informatics

Article

Automatic Translation Between Kreol Morisien and English Using the Marian Machine Translation Framework

Zaheenah Beebee Jameela Boodeea [1], Sameerchand Pudaruth [2,*], Nitish Chooramun [3] and Aneerav Sukhoo [4]

[1] Faculty of Information, Communication and Digital Technologies, University of Mauritius, Moka 80837, Mauritius; beebee.boodeea1@umail.uom.ac.mu
[2] ICT Department, Faculty of Information, Communication and Digital Technologies, University of Mauritius, Moka 80837, Mauritius
[3] Department of Computer Science and Digital Technologies, School of Architecture, Computing and Engineering, University of East London, London E16 2RD, UK; n.chooramun@uel.ac.uk
[4] Central Information Systems Division, Ministry of Information Technology, Communication and Innovation, Port Louis 11302, Mauritius; aneeravsukhoo@yahoo.com
* Correspondence: s.pudaruth@uom.ac.mu

Abstract: Kreol Morisien is a vibrant and expressive language that reflects the multicultural heritage of Mauritius. There are different versions of Kreol languages. While Kreol Morisien is spoken in Mauritius, Kreol Rodrige is spoken only in Rodrigues, and they are distinct languages. Being spoken by only about 1.5 million speakers in the world, Kreol Morisien falls in the category of under-resourced languages. Initially, Kreol Morisien lacked a formalised writing system, with many people using different spellings for the same words. The first step towards standardisation of writing Kreol Morisien was after the publication of the Kreol Morisien orthography in 2011 and Kreol Morisien grammar in 2012 by the Kreol Morisien Academy. Kreol Morisien obtained a national position in the year 2012 when it was introduced in educational organisations. This was a major breakthrough for Kreol Morisien to be recognised as a national language on the same level as English, French, and other oriental languages. By providing a means for Kreol Morisien speakers to connect with others, a translation system will help to preserve and strengthen the identity of the language and its speakers in an increasingly globalized world. The aim of this paper is to develop a translation system for Kreol Morisien and English. Thus, a dataset consisting of 50,000 parallel Kreol Morisien and English sentences was created, where 48,000 sentence pairs were used to train the models, while 1000 sentences were used for evaluation and another 1000 sentences were used for testing. Several machine translation systems such as statistical machine translation, open-source neural machine translation, a Transformer model with attention mechanism, and Marian machine translation are trained and evaluated. Our best model, using MarianMT, achieved a BLEU score of 0.62 for the translation of English to Kreol Morisien and a BLEU score of 0.58 for the translation of Kreol Morisien into English. To our knowledge, these are the highest BLEU scores that are available in the literature for this language pair. A high-quality translation tool for Kreol Morisien will facilitate its integration into digital platforms. This will make previously inaccessible knowledge more accessible, as the information can now be translated into the mother tongue of most Mauritians with reasonable accuracy.

Keywords: Kreol Morisien; English; MarianMT; BLEU score; parallel sentences

Academic Editor: Zhiwen Yu

Received: 11 December 2024
Revised: 2 February 2025
Accepted: 8 February 2025
Published: 10 February 2025

Citation: Boodeea, Z.B.J.; Pudaruth, S.; Chooramun, N.; Sukhoo, A. Automatic Translation Between Kreol Morisien and English Using the Marian Machine Translation Framework. *Informatics* **2025**, *12*, 16. https://doi.org/10.3390/informatics12010016

1. Introduction

In modern linguistics, machine translation is a revolutionary force that makes cross-cultural communication possible. In the context of Kreol Morisien and English, it serves as a vital tool for preserving the rich cultural heritage of Kreol Morisien while enhancing global accessibility to this unique language. By using machine translation with special focus on neural machine translation (NMT), we can break down language barriers, democratise knowledge, and promote the visibility of underrepresented languages on a global stage [1].

Given the varied language heritage and oral tradition, Kreol Morisien frequently has trouble being correctly represented in written or translated forms. We can ensure the preservation of the linguistic and cultural peculiarities of Kreol Morisien while also enabling smoother, more accurate translations between Kreol Morisien and English for both professional and educational situations by using the more advanced capabilities of NMT. This promotes increased global awareness and understanding of this lively language, in addition to empowering Kreol Morisien speakers.

Children having a substantial knowledge in their mother tongue often have a better insight of their place in the society by building pride in their own culture, along with improved feelings of comfort and confidence. This flows down into all parts of their lives, which includes their academic accomplishment [2]. Cummins (2001) found that it is easier for children to learn a second language and make progress in critical thinking and literacy skills if they master their mother tongue language [3].

Kreol Morisien's presence in the National Assembly emphasises how necessary it is to learn the language. Furthermore, it has been acknowledged that Kreol Morisien requires specialised software [4]. Online access to Kreol Morisien's live and recorded Parliament sessions provides materials for a global audience [5]. The intention of using Kreol Morisien in parliamentary talks is to remove linguistic barriers so that listeners of all language proficiency levels can understand and interact with the recordings. Thus, making Kreol Morisien more approachable and inclusive for all is the ultimate objective.

Kreol Morisien is neither a dialect nor a patois. It is now considered as a language in its own and has great value since it is used by Mauritians in everyday life. The Akademi Kreol Repiblik Moris (AKRM) was set up in 2019 with the aim of developing further the orthography, language rules, vocabulary, and norms of Kreol Repiblik Moris as the national language of the Republic of Mauritius [6]. In a newspaper article on 19 February 2021, the president of AKRM acknowledged that they have not made any progress for one year and one of the main reasons was due to the lack of a translation tool for the Kreol Morisien language [7]. On the 21st of February, we celebrated the International Mother Language Day to promote awareness of lingualism and cultural diversity. Given that the number of formal users of Kreol Morisien is steadily increasing, this illustrates how an effective translation tool for the language will undoubtedly help to establish Kreol Morisien as a complete and official language.

In this paper, the role and significance of Kreol Morisien are first explored. We will then review the existing Kreol Morisien machine translation systems. This paper will then outline the process of creating an enhanced dataset. Following this, the evaluation and the performance of the Moses tool, OpenNMT system, Transformer model, and MarianMT, as well as a comparison with ChatGPT and Google Translate, are performed.

2. Kreol Morisien

The fight for Kreol Morisien's recognition in Mauritius has been difficult for the language's proponents. Since Kreol Morisien was the mother tongue of many Mauritians, the adoption of it as the country's official language was promoted when Mauritius gained its independence [8]. Additionally, a number of literary works, including Shakespearian plays (e.g., Macbeth) and poems, were translated into Kreol Morisien [8,9]. In 1976, a number of politicians got involved by advocating the usage of Kreol Morisien in both the National Assembly and the media [10]. Beginning in 1982, another political party made a valiant effort to communicate both verbally and in writing in Kreol Morisien [11]. The Ledikasyon Pu Travayer movement published the "Mauritian Creole to English dictionary" in 1984 followed by another one with the translation to French in 1987 [12]. These opened avenues for further developments. The aim was to safeguard the preservation of the native language of Mauritius.

Mauritius has made praiseworthy efforts in conserving Kreol Morisien after a long history where the language has been making its way for being recognised officially. There have been many further efforts to widen the effect of Kreol Morisien in Mauritius. The first step towards standardisation of writing Kreol Morisien was the setting up of the technical committee in April 2004 with the objective of coming forward with a consistent writing system for Kreol Morisien. This led to the publication of "Grafi-larmoni" in September 2004 [13]. Hookoomsingh (2004) connected "Grafi-larmoni" to a unified orthography that made the language easier for people to understand [13]. The year 2010 marked the introduction of the committee Akademi Kreol Morisien, the purpose of which was to work on the techniques for teaching Kreol Morisien in primary schools. As a result, "Lortograf Kreol Morisien" and "Gramer Kreol Morisien" were established in 2011 and 2012, respectively, offering a recognised and proper method of writing Kreol Morisien [14].

