International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 9, No. 6 (Nov-2020)
http://dx.doi.org/10.12785/ijcds/0906012

Kreol Morisien to English and English to Kreol Morisien Translation System using Attention and Transformer Model

Zaheenah Boodeea[1] and Sameerchand Pudaruth[1]

[1]Department of Information and Communication Technologies, University of Mauritius, Mauritius

Received 16 Mar. 2020, Revised 19 Apr. 2020, Accepted 31 Jul. 2020, Published 1 Nov. 2020

Abstract: Translation from one language to another is extensively carried out by travellers, students and many other people. Unfortunately, less popular languages, such as Kreol Morisien (KM), are not catered for by popular translation systems. The objective of this work is to develop a system that translates English into Kreol Morisien and vice-versa. The size of the dataset which was used in the model consists of 19,650 pairs of English and Kreol Morisien sentences. The Kreol Morisien words are up to date since they were taken from the third edition of the Diksioner Morisien. A neural machine translation system has been used in this work. A transformer model with attention was then developed which was trained for several iterations. Evaluation of the system is done using the standard BLEU scores. The English to Kreol Morisien model achieves a BLEU score of 0.20 and the Kreol Morisien to English model achieves a BLEU score of 0.23. Both are much higher than existing systems. Furthermore, user evaluation has been carried out in the form of two surveys. Each survey consisted of 25 pairs of sentences in the source language and the target language. Responses were gathered from people with different age groups living in both rural and urban regions and including both students and professionals. We received more than 90 responses for each survey. Evaluation and testing of the translation model using the BLEU score showed that the model can produce satisfactory translations.

Keywords: Translation, English, Kreol Morisien, Neural Network, Transformer, Attention

1. Introduction

Today we are living in a global village where the use of the Internet has increased to such an extent that it is having a positive impact on the economy. Lots of websites and documents are accessible over the Internet. There are more than 4 billion of people who use the internet actively as of July 2019, which covers 56 percent of the population globally [1]. Yet, the soaring usage of the internet has led to a divide between those who can access it and those who cannot. Language happens to be one of the driving factors of the digital divide. Only 20% of the Earth's population speak English [2]. Despite this fact, English is used by the largest group of Internet users. Language barriers exist in Mauritius because many Mauritians are unable to express themselves properly using the English language and foreigners sometimes find it problematic to converse with Mauritians. The formers are usually comfortable only with their native language. Being the most used language in Mauritius, Kreol Morisien stands as a link between the different communities of the island.

Informally, Kreol Morisien (KM) is the medium of expression for most Mauritians. Moreover, over the last five years, the usage of the Kreol Morisien language has increased significantly in the official domain as well. A decade earlier, doing advertising in Kreol was considered vulgar but now it is considered 'cool'. There is reasonably well written Kreol text in many places which include newspapers, on the MBC, in posters, banners, posters and leaflets as well as within official document from the government. In 2018, the 5th bimonthly publication of Revi Lalit, a 32-page magazine in the Kreol Morisien language was published [3]. The taboo against the written Kreol Morisien language is slowly being eradicated.

Today, dictionaries and guidebooks are available on the market to provide a more in-depth insight, clarity and standardization when it comes to writing in Kreol. In 2019, the third edition of the Kreol Morisien dictionary, Diksioner Morisien, was published which consists of around 2500 new entries. The Kreol Morisien dictionary celebrates its tenth anniversary with the release of this new edition. Kreol Morisien is, therefore, becoming essential for Mauritians since it has already been formalised as an official language.

In 2012, the Kreol Morisien language was introduced at school as an optional subject mainly because of the

E-mail: beebee.boodeea@umail.uom.ac.mu, s.pudaruth@uom.ac.mu

increase in the number of failures at PSAC (Primary School Achievement Certificate) which was previously known as the Certificate of Primary Education (CPE) examinations [4]. Children can have a better grasp of early learning since they can better express themselves openly in their mother tongue. Nelson Mandela rightly said that "*If you talk to a man in a language he understands, that goes to his head. If you talk to him in his own language, that goes to his heart*."

