subsets will be used.

3.3 Experiments with Full Data Sets

Although we constrained data sets artificially for the integrity of pivot language comparison, we acknowledge that most MT practitioners would use all data available. To represent this scenario, we ran a second slate of experiments on full data sets. Information for these full sets is displayed under "full data" in Table 1. These experiments partially address a concern: although related African languages may be a better fit for knowledge transfer to a Creole language than the language's European lexifier, the European lexifier may often be a more practical choice, simply because European languages tend to have more abundant data than African languages. Evaluating results over full, unequal data sets partially addresses this concern (though experimentation with other cross-lingual transfer methods

Table 1: Data amounts for transfer language comparison, as well as full authentic and augmentation amounts. Length is measured in aligned sentence pairs. The sets presented under "constrained data" represent subsets of the sets under "full data."

bitext	**use**	**constrained data**	**full data**
HAT-ENG	HAT→ENG training (authentic)	5500	176485
HAT-ENG	HAT→ENG validation	500	2500
HAT-ENG	HAT→ENG test	2212	2212
FRA-ENG	HAT→ENG training augmentation	18821	835003
IBO-ENG	HAT→ENG training augmentation	18821	47298
FAT-ENG	HAT→ENG training augmentation	18821	55743
TWI-ENG	HAT→ENG training augmentation	18821	34351
JAM-FRA	JAM→FRA training (authentic)	4128	4128
JAM-FRA	JAM→FRA validation	500	500
JAM-FRA	JAM→FRA test	500	500
ENG-FRA	JAM→FRA training augmentation	18816	773314
IBO-FRA	JAM→FRA training augmentation	18816	33456
FAT-FRA	JAM→FRA training augmentation	18816	38554
TWI-FRA	JAM→FRA training augmentation	18816	23518

would address it more fully). Additionally, since we used the full authentic Haitian-English bitext, full-data experiments also display results for the practical application of Haitian MT.

In these full-data experiments we trained each language's BPE model on the full data from its corresponding set used for training (and not for validation or testing). Noting that English and French appear in multiple of the bitexts, we specify that we trained English and French BPE models on both sides of the English-French bitext consisting of 835003 segments.

For full-data Haitian MT experiments, we selected models to test by adhering to a stopping criterion with patience 2. Per our configuration, models were saved every 2 validation steps. Once validation accuracy did not increase twice in a row, we reverted back to the most recent saved model (either 1 or 2 validation steps earlier) for testing. In the case of Jamaican MT, we observed that validation accuracy oscillated more before convergence, and we accordingly used a patience value of 3. In this case, once the validation accuracy did not increase 3 times in a row, we reverted back to the model saved 2 or 3 validation steps earlier.

4 RESULTS AND ANALYSIS

We evaluated translation results using three metrics: BLEU (Papineni et al., 2002), chrF++ Popović (2017), and COMET Rei et al. (2020). Both BLEU and chrF++ were calculated with the `sacrebleu`[7] API, using `chrf-word-order` 2. To determine statistical significance of result improvements, we used `sacrebleu`'s built-in paired bootstrapping comparison for BLEU and chrF++ and COMET's `comet-compare` command line tool. We predetermined a threshold of $p = 0.05$ to reject null hypotheses. If one value is **bold** in a results table column, it means that that result was statistically significantly better than all others in its comparison group. If two or more values are **bold**, it means that they are all statistically significantly better than the next best score but not than each other. If no values are **bold**, no such relationship exists.

The results of our controlled experiments are in Table 2. As noted in § 3.2, we trained these models for 27000 steps. Igbo was best pivot language for Haitian MT, despite the general consensus of Haitian's hypothesized closer tie to Kwa languages Twi and Fante. (See § 1.) The best-performing pivot language for Jamaican was English, with Igbo nearly on par. As noted in § 1, Jamaican has hypothesized relationships with all the African languages included.