Kreol Morisien eventually obtained a national position in the year 2012 when it was introduced in all primary schools of the Republic of Mauritius as an optional subject [15]. This was a major breakthrough for Kreol Morisien to be recognised as a national language. In 2014, the University of Mauritius introduced its first BA (Hons) French and Creole programme. The introduction of Kreol Morisien at the secondary level was made in 2018 for the pupils who passed in Kreol Morisien for the Primary School Achievement Certificate examination [15]. One thousand two hundred twenty students participated in the Kreol Morisien NCE (National Certificate of Education) exam in 2021 [16]. In 2023, students took the first Kreol Morisien exams at School Certificate (SC) level. These exams are conducted in Kreol Morisien by the University of Mauritius and the Mauritius Examinations Syndicate [17].

The taboo against Kreol Morisien language being a language only for the illiterates is slowly being eliminated. In the past, marketing in Kreol Morisien was considered rude, but it is now rapidly becoming the norm. Currently, the media in Mauritius, including both traditional and digital, offer information and entertainment through Kreol Morisien language. "Zournal Kreol" is a news announcement in Kreol Morisien currently aired daily by the Mauritius Broadcasting Corporation (MBC). Additionally, "Senn Kreol" is a TV station owned by the MBC which broadcasts gastronomic shows and documentaries in Kreol Morisien. Several private radios such as Radio Plus started to use Kreol Morisien to communicate information to the general public. Even certain press articles such as "Anou koz parol" are published in Kreol Morisien.

Moreover, there are practically well-written Kreol Morisien texts in various places which comprise posters, banners and leaflets even within official documents from the government. There is a vast number of advertisements and messages being broadcasted in Kreol Morisien even by large and reputed organisations such as banks and insurance companies. Also, several public bodies such as the Central Electricity Board (CEB), the National Transport Authority, hospitals, the Central Water Authority (CWA), and the Traffic Management and Road Safety Unit raise awareness for the general community through campaigns in Kreol Morisien.

The evolution of Kreol Morisien, from being previously disregarded to gaining widespread acceptance in various spheres, demonstrates its growing importance. As Kreol Morisien continues to gain importance, a translation system will facilitate its incorporation into digital platforms.

3. Existing Kreol Morisien Machine Translation Systems

Using a deep learning method based on the Transformer model, Pudaruth et al. (2021) developed the first user-friendly automatic online translation system between Kreol Morisien and English [18]. Access to this ground-breaking translation technique was made possible by both the Morisia app 1.0 on the Google Play Store 26.4.21 and the corresponding website. The system used a dataset of 24,810 sentence pairs to build the translation models. One thousand unseen sentences were used to test the trained models. The BLEU score was used to assess the translations. The translation from Kreol Morisien to English scored 30.30, whereas the translation from English to Kreol Morisien scored 26.34.

Morisien MT 1.0 was developed by Dabre and Sukhoo [19]. They established a variety of baseline models using their own parallel corpora. The data for MorisienMT was created manually, specifically through books available in English translated to Kreol Morisien. The dataset contains 23,310 parallel sentences belonging to a mix of domains. They achieved a BLEU score of 22.9 for the translation of KM to English and a BLEU score of 22.6 for the translation of English to KM. The training data contain many dictionary entries but the evaluation data contain only full sentences. This showed that there is a possibility of leveraging dictionaries and pre-trained models for higher quality translation for Kreol Morisien.

Morisia 2.0 was developed by Pudaruth et al. [20] for the translation of Kreol Morisien to English and vice versa. A machine translation system using a Transformer model with a dataset of 50,000 parallel sentences has been developed. A BLEU score of 31.46 for translating from Kreol Morisien to English and 28.15 for translating from English to Kreol Morisien were achieved. This shows an improvement from the previous score in Morisia 1.0. The authors show that an expanded dataset has enhanced the translation quality.

4. Creation of the Dataset

We collected sentences from the internet and manually translated them. We have created two datasets (Dataset 1 and Dataset 2) for the Kreol Morisien and English translation models. Different numbers of sentences are used in the two datasets for training. Dataset 1 is a subset of Dataset 2. However, Datasets 1 and 2 both make use of identical evaluation and testing sets. The dataset used in this paper is completely different from the dataset used in Morisia 1.0 and 2.0.

4.1. Dataset 1

Dataset 1 consists of a total of 26,000 sentence pairs. The first 24,000 sentences were used for training the English to Kreol Morisien model and vice versa, and 1000 sentence pairs were used for validating the models. Another set of 1000 sentence pairs were used to test the trained models. The analysis of the training dataset is displayed in Table 1.

The third edition of Diksioner Morisien, which contains 19,400 original words, was released in 2019 [21]. Table 1 shows that the dataset contains 10,091 unique Kreol Morisien words, or about 52% of all the words in Kreol Morisien. However, only 10,084 of the more than 100,000 words in the English language were included in this dataset.

Table 1. Analysis of the training Dataset 1.

Statistics	English	Kreol Morisien
Number of sentences	24,000	24,000
Number of words overall	146,232	143,737
Total number of unique words	10,084	10,091
Length of shortest sentence	2	1
Shortest sentence	And besides	Kontan
Length of longest sentence	24	26
Longest sentence	Part of the charm of a big city lies in the variety of styles that can be seen in the architecture of its buildings.	Sa legliz devan lekel nou finn pase la finn kraze sink minit apre akoz enn tranbleman-de-ter ek plis ki san relizie finn antere vivan.
Average word count in a sentence	7.4	7.3

4.2. Dataset 2

Briefly, 48,000 sentence pairs were used for training out of the 50,000 sentence pairs used for Dataset 2. Dataset 1 is fully contained in Dataset 2. The size of the training set has doubled from Dataset 1 to Dataset 2. The same validation and testing sets were fed into Datasets 1 and 2 in order to evaluate the performance of the models. Table 2 shows the analysis of the dataset used in training.

Table 2. Analysis of the training Dataset 2.

Statistics	English	Kreol Morisien
Number of sentences	48,000	48,000
Number of words overall	321,687	305,708
Total number of unique words	13,326	16,734
Length of shortest sentence	2	1
Shortest sentence	And besides	Kontan
Length of longest sentence	26	26
Longest sentence	A qipao in the traditional style is made entirely by hand and requires labour-intensive manual tailoring techniques to enhance and exhibit every unique female form.	Sa legliz devan lekel nou finn pase la finn kraze sink minit apre akoz enn tranbleman-de-ter ek plis ki san relizie finn antere vivan.
Average word count in a sentence	7.8	7.4

Table 2 shows that the dataset consists of 16,734 unique Kreol Morisien words. This is about 86% of all the words in Kreol Morisien. However, only 13,326 unique English words were present in this dataset.

5. Methodology

Machine translation technology has made noteworthy advances in recent years, as deep learning has been integrated with natural language processing [22]. This section details the methodological framework employed in the development of Kreol Morisien to English and English to Kreol Morisien language models. Several approaches have been developed and are described below, including OpenNMT, statistical machine translation using Moses, a Transformer model, and MarianMT.

5.1. OpenNMT System

OpenNMT is an open-source toolkit for neural machine translation [23]. The main package required for the translation system is PyTorch in which the OpenNMT models have been developed. An encoder and decoder with six layers make up the model. A checkpoint is saved after every 1000 epochs, and 75,000 epochs are used to train the model. The training was completed in around 4 h.