Introducing this language at school has indeed been a success since the students have performed better in the Kreol Morisien language compared to other oriental languages as shown in Figure 1 [5].

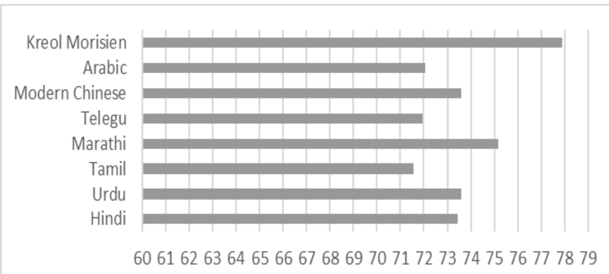

Figure 1. Percentage pass in different subjects at primary level

The Kreol Morisien language is becoming increasingly popular in the ICT field. It is used as a means of communication for sending text messages and emails. Moreover, the contents of certain websites are even displayed in Kreol. Also, Google has contributed towards the recognition of the Kreol Morisien language by providing the Kreol version of its search engine. This is an acknowledgement that the Kreol Morisien is a language and not a patois [6]. With the emergence of new technologies, a translation engine is crucial as it would help overcome language barriers, and facilitate the task of users by performing their needed translation faster.

The fact cannot be denied that the Kreol Morisien language also plays a significant role, especially in the advertising field. Certain articles in the newspaper are nowadays published in Kreol Morisien. There are also a multiple number of programs, including news bulletins, known as the "Zournal Kreol" which are aired on the national television in Kreol Morisien. Furthermore, "Senn Kreol", a TV Channel in Mauritius, which is owned by the Mauritius Broadcasting Corporation (MBC), broadcast shows such as cooking shows and documentaries in the Kreol Morisien language.

This paper attempts to resolve the identified issues by developing a machine translation system that will enable people to switch to and from English and Kreol Morisien languages easily. The structure of the paper is organised as follows. Section 2 shows the relation of Kreol Morisien with other languages. Section 3 describes the relevant machine learning techniques. Some existing machine translation systems are described in Section 4. Section 5 describes the methodology. Testing and evaluation are performed in Section 6 and we conclude the paper in Section 7.

2. Relation of Kreol Morisien With Other Languages

A. Similarities of Kreol Morisien with French language

Kreol Morisien is a French-based Creole language. Some examples of French and Kreol Morisien that are written and pronounced similarly are shown in Table 1. The corresponding English words are also given.

TABLE I. Words With Similar Orthography In French And Kreol Morisien

French	Kreol Morisien	English
four	four	oven
gaz	gaz	gas
selon	selon	according
par	par	by

While writing in the French language, accents are used more as compared to Kreol Morisien. There are many words which have same the pronunciation in French and Kreol, but with different graphemes. Some examples are shown in Table 2. The corresponding English words are also provided.

TABLE II. Words With Different Orthography But Same Pronounciation In French And Kreol Morisien

French	Kreol Morisien	English
insecte	insekt	insect
sale	sal	dirty
oui	wi	yes
dire	dir	tell

B. Differences between Kreol Morisien and English

Differences exist between the English language and the Kreol Morisien language. For example, objects are sometimes moved before the adjectives when translating from English language to Kreol Morisien [7]. Another example is that there is difference between singular and plural [7]. The plural does not take "s" when translating English to Kreol Morisien. Instead, the word "boukou" is used to mean many.

C. Differences between Kreol Morisien (KM) and Haitian Kreol

Haitian Kreol is a French-based Kreol spoken mostly in Haiti, a country located in the Latin American regions. Although Kreol Morisien is also based on the French language, Kreol Morisien and Haitian Kreol are very different. Some of these differences in their vocabulary are shown in Table 3.