[7]https://github.com/mjpost/sacrebleu

Table 2: Comparison of augmentation language in our scientific language comparison experiments. Statistically significant best results are **bold**. These results were derived from training data described under "constrained data" in Table 1.

src→tgt	aug lang	BLEU	chrF++	COMET
HAT→ENG	none	19.4	39.2	-0.642
HAT→ENG	FRA (lexifier)	29.8	49.3	-0.317
HAT→ENG	IBO (substrate)	**31.4**	**50.3**	**-0.271**
HAT→ENG	FAT (substrate)	29.8	49.3	-0.315
HAT→ENG	TWI (substrate)	29.5	49.1	-0.324
JAM→FRA	none	10.4	26.1	-1.154
JAM→FRA	ENG (lexifier)	**16.5**	**36.5**	**-0.952**
JAM→FRA	IBO (substrate)	**16.3**	35.3	-0.985
JAM→FRA	FAT (substrate)	14.8	34.4	-1.026
JAM→FRA	TWI (substrate)	15.3	34.0	-1.017

Our additional experiments using full datasets are in Table 3. In these experiments Fante performed generally the best for Haitian, and Igbo for Jamaican, though trends are not outstanding. Note that Haitian baseline performance is strong enough with full data that ordinary bilingual source training is not helpful for all pivot languages. (It is harmful with French.) Interestingly, English augmentation for Jamaican yielded a low BLEU score but high chrF++ and COMET scores. This can perhaps be explained by properties of our data. The French sentences we used were translated from original English sources (the same sentences we used for English augmentation), so their word order and structures correspond to the original English source structures in a predictable way. Jamaican sentences have very different word order from English sentences. So it is conceivable that the model learned word correspondences that do not hold for Jamaican-French translations, like those in the test set. As a result, many of the model's translations from Jamaican are good paraphrases of the French targets, but with different word ordering. Due to its calculation process, BLEU is not forgiving of word order variations even to paraphrase (Papineni et al., 2002). This issue could potentially be mitigated by further data balancing or fine-tuning. See § 5. It also highlights a structural advantage that the African languages have in augmenting Jamaican MT.

Our results considered in aggregate clearly demonstrate that West African languages with theoretical ties to African-diaspora Creole languages do have potential to transfer useful knowledge in MT. In some cases they even outperform the European lexifier languages that are more frequently used in MT applications.

Importantly, the uniqueness of our test set means that our translation scores cannot be compared directly to previous work. Furthermore, the narrowness of the text domain in our data allows for frequently repeated phrases and sub-phrases, which may make our test set easier than others for MT. We draw attention, however, to the impressive performance of our Haitian→English model trained on full datasets and augmented with Fante data. (See Table 3.) The winners of the WMT 2011 shared task and later researchers, though using a test set unrelated to ours, reported BLEU scores of no more than 39.86 (Dholakia & Sarkar, 2014; Heafield & Lavie, 2011). More notably, Robinson et al. (2022) used a test set that, while not identical to ours, was composed almost entirely of data from the same source. Their work reported maximal Haitian BLEU scores of 46.88. Our model's performance, while not conclusive of direct improvement, encouragingly suggests our training configuration's suitability to Haitian MT for future work.

5 Conclusion and Future Work

Our results suggest that African-diaspora Creole languages are typologically influenced by Niger-Congo languages. In multiple settings, Haitian and Jamaican Creole benefited more from West African language augmentation than from their own European lexifiers for MT. Our results suggest that barring external factors such as dataset size and characteristics, Igbo is a good pivot language choice for Haitian MT, and English and Igbo are good choices for Jamaican. When we included full

Table 3: Experiments with full data sets. Statistically significant best results are **bold**. These results were derived from training data described under "full data" in Table 1.

src→tgt	aug lang	training steps	BLEU	chrF++	COMET
HAT→ENG	none	54000	67.3	**78.5**	0.5531
HAT→ENG	FRA (lexifier)	57000	59.3	73.4	0.5242
HAT→ENG	IBO (substrate)	63000	67.4	**78.3**	0.5456
HAT→ENG	FAT (substrate)	99000	**68.0**	**78.5**	**0.5576**
HAT→ENG	TWI (substrate)	48000	66.5	77.6	0.5413
JAM→FRA	none	75000	11.1	26.7	-1.149
JAM→FRA	ENG (lexifier)	18000	12.9	35.8	**-0.860**
JAM→FRA	IBO (substrate)	57000	**18.0**	37.3	**-0.865**
JAM→FRA	FAT (substrate)	42000	17.1	36.8	-0.899
JAM→FRA	TWI (substrate)	69000	16.1	35.4	-0.964

datasets, we achieved optimal overall performance for Jamaican by using Igbo and for Haitian by using Fante.