5.2. Statistical Machine Translation Using the Moses Tool

Statistical machine translation is a traditional approach in the field of language translation. The statistical method for machine translation is implemented in the open-source toolkit known as Moses [24]. Moses facilitates the development of statistical translation models by leveraging large bilingual corpora. Figure 1 depicts a high-level picture of Moses' different stages.

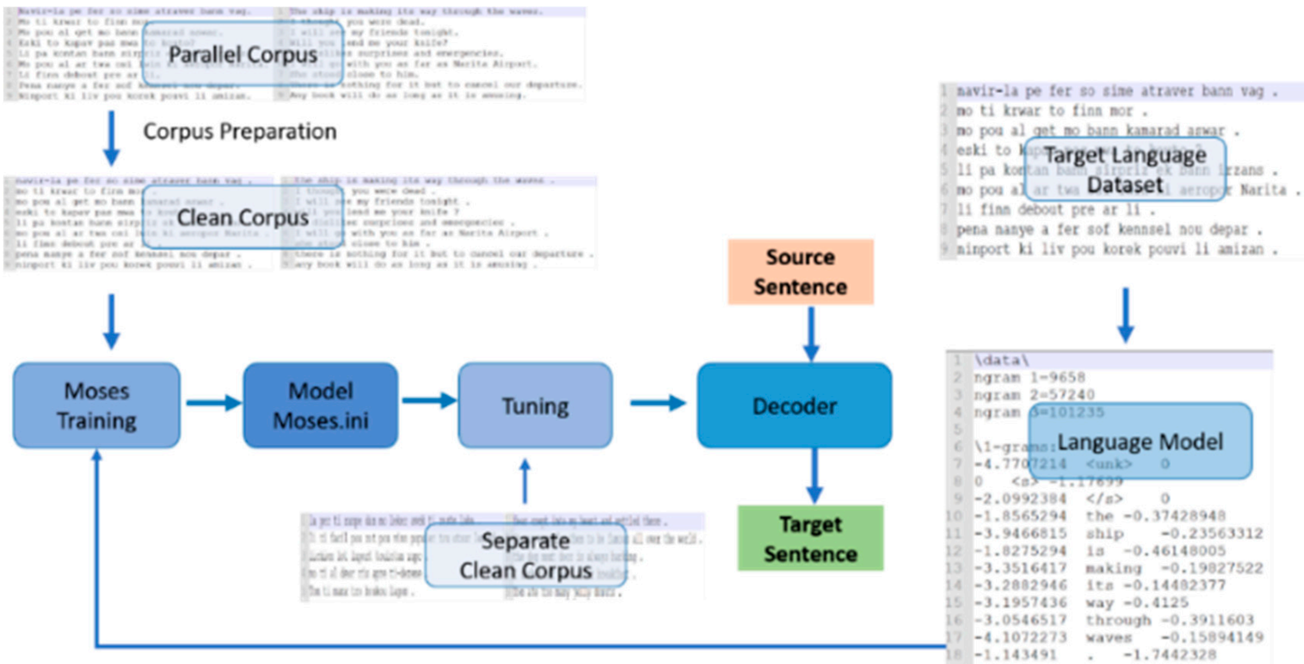

Figure 1. Overview of Moses.

5.3. Application of Deep Learning Algorithm

Previously, translation relied on interpreting a sentence and compressing all the data into a fixed-length vector [25]. A long sentence leads to information loss. The attention mechanism has altered the technique and is one of the innovations in machine translation [26]. When compared to more conventional machine translation methods, NMT (neural machine translation) has produced higher quality translations. The self-attention mechanism model is paired with a feed-forward neural network on the encoder and decoder sides of the transformer, and it takes the role of the original recurrent neural network or convolutional neural network [27].

The Spyder IDE from Anaconda was used for conducting these experiments. A separate virtual environment was created from Anaconda. The Python programming language was used. The Tensor2Tensor (T2T) library, developed by Google, was used for training the deep learning models. The models were trained for 80,000 steps for both

Dataset 1 and Dataset 2. After each 20,000 steps, a model checkpoint was saved. Table 3 demonstrates how the translation system gets better until it approaches the best English-to-Kreol Morisien translation.

Table 3. Progress of English-to-Kreol Morisien translation.

	Sentences	**Dataset 1—Model Trained on 24,000 Sentences**	**Dataset 2—Model Trained on 48,000 Sentences**
1	Source sentence:	The box is rotten.	
	Translated sentence 1:	Sa-la se.	Sa-la se.
	Translated sentence 2:	Bwat-la pouri.	Bwat-la pouri.
	Translated sentence 3:	Bwat-la pouri.	Bwat-la pouri.
	Final Translation:	Bwat-la pouri.	Bwat-la pouri.
	Target Translation:	Bwat-la pouri.	
2	Source sentence:	People laughed at the boy.	
	Translated sentence 1:	Zot ti fer-la.	Zot ti fer-la.
	Translated sentence 2:	Lopinion lor garson-la.	Lopinion lor garson-la.
	Translated sentence 3:	Dimounn ti riy a garson-la.	Dimounn ti riy a garson-la.
	Final Translation:	Dimounn ti riy garson-la.	Dimounn ti riy garson-la.
	Target Translation:	Bann dimounn ti riy garson-la.	
3	Source sentence:	Parents are responsible for their children's education.	
	Translated sentence 1:	Bann zot.	Bann zot.
	Translated sentence 2:	Bann paran responsab pou zot zenfan.	Bann paran responsab pou zot zenfan.
	Translated sentence 3:	Bann paran responsab pou zot zanfan so ledikasion.	Bann paran responsab pou zot zanfan so ledikasion.
	Final Translation:	Bann paran responsab pou zot zanfan so ledikasion.	Bann paran responsab pou zot zanfan so ledikasion.
	Target Translation:	Bann paran responsab pou zot zanfan so ledikasion.	

Notable progress in translation quality is shown when the model's translation performance is assessed across the three sentences. The model's translations for sentence 1, "The box is rotten." contained partial or incorrect sentences at first, but by the last iteration, it was consistently translated to "Bwat-la pouri.", which is aligned with the manual translation. The word "Bann" is missing from the automatic translation in both models. However, this is not a mistake per se, as it is possible to omit the qualifier "bann" in many cases where it is clear that we are referring to many objects and not just one. The third sentence has been translated correctly by both models.

Table 4 demonstrates how the translation system gets better until it approaches the best Kreol Morisien-to-English translation.

Table 4 shows the various stages of refinement across three source sentences for KM-to-English translation for both models. The first two sentences can be considered as perfect translations. However, both the smaller model and the larger one were not able to correctly translate the longer sentence (sentence 3). The major mistake was the inability to understand the word "san" which means "without".

Table 4. Progress of Kreol Morisien-to-English translation.

	Sentences	Dataset 1—Model Trained on 24,000 Sentences	Dataset 2—Model Trained on 48,000 Sentences
1	Source sentence:	To bizin aret bwar.	
	Translated sentence 1:	You should.	You should.
	Translated sentence 2:	You must stop drinking.	You must stop drinking.
	Translated sentence 3:	You must stop drinking.	You must stop drinking.
	Final Translation:	You should stop drinking.	You should stop drinking.
	Target Translation:	You must stop drinking.	
2	Source sentence:	Nou pe fer enn fet tanto.	
	Translated sentence 1:	We are a day.	We are a day.
	Translated sentence 2:	We are making a party tonight.	We are making a party tonight.
	Translated sentence 3:	We are doing a party tonight.	We are doing a party tonight.
	Final Translation:	We are doing a party tonight.	We are doing a party tonight.
	Target Translation:	We are doing a party tonight.	
3	Source sentence:	To ti bizin reflesi avan to al dan lapli san parapli.	
	Translated sentence 1:	You must must you to go go go go go go.	You must must you to go go go go go go.
	Translated sentence 2:	You had better think it is before you were raining.	You had better think it is before you were raining.
	Translated sentence 3:	You had to think before you went to rain in the umbrella.	You had to think before you went to rain in the umbrella.
	Final Translation:	You had to think it before you go to rain in umbrella.	You had to think it before you go to rain in umbrella.
	Target Translation:	You had to think before you go in the rain without an umbrella.	