TABLE III. DIFFERENCES BETWEEN KREOL MORISIEN AND HAITIAN KREOL

English	Kreol Morisien	Haitian Kreol
You are beautiful.	Ou zoli.	Ou bèl.
I am eating an apple.	Mo pe manz enn pom.	Mwen manje yon pòm.
I am hungry.	Mo pe gagn fin.	Mwen grangou.
I am coming to the shop.	Mo pe vinn laboutik.	Mwen vini nan boutik la.

3. MACHINE TRANSLATION

The concept of machine translation has been around since the middle of last century. Machine translation is a sub-field of computational linguistics which does the translation of one language to another through a computer algorithm [8]. A significant amount of data can be translated in a very short amount of time using machine translation, and it is much less costly as compared to human translators [9].

Machine translation systems can be implemented using different approaches which include rule-based algorithms, statistical systems, knowledge-based approaches and deep learning. Rule-based translation consists of grammatical rules and programs to translate a source language into a target language. The first stage is to get the word classes for each word, as shown in Table 4.

TABLE IV. WORD CLASSES

Word	Word classes
a	indefinite article
boy	noun
eats	verb
an	indefinite article
orange	noun

Next, the syntactic information about verbs and other nouns are obtained. The final stage is the mapping of the words into their appropriate inflected forms as shown in Table 5~~V~~.

TABLE V. ENGLISH TO KREOL MORISIEN MAPPINGS

English	Kreol Morisien
a	enn
boy	garson
eats	manz
an	enn
orange	zoranz

The statistical approach uses a probabilistic process based on statistics derived from a large corpus of parallel sentences. If given enough training data, translation systems using this approach perform much better than rule-based systems [10]. KANT is an example of a Knowledge Based Machine Translation (KMBT) system which was developed on a big scale for multilingual translation [11].

Deep learning is a subcategory of machine learning while machine learning is an important branch of Artificial Intelligence. Deep learning also creates patterns which are then used to make decisions [12]. A Recurrent Neural Network (RNN) is a form of deep learning which is designed in such a way that the inputs and outputs are sequences of text. Words that are generated at each position in a sentence in the target language are dependent in some manner on all the previous words in that sentence. This recurrence is used as a form of memory.

The aim of a sequence-to-sequence model is to translate text where the input and the output are sentences. The latter is a series of words which goes inside and outside of a model [13][14][15]. Figure 2 shows how neural machine translation from English to Kreol Morisien using an LSTM (Long Short-Term Memory) model takes place using an Encoder-Decoder network

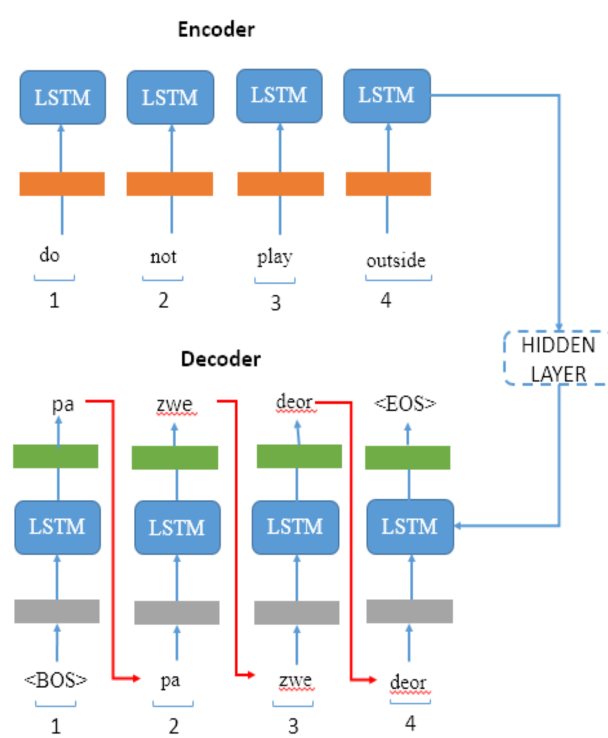

Figure 2. Translation using an Encoder-Decoder network [16]