In addition to our experimental conclusions, we contribute practical training specifications combined with cross-lingual augmentation that yield notably high translation scores for low-resource Haitian MT, given the restricted domain of our dataset. We encourage and hope to participate in future efforts to apply these techniques to build more domain adaptable Haitian MT models and improve on previously established benchmarks.

Exploration of African substrate pivot languages can be continued. As noted in §2, we chose a relatively simple cross-lingual transfer method for optimal controlled comparison of pivot language choice. However, we remark that given the size of data available, further translation improvements can be made. Particularly in settings where large bitexts involving high-resource languages such as English and French are available and outnumber authentic data, performance may be improved by up-sampling the authentic bitext to avoid overfitting on augmentation data. Performance may also be reasonably improved by multilingual (as opposed to bilingual) training, and by separating training into a pre-training step and fine-tuning on authentic data. Pivot language choice may be explored in these settings to yield improvements over existing methods and provide further insights about language relationships.

We recognize that conclusions about the relationships between African-diaspora Creole languages and their African substrates cannot be concluded fully by MT experimentation. In the future these language pairs should be subjected to tests in other NLP applications, as well as more in-depth analysis and scrutiny regarding how their hypothesized shared structures benefit these applications.

We acknowledge the limited scope of this current work to the exploration of two African-diaspora Creole languages and three Niger-Congo languages. In the future we hope to collaborate with other researchers to expand this study to include other African languages such as Fongbe and Wolof and other African and African-diaspora Creoles such as Papiamento, Cape Verdean Creole, Louisiana Creole, and Seychellois Creole. We recognize that many of these languages have extremely scarce resources and hope to collaborate with data collection efforts to facilitate further discoveries.

ACKNOWLEDGMENTS

We acknowledge the contributions of Kalvin Chang and Yanlin Feng in assisting this work. This material is based on research sponsored in part by the Air Force Research Laboratory under agreement number FA8750-19-2-0200. The U.S. Government is authorized to reproduce and dis- tribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Air Force Research Laboratory or the U.S. Government.

REFERENCES

Cedric Audebert. The recent geodynamics of haitian migration in the americas: refugees or economic migrants? *Revista Brasileira de Estudos de População*, 34:55–71, 2017.

Angela Bartens. The making of languages and new literacies: San andrés-providence creole with a view on jamaican and haitian. *Lingüística y Literatura*, 42(79):237–256, 2021.

Anne-Marie Brousseau. One substrate, two creoles. In Claire Lefebvre (ed.), *Creoles, their Substrates, and Language Typology*, pp. 105–153. John Benjamins, Amsterdam, 2011.

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

Xilun Chen, Ahmed Hassan Awadallah, Hany Hassan, Wei Wang, and Claire Cardie. Multi-source cross-lingual model transfer: Learning what to share. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 3098–3112, Florence, Italy, July 2019. Association for Computational Linguistics. doi: 10.18653/v1/P19-1299. URL `https://aclanthology.org/P19-1299`.

Raj Dabre and Aneerav Sukhoo. Morisienmt: A dataset for mauritian creole machine translation. *arXiv e-prints*, pp. arXiv–2206, 2022.

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

Rohit Dholakia and Anoop Sarkar. Pivot-based triangulation for low-resource languages. In *Proceedings of the 11th Conference of the Association for Machine Translation in the Americas: MT Researchers Track*, pp. 315–328, 2014.