5.4. MarianMT

One important tool in the field of neural machine translation (NMT) is MarianMT, which is a component of the Marian machine translation (MT) framework. It has been extensively used for both research and real-world applications and is renowned for its highly efficient translation models [28]. MarianMT was created in response to the demand for scalable, effective, and low-latency machine translation models to handle a variety of issues in multilingual environments.

MarianMT's start is linked to the development of deep learning models in the translation domain, specifically after the triumph of attention mechanisms like the Transformer model [29] and sequence-to-sequence models [30]. MarianMT offers a versatile, high-performing translation engine by integrating these developments. Junczys-Dowmunt et al. (2018) initially presented the Marian framework in a study that described its architecture and included important advances, including the capacity to operate on a single GPU or multiple GPU configurations [28].

MarianMT's effective support for large-scale translation models is what makes it significant. This fully independent neural machine translation system offers quick training times, compatibility with multiple architectures, such as Transformer, and integration with SentencePiece for subword tokenization, which is a crucial feature for translating texts with uncommon vocabulary and low-resource languages [31]. MarianMT employs cutting-edge methods that dramatically improve translation quality between languages, such as shared encoder–decoder parameters in multilingual models [32]. This is significant because

MarianMT addresses a major shortcoming of previous NMT systems that had trouble with less common language pairs by being able to handle both high- and low-resource languages. This is a significant advantage for the translation involving the Kreol Morisien language, which is considered as an under-resourced language with its limited dataset available. Dataset 1 and Dataset 2 have been used to train the Kreol Morisien-to-English models and vice versa using MarianMT.

The code has been set up using a sequence-to-sequence Transformer model from the Hugging Face library. The learning rate is set to 2×10^{-5}, and the training and evaluation batch sizes are both set to 16. The model will use a weight decay of 0.01, and training will run for five epochs. The trainer is then initialized with these settings, along with the training and evaluation datasets, as well as the tokenizer. The learning rate and epoch size adjustments in the training configuration do not significantly affect the final result, as measured by the BLEU and METEOR scores. Despite modifying the learning rate and the number of training epochs, the scores remain consistent, signifying that these changes may not lead to substantial improvements or degradation in translation quality.

6. Results and Evaluation

In this section, the performance of the Moses tool, the OpenNMT system, the Transformer model, and MarianMT are analysed and discussed.

6.1. Evaluation Using BLEU Score

The BLEU (Bilingual Evaluation Understudy) score counts the number of matches by performing the comparison between the n-gram of the machine translation and the n-gram of the human translation [33]. A greater number of matches implies a better machine translation. Table 5 and Figure 2 show the BLEU scores for each system.

Table 5. The BLEU scores of Moses tool, OpenNMT system, Transformer system, and MarianMT.

Model	Dataset	English to Kreol Morisien	Kreol Morisien to English
Moses Tool	Dataset 1	0.126	0.204
	Dataset 2	0.14	0.23
OpenNMT system	Dataset 1	0.154	0.172
	Dataset 2	0.17	0.24
Transformer with attention mechanism	Dataset 1	0.21	0.19
	Dataset 2	0.30	0.31
MarianMT	Dataset 1	0.59	0.52
	Dataset 2	0.62	0.58

The results show that Moses and OpenNMT are not suitable for translating English to Kreol Morisien as the BLEU scores are less than 0.20 in all the experiments. However, the translation of Kreol Morisien to English is significantly better, although the BLEU scores are still low. With the Transformer model, we managed to reach a BLEU score of 0.30 for English to KM and 0.31 for KM to English. The models can translate simple and short sentences with good accuracy. However, they have difficulty handling entities such as names of people and places. Using MarianMT, we were able to achieve a BLEU score of 0.62 for English-to-KM translation and 0.58 for KM-to-English translation. MarianMT is able to handle named entities very well. However, it has some difficulties with more complex entities, especially those involving names of places which consists of two or more words.

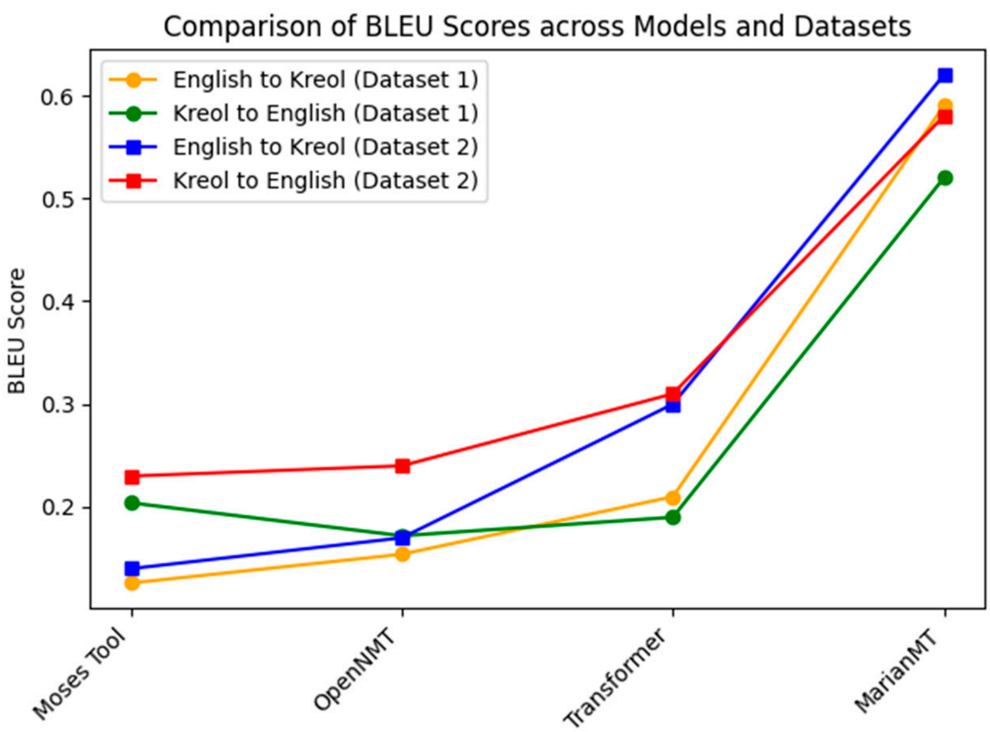

Figure 2. BLEU scores for the translation systems.

6.2. Evaluation Using METEOR

METEOR (Metric for Evaluation of Translation with Explicit Ordering) performs the machine translation evaluation by calculating a score based on exact word matches between the machine translation and a given reference translation [34]. Table 6 shows the METEOR scores across the different models and datasets.

Table 6. Comparison of METEOR scores across the different models and datasets.

Model	Dataset	English to Kreol Morisien	Kreol Morisien to English
Moses tool	Dataset 1	0.11	0.14
	Dataset 2	0.14	0.20
OpenNMT system	Dataset 1	0.16	0.17
	Dataset 2	0.17	0.20
Transformer with attention mechanism	Dataset 1	0.20	0.20
	Dataset 2	0.28	0.33
MarianMT	Dataset 1	0.49	0.53
	Dataset 2	0.54	0.60

Table 6 and Figure 3 show that the Moses tool delivers the lowest scores, reflecting its limited capabilities in this domain, whereas MarianMT consistently outperforms other models across both datasets and translation directions, achieving the highest scores (up to 0.60). This suggests its robustness and adaptability for the translation of Kreol Morisien to English and vice versa. Section 6.4 compares the translated sentences produced by MarianMT against ChatGPT and Google Translate.