4. Literature Review

There have been many attempts to show the performance of neural machine translation approaches on Indian languages. Jadhav (2020) [17] introduced a Marathi to English neural machine translation using the Transformer model which uses a dataset of around 1 million parallel sentences. The system achieves a BLEU score of 0.72 for sentences having a word count of less than 15. A BLEU score of 0.30 was achieved on longer sentences. For sentences less than 15 words, the transformer model achieved a higher BLEU score than Google Translate. They plan to support multiple Indian languages for the English translation task.

AnglaHindi was developed in 2003 by Sinha and Jain [18]. It is an English to Hindi machine translation system. AnglaHindi uses an abstracted example-based approach to get more precise translation for nouns and verbs that are commonly encountered. A partial hybridization of rule-based and example-based methods have been incorporated. Ambiguities are resolved using an appropriate distance function.

The Arabic language exhibits a rich morphology and implementing a machine translation system for this language is quite challenging. Salem *et al.* (2008) introduced a machine translation system called UniArab which is based on the Role and Reference Grammar (RRG) linguistic model [19]. The latter maps the semantic representation of a sentence to its equivalent syntactic representation. The model creates a logical structure from an Arabic sentence. The equivalent sentences in the target language is then generated based on this logical structure. Among the challenges faced while developing the system, the main one is that in Arabic, there are some words which hold the meaning of the entire sentence. Therefore, the translation system should have a robust analysis to realise that this one word is equivalent to a full sentence in English.

Almahairi *et al*. (2016) introduced another Arabic to English neural machine translation system and compared it with a standard phrase-based translation system [20]. The results show that the phrase-based and neural translation systems perform comparably well to each other and that proper preprocessing of Arabic script has the same outcome on both systems. However, the authors conclude that the neural machine translation meaningfully outperforms the phrase-based system on an out-of-domain test making it attractive for real-world deployment.

Costa-Jussà and Banchs (2011) developed a phrase-based statistical Haitian-Creole to English machine translation system which was used as a means of communication by people in Haiti during the 2010 earthquake [21]. This system was ranked first in the 6th Workshop on Statistical Machine Translation in 2011.

The performance of some recently developed models using deep learning for the translation of Japanese to English was examined by Greenstein and Penner [22]. An RNN search model was trained on a subgroup of approximately 150,000 sentence pairs. With little training, the model was able to perform the translations with relatively high accuracy and adapts to grammatical complexity with ease.

Although a lot of work has been done on the translation of popular spoken languages, inadequate attention has been given to machine translation for less popular and under-resourced African languages. By making use of modern neural machine translation techniques, Martinus and Abbott trained several deep learning models to translate English to five official South African languages [23]. For each language, two neural machine translation architectures were used, namely, ConvS2S and Transformer. It was noted that the Transformer model performed better than the ConvS2S model for all languages. The system offered good quality translation for languages like Xitsonga, Setswana and Afrikaans since they had good quality data. However, more work has to be done for languages like isiZulu and Northern Sotho since they had little data.

Xhosa is a language which is spoken by around 16% of the population in South Africa [24]. Mueller and Lal (2019) proposed a sentence-level process to perform the translation of Xhosa to English [25]. The dataset consists of 20544 English and Xhosa parallel sentences for training and 1956 sentences for testing. An attention-based neural machine translation model using transformers as its foundation was used. The sentence adapted models achieve slightly higher BLEU scores compared to standard neural approaches.

Swahili plays an important part in education in several African countries such as Uganda and Kenya [26]. Guy *et al*. (2011) introduced an English to Swahili and Swahili to English statistical machine translation system called SAWA using the Moses package [27]. The dataset consists of more than two million words. The dataset was randomly divided into a 90% training set and a 10% testing set. For English to Swahili translation, the BLEU score for Google Translate was 0.26 while for SAWA it was 0.20. Moreover, for Swahili to English translation, a BLEU score of 0.29 was obtained for Google Translate and 0.35 for SAWA.