Jules Faine. *Philologie creole: Études historique et étymologiques sur la langue creoles d'Haïti*. Imprimerie de l'Etat, Port-au-Prince, 1937.

Joseph Farquharson. *The African lexis in Jamaican: Its linguistic and sociohistoircal significance*. PhD thesis, University of the West Indies, 2012.

Robert E Frederking, Ralf D Brown, and Christopher Hogan. The diplomat rapiddeployment speech mt system. *MT Summit (1997)*, pp. 261–262, 1998.

Jiatao Gu, Hany Hassan, Jacob Devlin, and Victor O.K. Li. Universal neural machine translation for extremely low resource languages. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pp. 344–354, New Orleans, Louisiana, June 2018. Association for Computational Linguistics. doi: 10.18653/v1/N18-1032. URL `https://aclanthology.org/N18-1032`.

Robert Hall. Creole languages and genetic relationships. *Word*, 14:367–373, 1958.

Harald Hammarström, Sebastian Nordhoff, Robert Forkel, Martin Haspelmath, and Sebastian Bank. Glottolog. Online database, 2023. URL `https://glottolog.org/`.

Kenneth Heafield and Alon Lavie. Cmu system combination in wmt 2011. In *Proceedings of the Sixth Workshop on Statistical Machine Translation*, pp. 145–151, 2011.

Jessica Heinzelman and Carol Waters. *Crowdsourcing crisis information in disaster-affected Haiti*. JSTOR, 2010.

Yunsu Kim, Yingbo Gao, and Hermann Ney. Effective cross-lingual transfer of neural machine translation models without shared vocabularies. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 1246–1257, Florence, Italy, July 2019. Association for Computational Linguistics. doi: 10.18653/v1/P19-1120. URL `https://aclanthology.org/P19-1120`.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017.

Guillaume Klein, Yoon Kim, Yuntian Deng, Jean Senellart, and Alexander Rush. OpenNMT: Open-source toolkit for neural machine translation. In *Proceedings of ACL 2017, System Demonstrations*, pp. 67–72, Vancouver, Canada, July 2017. Association for Computational Linguistics. URL https://www.aclweb.org/anthology/P17-4012.

Silvia Kouwenberg. The problem of multiple substrates: The case of Jamaican Creole. In Susanne Michaelis (ed.), *Roots of creole structures: Weighing the contribution of substrates and superstrates*, pp. 1–27. John Benjamins, 2008.

Taku Kudo and John Richardson. SentencePiece: A simple and language independent subword tokenizer and detokenizer for neural text processing. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pp. 66–71, Brussels, Belgium, November 2018. Association for Computational Linguistics. doi: 10.18653/v1/D18-2012. URL https://aclanthology.org/D18-2012.

Claire Lefebvre. Substrate features in the properties of verbs in three atlantic creoles: Haitian Creole, Saramaccan and Papiamentu. In Claire Lefebvre (ed.), *Creoles, their Substrates, and Language Typology*, pp. 127–154. John Benjamins, Amsterdam, 2011.

Heather Lent, Emanuele Bugliarello, and Anders Søgaard. Ancestor-to-creole transfer is not a walk in the park. In *Proceedings of the Third Workshop on Insights from Negative Results in NLP*, pp. 68–74, Dublin, Ireland, May 2022a. Association for Computational Linguistics. doi: 10.18653/v1/2022.insights-1.9. URL https://aclanthology.org/2022.insights-1.9.

Heather Lent, Kelechi Ogueji, Miryam de Lhoneux, Orevaoghene Ahia, and Anders Søgaard. What a creole wants, what a creole needs. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pp. 6439–6449, Marseille, France, June 2022b. European Language Resources Association. URL https://aclanthology.org/2022.lrec-1.691.

Rhoda Margesson and Maureen Taft-Morales. Haiti earthquake: Crisis and response. Library of Congress Washington DC Congressional Research Service, 2010.

Salikoko Mufwene. Socio-economic historical arguments for a gradual and heterogeneous development of patois in Jamaica. In *Biennial Meeting of Society for Caribbean Linguistics*, 2002.