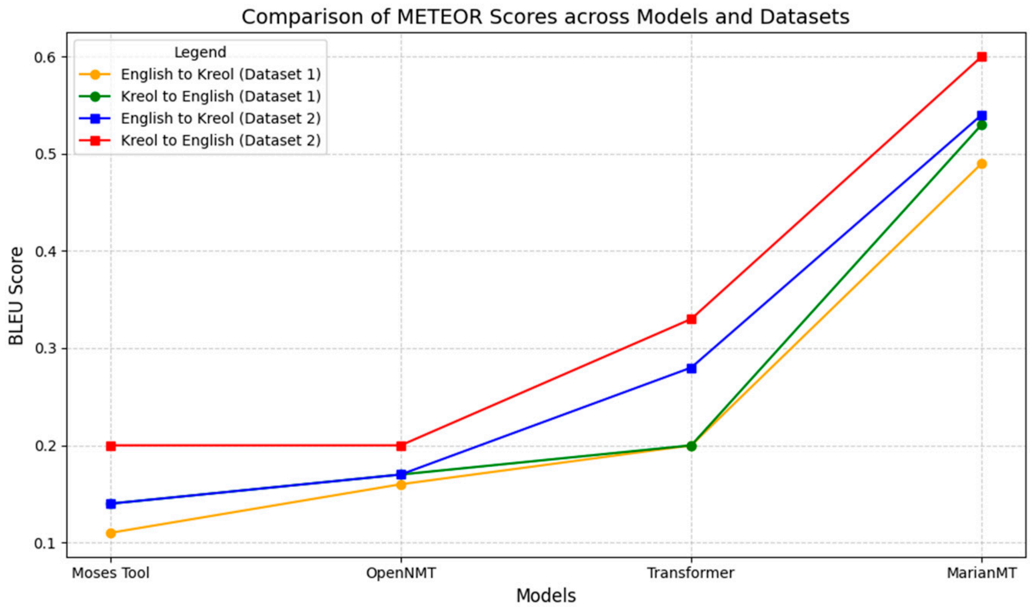

Figure 3. METEOR scores for the translation systems.

6.3. Evaluation Using TER

The TER (Translation Error Rate) score measures the amount of editing that a translator would have to perform to change a translation so it exactly matches the expected translation. Table 7 shows the TER scores across the different models and datasets.

Table 7. Comparison of TER scores across the different models and datasets.

Model	Dataset	English to Kreol Morisien	Kreol Morisien to English
Moses tool	Dataset 1	0.14	0.20
	Dataset 2	0.15	0.21
OpenNMT system	Dataset 1	0.15	0.17
	Dataset 2	0.19	0.22
Transformer with attention mechanism	Dataset 1	0.27	0.21
	Dataset 2	0.38	0.35
MarianMT	Dataset 1	0.52	0.50
	Dataset 2	0.49	0.51

Table 7 and Figure 4 show that MarianMT consistently achieves the highest BLEU scores across all datasets, which indicates superior translation performance. Also, Dataset 2 translations (both Kreol to English and English to Kreol) obtain higher BLEU scores than Dataset 1. This indicates that increasing the number of parallel sentences in the dataset has a positive influence on the TER score. The next section compares the translated sentences produced by MarianMT against a chatbot and Google Translate.

6.4. Comparison of Translated Sentences Against a Chatbot (ChatGPT-4) and Google Translate

AI-based translation systems are readily accessible. However, translators such as DeepL do not cater for the Kreol Morisien language. Google Translate added the Kreol Morisien language in 2024. This section compares and analyses the translated sentences against ChatGPT-4 and Google Translate for English to Kreol Morisien and vice versa. Table 8 shows the English-to-Kreol Morisien translations.

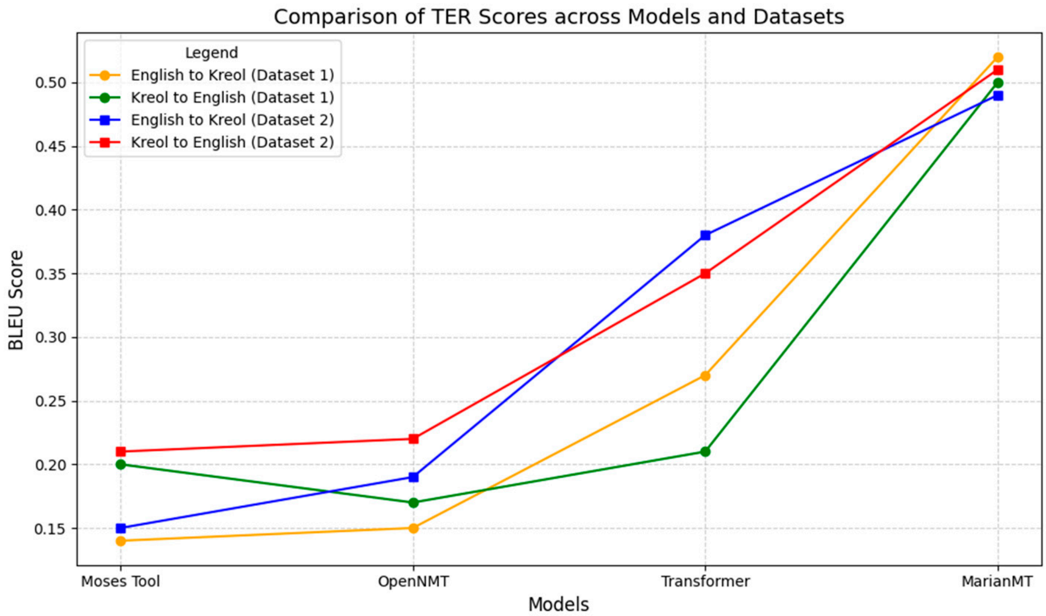

Figure 4. TER scores for the translation systems.

Table 8. English-to-Kreol Morisien comparison.

#	Language	Sentence
1	English	Tom will go there tomorrow.
	Kreol Morisien	Tom pou al laba demin.
	MarianMT	Tom pou al laba demin.
	ChatGPT	Tom pou al laba dime.
	Google Translate	Tom pou al laba demin.
2	English	There were ten police officers on the spot.
	Kreol Morisien	Ti ena dis lapolis lor terain.
	MarianMT	Ti ena dis lapolis lor plas.
	ChatGPT	Ti ena dis polisye lor plas.
	Google Translate	Ti ena dis polisie lor plas.
3	English	The rumor may be true.
	Kreol Morisien	Rimer-la kapav vre.
	MarianMT	Rimer-la kapav vre.
	ChatGPT	Lavi razer kapav vre.
	Google Translate	Sa rimer-la kapav vre.
4	English	The milk has gone bad.
	Kreol Morisien	Dile-la in gate.
	MarianMT	Dile-la finn fer move.
	ChatGPT	Lait-la finn gate.
	Google Translate	Dile-la finn vinn move.
5	English	Maria's late husband was a violinist.
	Kreol Morisien	Maria so misie ki finn mor ti enn violonis.
	MarianMT	Maria so misie tar ti enn violonis.
	ChatGPT	Mari defen Maria ti enn violinist.
	Google Translate	Mari so defen mari ti enn violonis.