The national language of the Philippines is Filipino [28]. Ang *et al*. (2015) developed a system that translates Filipino to English and vice-versa [29]. The dataset contains around 22000 sentences. The system was implemented using a Moses-based statistical machine translation approach. Training was performed using feedback from users for a period of three months. Over the course of 100 training steps, the translation accuracy sharply improved within the first few repetitions, which then gradually decreased due to over-fitting. The highest BLEU score for the English to Filipino translation was 0.34 and 0.39 for the other way.

Oo *et al*. developed a system to translate the Burmese language to the Arakanese language [30]. Arakanese is a language which closely related to Burmese. 14,076 parallel sentences were used for training, 2485 sentences for testing and 1812 for evaluation. Three neural machine translation models were implemented: recurrent neural networks (RNN), transformer, and convolutional neural networks (CNN). The best performance was obtained with CNN. The model achieved a BLEU score of 0.84 for Arakanese to Burmese and 0.81 for the other way.

Coming to Mauritius, the first English to Kreol Morisien machine translation was presented by Pudaruth *et al.* [9]. To perform translation, they used a greedy rule-based approach. First, the text was split into an array of sentences using a greedy algorithm. If the greedy algorithm is not successful, morphological analysis is performed to check for equivalent translation by extracting the root of words. Finally, the words were reordered to match the structure of the target language.

Sukhoo *et al*. (2014) introduced another Kreol Morisien to English machine translation system named *Anou Tradir* [31]. The latter is a statistical machine translation (SMT) system which was developed using a phrase-based approach. The languages involved were English, French, and Kreol Morisien. French was used as a bridge language since it is closer to Kreol Morisien. A 'transfer method' was used to translate English sentences to Kreol Morisien. Sukhoo *et al*. (2014) developed another statistical machine translation with a similar dataset using generic translation models developed by IBM (International Business Machines) [7]. Both systems were not evaluated adequately.

The major issues which were found in some of the systems are the size and the quality of the corpus to train the machine translation system. Furthermore, it can be seen that very little consideration has been given to the translation of the Kreol Morisien language. The existing systems to translate English to Kreol Morisien were implemented using either rule-based or statistical approaches. Also, the Kreol Morisien language used for the creation of the datasets in the existing systems was not standardized. In the last five years, no new recent approaches have been implemented for the translation of Kreol Morisien to English. Thus, we propose to use a neural machine translation approach to translate English to Kreol Morisien and vice-versa. A transformer model with attention will be developed for training the models. The third edition of the Diksioner Morisien (dictionary), which is the most recent one, will be used to learn the Kreol Morisien (KM) language before the creation of the dataset.

5. METHODOLOGY

A. Description of the dataset

Nineteen thousand, sixty hundred and fifty (19,650) English and Kreol Morisien sentences are used for training, another 1000 pairs are used for testing and an unseen set of 1000 sentences is kept aside for validation. The details of the training set are shown in Table 6.

TABLE VI. DATASET

Statistics	English	Kreol Morisien
Number of sentences	19650	19650
Total number of words	146232	143737
Number of unique words	10084	10091
Length of shortest sentence	1	1
Length of the longest sentence	26	27
Average number of words in a sentence	7.4	7.3

B. Why Attention?

Attention is one among the most dominant ideas in the machine learning and deep learning community. The use of attention mechanisms is increasing to improve the performance of neural machine translation by focusing on the sub-parts of the sentence during translation [32]. It was mainly designed for sequence-to-sequence models. In a sequence-to-sequence model, the encoder inputs are discarded. Only the final states are used as input to the decoder. This method works well for short sentences but as the length of the sentence increases, using a single vector becomes a bottleneck. It becomes very difficult to encapsulate longer sentences into a single vector [32].