Salikoko Mufwene. Pidgins and creoles. In Braj B. Kachru, Yamuna Kachru, and Cecil L. Nelson (eds.), *The Handbook of World Englishes*, pp. 313–327. Blackwell, Hoboken, NJ, 2006.

Graham Neubig and Junjie Hu. Rapid adaptation of neural machine translation to new languages. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 875–880, Brussels, Belgium, October-November 2018. Association for Computational Linguistics. doi: 10.18653/v1/D18-1103. URL https://aclanthology.org/D18-1103.

Kelechi Ogueji and Orevaoghene Ahia. Pidginunmt: Unsupervised neural machine translation from west african pidgin to english, 2019.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pp. 311–318, 2002.

Martin Popel and Ondř ej Bojar. Training tips for the transformer model. *The Prague Bulletin of Mathematical Linguistics*, 110(1):43–70, apr 2018. doi: 10.2478/pralin-2018-0002. URL https://doi.org/10.2478%2Fpralin-2018-0002.

Maja Popović. chrf++: words helping character n-grams. In *Proceedings of the second conference on machine translation*, pp. 612–618, 2017.

Andrew Rasmussen, Eddy Eustache, Giuseppe Raviola, Bonnie Kaiser, David J Grelotti, and Gary S Belkin. Development and validation of a haitian creole screening instrument for depression. *Transcultural psychiatry*, 52(1):33–57, 2015.

Ricardo Rei, Craig Stewart, Ana C Farinha, and Alon Lavie. COMET: A neural framework for MT evaluation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 2685–2702, Online, November 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.emnlp-main.213. URL `https://aclanthology.org/2020.emnlp-main.213`.

John R Rickford and John McWhorter. Language contact and language generation: Pidgins and creoles. In Florian Coulmas (ed.), *The handbook of sociolinguistics*, pp. 238–256. Blackwell, Malden, MA, 1997.

Nathaniel Robinson, Cameron Hogan, Nancy Fulda, and David R. Mortensen. 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)*, pp. 35–42, Gyeongju, Republic of Korea, October 2022. Association for Computational Linguistics. URL `https://aclanthology.org/2022.loresmt-1.5`.

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

Rico Sennrich. Perplexity minimization for translation model domain adaptation in statistical machine translation. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 539–549, 2012.

Rico Sennrich, Barry Haddow, and Alexandra Birch. Improving neural machine translation models with monolingual data. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 86–96, Berlin, Germany, August 2016. Association for Computational Linguistics. doi: 10.18653/v1/P16-1009. URL `https://aclanthology.org/P16-1009`.

Sara Stymne. Clustered word classes for preordering in statistical machine translation. In *Proceedings of the Joint Workshop on Unsupervised and Semi-Supervised Learning in NLP*, pp. 28–34, 2012.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Viveka Velupillai. Pidgins, creoles and mixed languages. *Pidgins, Creoles and Mixed Languages*, pp. 1–626, 2015.

Mengzhou Xia, Xiang Kong, Antonios Anastasopoulos, and Graham Neubig. Generalized data augmentation for low-resource translation. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 5786–5796, Florence, Italy, July 2019. Association for Computational Linguistics. doi: 10.18653/v1/P19-1579. URL `https://aclanthology.org/P19-1579`.

Xiaolin Xing, Yu Hong, Minhan Xu, Jianmin Yao, and Guodong Zhou. Taking actions separately: A bidirectionally-adaptive transfer learning method for low-resource neural machine translation. In *Proceedings of the 29th International Conference on Computational Linguistics*, pp. 4481–4491, Gyeongju, Republic of Korea, October 2022. International Committee on Computational Linguistics. URL `https://aclanthology.org/2022.coling-1.395`.

Barret Zoph, Deniz Yuret, Jonathan May, and Kevin Knight. Transfer learning for low-resource neural machine translation. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 1568–1575, Austin, Texas, November 2016. Association for Computational Linguistics. doi: 10.18653/v1/D16-1163. URL `https://aclanthology.org/D16-1163`.