Table 8. *Cont.*

#	Language	Sentence
6	English	I'm getting a new house built.
	Kreol Morisien	Mo pe ranz enn nouvo lakaz.
	MarianMT	Mo pe fer enn nouvo lakaz aranze.
	ChatGPT	Mo pe fer konstrir enn nouvo lakaz.
	Google Translate	Mo pe fer konstrir enn nouvo lakaz.
7	English	Alyssa is an event planner working in Washington.
	Kreol Morisien	Alyssa se enn planner levennman ki pe travay dan Washington.
	MarianMT	Alyssa enn plan dan evennman travay dan Washington.
	ChatGPT	Alyssa enn planifikater levennman ki travay Washington.
	Google Translate	Alyssa se enn planner levennman ki travay dan Washington.
8	English	Tania lives at Pointe aux Piments.
	Kreol Morisien	Tania res dan Pointe aux Piments.
	MarianMT	Tania res dan frazid-la.
	ChatGPT	Tania res Pointe aux Piments.
	Google Translate	Tania res dan Pointe aux Piments.
9	English	The sega is the traditional dance of Mauritius.
	Kreol Morisien	Sega, se lamizik tradisionel pou Moris.
	MarianMT	Sega enn lamizik tradisionel pou Moris.
	ChatGPT	Sega se dans tradisionel Moris.
	Google Translate	Sega se ladans tradisionel Moris.
10	English	The dodo is the national bird of Mauritius.
	Kreol Morisien	Dodo se zwazo nasional pou Moris.
	MarianMT	Dodo se zwazo nasional pou Moris.
	ChatGPT	Dodo se zwazo nasional Moris.
	Google Translate	Dodo se zwazo nasional Moris.
11	English	Actions speak louder than words.
	Kreol Morisien	Aksion koz pli for ki parol.
	MarianMT	Bann aksion dir pli for ki bann mo.
	ChatGPT	Aksyon koz pli for ki parol.
	Google Translate	Bann aksion koz pli for ki bann parol.
12	English	Many hands make light work.
	Kreol Morisien	Bann lame ansam fer travay vinn fasil.
	MarianMT	Boukou lame fer mwins travay.
	ChatGPT	Plizir lame fer travay vinn pli fasil.
	Google Translate	Boukou lame fer travay leze.
13	English	Do not count your chickens before they hatch.
	Kreol Morisien	Pa kont ou poul avan ki zot eklo.
	MarianMT	Pa kont to bann poul avan ki zot vini.
	ChatGPT	Pa kont ou poul avan zot kokin.
	Google Translate	Pa kont ou bann poul avan zot eklo.

Table 8. *Cont.*

#	Language	Sentence
14	English	Where there is smoke, there is fire.
	Kreol Morisien	Kot ena lafime, ena dife.
	MarianMT	Kan ena lafime, ena dife.
	ChatGPT	Kot ena lafime, ena dife.
	Google Translate	Kot ena lafime, ena dife.
15	English	That is a tough nut to crack.
	Kreol Morisien	Sa enn problem difisil pou rezoud.
	MarianMT	Sa li difisil pou krake.
	ChatGPT	Sa enn gro zafer pou konpran.
	Google Translate	Sa li enn nouritir difisil pou kase.

Each English sentence translation to Kreol Morisien is compared with the translation model, ChatGPT, and Google Translate. The translations provided by the model show that some sentences are closer to the target translations, while others deviate significantly. The first sentence has been translated perfectly. In the second sentence, the use of "plas" instead of the more contextually appropriate "terain" produces a satisfactory translation since the context remains unchanged. The third sentence has been perfectly translated. The model can also translate phrases which include cultural nuances correctly such as translation numbers 9 and 10. Idiomatic sentences have also been well handled by the translator. This includes translation numbers 11 to 15. Although some rare words such as "hatch" have not been correctly translated, the meanings of the phrases have been retained.

Translation numbers 4 and 5 show that the translations have been performed literally without recognising the specific phrasing used for rotten food or deceased person, making the translated sentence sound unnatural. In translation number 6, the translated sentence misses the passive construction. This shows the model's challenge in handling some sentence structures. More complex sentences like number 7 are also problematic. The translation is wrong, both in meaning and syntax, as it completely misrepresents the role and action of the subject. This shows that the model struggles with rare entities, in this case, names of people. The name of a person has been incorrectly translated, while the name of a place has been correctly decoded. In sentence number 8, the name of a place has been incorrectly handled. Overall, while the model performs decently in capturing the basic meaning of sentences, it often fails to convey idiomatic expressions, leading to less fluent and sometimes inaccurate translations. Thus, in our future works, we intend to develop a hybrid machine learning system which will involve the inclusion of some pre-processing and post-processing steps using a rule-based approach to handle entities and unknown words.

Generally, the ChatGPT and Google Translate models show fair translations. However, some issues can still be encountered. For sentence number 3, ChatGPT's output introduces the term "razer" which is unconnected to the context and creates misperception. In translation number 4, ChatGPT's translation, introduces French vocabulary to the Kreol Morisien sentence which is incorrect. Google Translate translates "bad" to "move" ("become bad"), which is an odd expression in this context. For translation number 5, ChatGPT's translation, "Mari defen Maria ti enn violonist", introduces multiple issues. The word "defen" ("late") is redundantly linked to "Maria", creating confusion. Additionally, "violonist" deviates from the preferred spelling "violonis." For the case of Google Translate, the entity name "Maria" has been wrongly written. For translation number 13, ChatGPT incorrectly translates "hatch" to "kokin" ("steal"), which is incorrect and changes the meaning of the sentence. For translation number 15, Google Translate incorrectly translates the idiomatic sentence.

The 15 sentences are evaluated using the BLEU Score. MarianMT achieved a BLEU score of 0.37, while ChatGPT-4 scored 0.31. Google Translate also obtained a BLEU score of 0.37. This demonstrates our model's performance in producing precise and reliable translations. Although the BLEU score obtained by our model and Google Translate is the same, the output is not exactly the same. In some cases, MarianMT performs better than Google Translate, such as translation number 15, where Google Translate incorrectly translates "problem" to "nouritir" ("food"). There are also sentences where Google Translate gives a good translation such as translation numbers 13 and 14. The score obtained by ChatGPT-4 is lower because there are instances where there are spelling mistakes, such as in translation number 11, where the word "aksion" ("action") has been incorrectly written. However, there are places where ChatGPT-4 has produced good translations such as translation number 12. Increasing the dataset will undoubtedly improve the BLEU score and the translation quality. Table 9 shows the Kreol Morisien-to-English translations using the different models.

Table 9. Kreol Morisien-to-English comparison.

#	Language	Sentence
1	Kreol Morisien	To kapav azir kouma to swete.
	English	You may act however you wish.
	MarianMT	You may act however you wish.
	ChatGPT	You can act as you wish.
	Google Translate	You can act as you wish.
2	Kreol Morisien	Mo konze finn pas bien vit.
	English	My holiday went by quickly.
	MarianMT	My holiday has been very fast.
	ChatGPT	My holiday went by very quickly.
	Google Translate	My holiday passed quickly.
3	Kreol Morisien	Eski sa ti to vakans?
	English	How was your vacation?
	MarianMT	What was it your vacation?
	ChatGPT	Was that your vacation?
	Google Translate	Was this your holiday?
4	Kreol Morisien	Eski nou kapav ena enn latab dan seksion kot pa fime la ?
	English	Could we have a table in the non-smoking section?
	MarianMT	Can we have a table in the section where it is not smoking?
	ChatGPT	Can we have a table in the non-smoking section?
	Google Translate	Can we have a table in the non-smoking section.
5	Kreol Morisien	Sa bon nouvel la finn amen larm dan zot lizie.
	English	The good news brought tears to their eyes.
	MarianMT	The good news brought tears to their eyes
	ChatGPT	That good news brought tears to their eyes.
	Google Translate	The good news brought tears to their eyes.