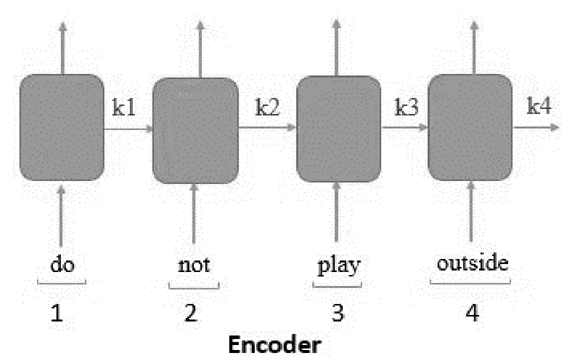

Figure 3. Encoder diagram showing the intermediate states [32]

The intermediate states of the encoder are shown in Figure 3. The states k1, k2, k3 and k4 represents vectors of fixed length. k1 represents the information found in the start of the sequence and k4 stores the information found in the later part of the sequence.

Suppose we want to predict the first word in the target sentence, that is, “pa”. Using the attention mechanism, all the states of the encoder are not required to predict this word. The probability that the data about the word “pa” will be found in the first few states, for example k1 and k2, is high. Therefore, we want the decoder to focus more on the first few states. Since we are predicting the first word of the sentence, there is no current internal state. Hence, the last state of the encoder, that is k4, will be used as the previous decoder state [32]. Using all the encoder states and the current state of the decoder, a simple feed forward neural network can be trained as follows [32]: a score is generated for each state, the attention weight is computed, the context vector is calculated, the context vector is concatenated with the output of the previous time step and finally the output is decoded by the decoder. These steps are shown in Figure 4.

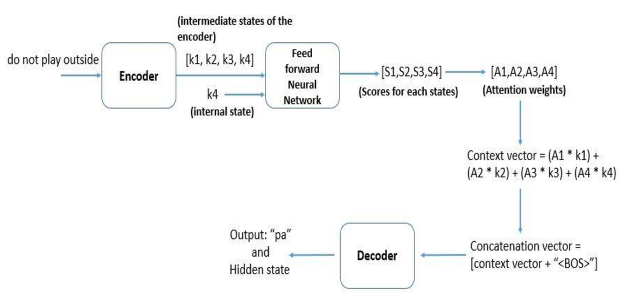

Figure 4. Steps for predicting the first word in a target sentence [32]

Figure 5 illustrates a possible example to show how the weight is given to each hidden state when translating the sentence “do not play outside” to Kreol Morisien. The darker the colour implies that more weight is associated to each word.

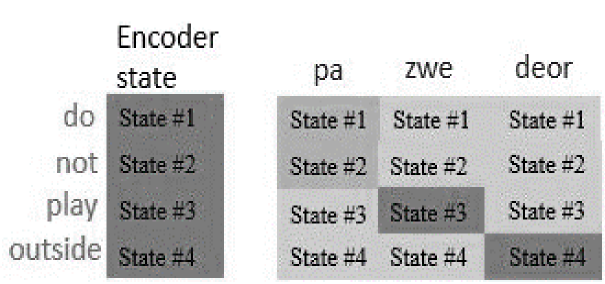

Figure 5. Example to show how the weight is given to each word [32]

C. Transformer model

Figure 6 shows the architecture of an encoder-decoder system. Firstly, the input of the encoder passes through the self-attention layer [33]. The outputs of the self-attention layer are then fed to a feed forward neural network [33]. Self-attention is an *attention mechanism* which relates different positions of a single sequence to produce a complete representation of the sequence [33]. The idea behind using self-attention is explained in Figure 7.