Table 9. *Cont.*

#	Language	Sentence
6	Kreol Morisien	Mo finn tann John koz ar misie Brown.
	English	I heard John speak to Mr Brown.
	MarianMT	I heard John talk to Mr Brown.
	ChatGPT	I heard John talking to Mr. Brown.
	Google Translate	I heard John talking to Mr. Brown.
7	Kreol Morisien	Bann zom mo zwenn la bien zanti.
	English	The men I meet are very kind.
	MarianMT	Men’s meeting are very kind.
	ChatGPT	The men I met there are very kind.
	Google Translate	The men I meet are very nice.
8	Kreol Morisien	Bin la pa ti ase so e mo pa finn kapav amiz sa.
	English	The bath was not hot enough and I was unable to enjoy it.
	MarianMT	The Bin was not hot enough and I couldn’t enjoy it.
	ChatGPT	The bath wasn’t warm enough, and I couldn’t enjoy it.
	Google Translate	The bin wasn’t hot enough and I couldn’t have fun with it.
9	Kreol Morisien	Sega, se lamizik tradisionel pou Moris.
	English	The sega is the traditional dance of Mauritius.
	MarianMT	Sega is the song traditional for Mauritius.
	ChatGPT	Sega is the traditional music of Mauritius.
	Google Translate	Sega is the traditional music of Mauritius.
10	Kreol Morisien	Dodo ti zwazo nasional pou Moris.
	English	The dodo is the national bird of Mauritius.
	MarianMT	Dodo is a national bird for Mauritius.
	ChatGPT	The dodo was the national bird of Mauritius.
	Google Translate	The dodo was the national bird of Mauritius.
11	Kreol Morisien	Aksion koz pli for ki parol.
	English	Actions speak louder than words.
	MarianMT	Actions speak more loud than words.
	ChatGPT	Actions speak louder than words.
	Google Translate	Actions speak louder than words.
12	Kreol Morisien	Bann lame ansam fer travay vinn fasil.
	English	Many hands make light work.
	MarianMT	Many hands do the work easy.
	ChatGPT	Many hands make light work.
	Google Translate	Hands together make work easy.
13	Kreol Morisien	Pa kont ou poul avan ki zot eklo.
	English	Do not count your chickens before they hatch.
	MarianMT	Do not count your hen before them.
	ChatGPT	Don’t count your chickens before they hatch.
	Google Translate	Don’t count your chickens before they hatch.

Table 9. *Cont.*

#	Language	Sentence
14	Kreol Morisien	Kot ena lafime, ena dife.
	English	Where there is smoke, there is fire.
	MarianMT	Where there is smoke, there is fire.
	ChatGPT	Where there is smoke, there is fire.
	Google Translate	Where there is smoke, there is fire.
15	Kreol Morisien	Sa enn problem difisil pou rezoud.
	English	That is a tough nut to crack.
	MarianMT	This is a difficult problem to solve.
	ChatGPT	That is a difficult problem to solve.
	Google Translate	This is a difficult problem to solve.

Table 9 shows that the Kreol Morisien-to-English translations show a varied level of accuracy as compared to the target sentences. In many cases, as in sentence numbers 1 and 5, the model's output is completely correct and matches the expected sentence, which shows that the model can produce accurate translations for shorter and simpler sentences. It can be seen that even if some sentences do not match the expected ones completely, the translated sentences are not incorrect, as in the case for translation number 2. The words "bien vit" have been translated to "very fast" instead of "quickly", which still gives a perfect translation output. Sentences that include cultural instances, such as sentence numbers 9 and 10, are correctly translated. Idiomatic translations, such as translation numbers 11 to 15, have also been performed correctly and the meaning of the sentence is retained.

Sentence number 3 has been translated literally and does not capture the intended meaning of the original sentence. Sentence number 4 is satisfactory as it is not incorrect, but it lacks fluidity of the target language. In number 6, "tande" has been translated to "talk" which is grammatically correct, and the translated sentence makes sense in English. The translated sentence number 7 is grammatically incorrect and confusing. The structure of the translation misinterprets the phrase "bann zom mo zwenn la" (the men I meet) and introduces a possessive form ("Men's meeting"), which distorts the meaning completely. In the translated sentence number 8, the model incorrectly renders "bin" as "Bin", which could be a misunderstanding of the word in context.

ChatGPT and Google Translate perform the translation from Kreol Morisien to English fairly well. However, for translation number 8, Google Translate did not translate the word "bin" ("bath") and this affects the translation quality. This shows that some words are not included in the Google Translate model, which makes it difficult to attain a perfect translation.

Evaluation of the sentences in Table 9 were performed using the BLEU score. MarianMT obtains a BLEU score of 0.73, while Google Translate achieves a BLEU score of 0.76. ChatGPT outperforms by obtaining a BLEU score of 0.84. In general, the Kreol Morisien-to-English MarianMT model performs well with simpler sentences but struggles with rare words, such as translation number 13, where the model did not translate the word "hatch". This suggests that while it can provide translations that are often close to the expected result, improvements are needed to handle rare words. However, there are instances where MarianMT produces a good translation, although it does not match the expected output, such as translation number 12, where the sentence "many hands do the work easy" is correct. Increasing the vocabulary and the dataset will surely improve the BLEU score.

7. Conclusions and Future Work

In this paper, we present a new dataset consisting of 50,000 parallel Kreol Morisien and English sentences, and 48,000 sentence pairs were used to build the models, 1000 for evaluation and another 1000 for testing the models. Our best models achieved a BLEU score of 0.62 for the translation of English to Kreol Morisien and a BLEU score of 0.58 for the translation of Kreol Morisien to English using the MarianMT framework. These are the highest BLEU scores that are available in the literature between Kreol Morisien and English. The MarianMT model has demonstrated its ability to produce good quality translations. By comparing MarianMT against ChatGPT and Google Translate, it can be seen that our model performs fairly well and can produce some translations that are similar to manually translated sentences. However, the model struggles with entities and unknown words. ChatGPT and Google Translate handle entities fairly well because they are trained on a larger dataset.

As future work, we intend to explore and develop a hybrid rule-based machine translation system. The pace of research and development in hybrid machine translation systems is moving very fast with new improvements for the translation of under-resourced languages. We plan to perform some pre-processing steps using a rule-based approach to handle entities and rare or unknown words and standardise the English language. Neural machine translation will still be used to perform the core translation. After the translation step, post-processing steps using a rule-based approach will be applied to restructure the sentence.

We believe that using a hybrid rule-based and deep learning algorithm can have a positive influence on the BLEU scores and on the overall quality of the translated texts. A translation system for English to Kreol Morisien and vice versa can be used in several contexts, such as public notices, to ensure accessibility to all citizens, facilitate the learning process for Kreol Morisien speakers, and enhance tourism experiences by fostering a greater cultural exchange. An efficient translator will promote cultural preservation and societal inclusivity and maintain its relevance in a digital age.

Author Contributions: Conceptualization, S.P., N.C. and A.S.; methodology, Z.B.J.B.; software, Z.B.J.B.; validation, S.P., N.C. and A.S.; formal analysis, Z.B.J.B.; investigation, Z.B.J.B.; resources, S.P., N.C. and A.S.; data curation, Z.B.J.B.; writing—original draft preparation, Z.B.J.B.; writing—review and editing, Z.B.J.B., S.P., N.C. and A.S.; visualization, Z.B.J.B.; supervision, S.P., N.C. and A.S.; project administration, S.P., N.C. and A.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The dataset presented in this article is not readily available because the data are part of an ongoing study. Requests to access the dataset should be directed to the authors.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Li, B.; Weng, Y.; Xia, F.; Deng, H. Towards better Chinese-centric neural machine translation for low-resource languages. *Comput. Speech Lang.* **2024**, *84*, 101566. [CrossRef]
2. Savage, C. The Importance of Mother Tongue in Education0000000000 2019. Available online: https://ie-today.co.uk/people-policy-politics/the-importance-of-mother-tongue-in-education/#:~:text=Language%20and%20mother%20tongue%20also,sense%20of%20wellbeing%20and%20confidence (accessed on 19 September 2024).
3. Cummins, J. Bilingual children's mother tongue: Why is it important for education? *Int. Small Bus. J.* **2001**, *7*, 15–20.
4. Le Mauricien. Introducing Kreol Morisien in Parliament. 2020. Available online: https://www.lemauricien.com/le-mauricien/introducing-kreol-morisien-in-parliament/370057/ (accessed on 19 September 2024).