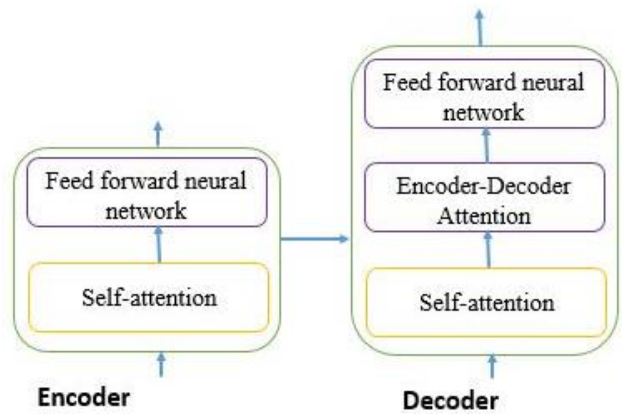

Figure 6. Transformer model [34]

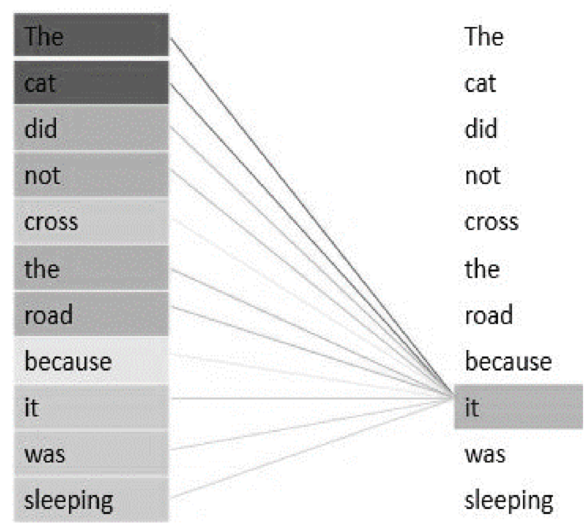

Figure 7. Applying self-attention on a sentence [35]

The transformer uses self-attention to understand relevant words in the sentence. To decode the final sequence, we use the decoder. The latter consists of three layers: the self-attention layer, the feed forward neural network layer and between these two layers is the attention layer.

D. System Model

Figure 8 shows the system model which depicts the different phases in the system development process. The dataset normally contains unstructured data, so cleaning needs to be done before going ahead with the implementation. The cleaning operations done are as follows: all punctuation characters are removed, all Unicode characters are normalised to ASCII, all texts are converted to lowercase and any non-alphabetic tokens are removed.

Next, the tokenizer splits the sentences into a list of words and then converts the latter to integers. Each sequence must have the same length. Since the sentences in the text file are of different lengths, the maximum length of the sentences in the dataset is found and then padding is done at the end of the sentences so that eventually, all the sentences are of the same length.

The models are trained on 19,650 pairs of parallel sentences. For both English to Kreol Morisien and Kreol Morisien to English translation, the models have been trained for 25000 steps, where each thousand step took around one hour to complete.