5. Mauritius National Assembly. 2020. Available online: https://mauritiusassembly.govmu.org/SitePages/Index.aspx (accessed on 17 September 2024).
6. Govmu.org. Updated Edition of Report on Spelling and Grammatical Rules of Creole Language of the Republic of Mauritius Launched. 2024. Available online: https://www.govmu.org/EN/newsgov/SitePages/Updated-edition-of-Report-on-spelling-and-grammatical-rules-of-Creole-Language-of-the-Republic-of-Mauritius-la.aspx (accessed on 19 September 2024).
7. L'Express Akademi Kreol Repiblik Moris: Ignition Delays. L'Express 19 February 2021. Available online: https://www.lexpress.mu/article/389161/akademi-kreol-repiblik-moris-retards-lallumage (accessed on 10 September 2024).
8. Eisenlohr, P. Creole publics: Language, cultural citizenship, and the spread of the nation in Mauritius. *Comp. Stud. Soc. Hist.* **2007**, *49*, 968–996. [CrossRef]
9. Mooneeram, R. The contribution of creative writing to the standardization of Mauritian Creole. *Lang. Lit.* **2007**, *16*, 245–261. [CrossRef]
10. Le Mauricien [Sylvio Michel, President of the Fraternal Greens: "Less Misery, If the Americans Had Recruited the Chagossians"]. Le Mauricien, 1 September 2018. Available online: https://www.lemauricien.com/le-mauricien/sylvio-michel-president-des-verts-fraternels-moins-de-misere-si-les-americains-avaient-recrute-les-chagossiens/228630/ (accessed on 19 September 2024).
11. The Alfa King Diary. Rodrigues: Advocacy for Educational Tourism. 2020. Available online: https://alfaking4fr.wordpress.com/category/actualites/ (accessed on 16 September 2024).
12. Betchoo, N.K. A Review of the Recognition of Kreol as a National Language in Mauritius. *Asian J. Basic Sci. Res.* **2019**, *1*, 36–39.
13. Hookoomsing, V. Grafi-Larmoni: A Harmonized Writing System for the Mauritian Creole Language. 2004. Available online: https://z-lib.id/book/a-harmonized-writing-system-for-the-mauritian-creo (accessed on 19 September 2024).
14. Harmon, J.D. Heritage Language & Identity Construction: A Critical Ethnography of Kreol Morisien as an Optional Language in Primary Education Within the Republic of Mauritius. Ph.D. Thesis, University of the Western Cape, Cape Town, South Africa, 2014.
15. Miller, A. Kreol in Mauritian Schools: Mother Tongue Language Education and Public. Ph.D. Thesis, Yale University, New Haven, CT, USA, 2015.
16. L'Express. Morisian Kreol at School Certificate: Cambridge International no Longer Needed. 2021. Available online: https://lexpress.mu/s/article/393615/kreol-morisien-au-school-certificate-plus-besoin-cambridge-international (accessed on 19 September 2024).
17. Le Matinal School Certificate: The Success Pass Is 73.71%. 2024. Available online: https://english.lematinal.media/school-certificate-the-success-pass-is-73-71/ (accessed on 24 September 2024).
18. Pudaruth, S.; Sukhoo, A.; Kishnah, S.; Armoogum, S.; Gooria, V.; Betchoo, N.K.; Chady, F.; Ramoogra, A.; Hanoomanjee, H.; Khodabocus, Z. Morisia: A Neural Machine Translation System to Translate between Kreol Morisien and English. *InTRAlinea Online Transl. J.* **2021**, *23*, 1–6.
19. Dabre, R.; Sukhoo, A. MorisienMT: A Dataset for Mauritian Creole Machine Translation. *arXiv* **2022**, arXiv:2206.02421.
20. Pudaruth, S.; Armoogum, S.; Betchoo, N.K.; Sukhoo, A.; Gooria, V.; Peerally, A.; Khodabocus, M.Z. A neural machine translation system for Kreol Repiblik Moris and English. *IAES Int. J. Artif. Intell.* **2024**, *13*, 4976–4987. [CrossRef]
21. Carpooran, A. *Diksioner Morisien*, 3rd ed.; Editions Le Printemps: Vacoas, Mauritius, 2019.
22. Yang, S.; Wang, Y.; Chu, X. A survey of deep learning techniques for neural machine translation. *arXiv* **2020**, arXiv:2002.07526.
23. Klein, G.; Kim, Y.; Deng, Y.; Senellart, J.; Rush, A.M. Opennmt: Open-source toolkit for neural machine translation. *arXiv* **2017**, arXiv:1701.02810.
24. Nidhi, R.; Singh, T. SMT Algorithms for Indian Languages—A Case Study of Moses and MT Hub for English-Maithili Language Pair. In Proceedings of the ICETIT 2019, Delhi, India, 21–22 June 2019; Springer: Cham, Switzerland, 2020; pp. 269–279.
25. Choi, H.; Cho, K.; Bengio, Y. Fine-grained attention mechanism for neural machine translation. *Neurocomputing* **2018**, *284*, 171–176. [CrossRef]
26. Wang, H.; Wu, H.; He, Z.; Huang, L.; Church, K.W. Progress in machine translation. *Engineering* **2022**, *18*, 143–153. [CrossRef]
27. Jiang, H.; Zhao, S.; Fang, D.; Zhang, C.; Duan, J.A. Comparative Study on Transformer versus Sequence to Sequence in Machine Translation. In *Modern Industrial IoT, Big Data and Supply Chain*; Springer: Singapore, 2021; pp. 89–101.
28. Junczys-Dowmunt, M.; Grundkiewicz, R.; Dwojak, T.; Hoang, H.; Heafield, K.; Neckermann, T.; Seide, F.; Germann, U.; Aji, A.F.; Bogoychev, N.; et al. Marian: Fast neural machine translation in C++. *arXiv* **2018**, arXiv:1804.00344.
29. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 6000–6010.
30. Sutskever, I.; Vinyals, O.; Le, Q.V. Sequence to Sequence Learning with Neural Networks. *arXiv* **2014**, arXiv:1409.3215.
31. Kudo, T. Sentencepiece: A simple and language independent subword tokenizer and detokenizer for neural text processing. *arXiv* **2018**, arXiv:1808.06226.

32. Soliman, A.; Shaheen, S.; Hadhoud, M. Leveraging pre-trained language models for code generation. *Complex Intell. Syst.* **2024**, *10*, 3955–3980. [CrossRef]
33. Khandelwal, U.; Fan, A.; Jurafsky, D.; Zettlemoyer, L.; Lewis, M. Nearest neighbor machine translation. *arXiv* **2020**, arXiv:2010.00710.
34. Agarwal, A.; Lavie, A. Meteor, m-bleu and m-ter: Evaluation metrics for high-correlation with human rankings of machine translation output. In Proceedings of the Third Workshop on Statistical Machine Translation, Columbus, OH, USA, 19 June 2008; pp. 115–118.