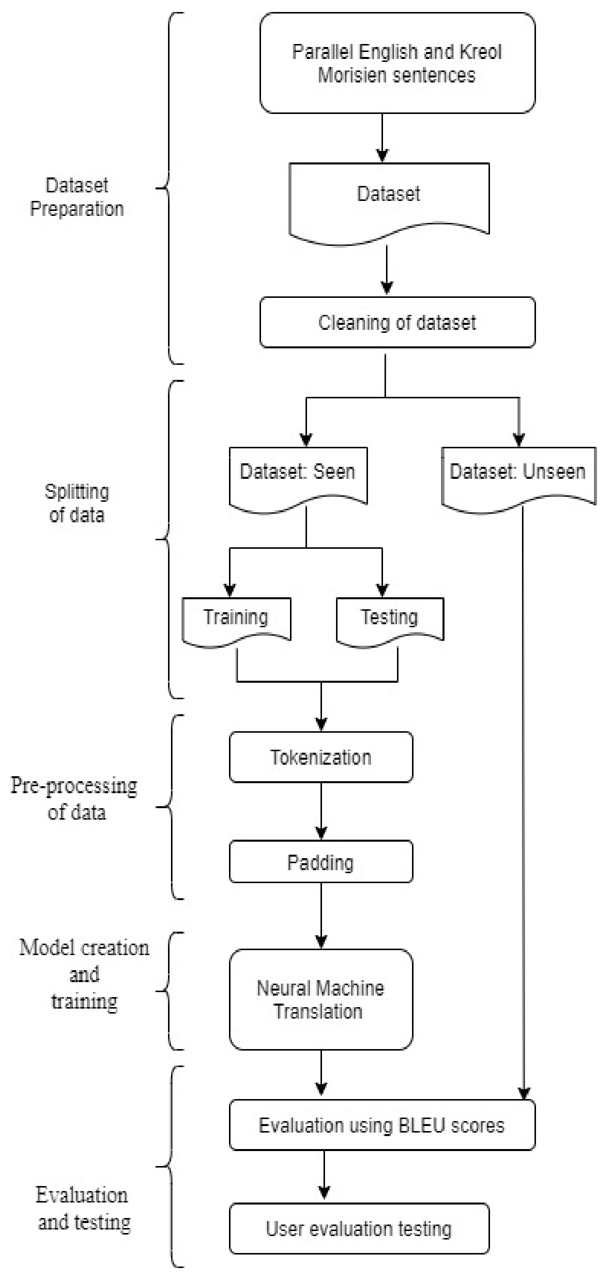

Figure 8. System Design

6. Testing and Evaluation

Evaluation of the quality of a text which has been translated from one language to another can be done using the BLEU algorithm. The BLEU score lies between 0 and 1. A perfect match results in a score of 1, whereas a perfect mismatch results in a score of 0 [36].

Figure 9 shows how the BLEU score varies as the number of training steps increases for the translation of English to Kreol Morisien. Figure 10 illustrates how the BLEU score varies as the number of raining steps increases for the translation of Kreol Morisien to English.

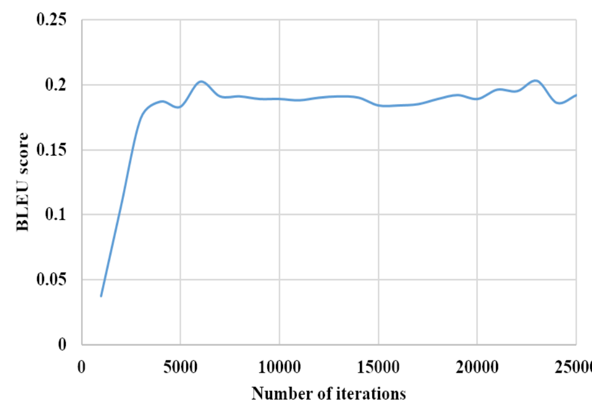

Figure 9. BLEU scores for English to Kreol Morisien translation

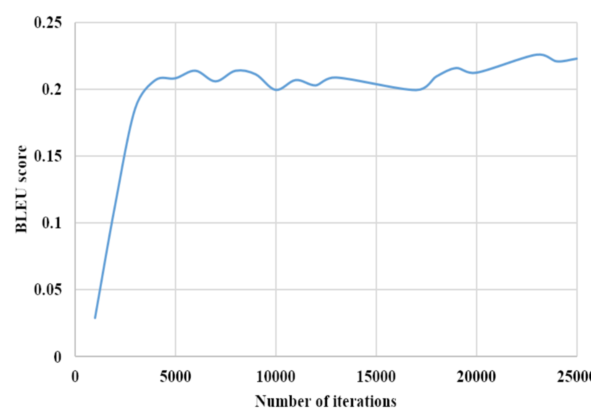

Figure 10. BLEU scores for Kreol Morisien to English translation

The English to Kreol Morisien model achieved a BLEU score of 0.20 and the Kreol Morisien to English model achieved a BLEU score of 0.23.

The examples in Table 7 show how the translation from English to Kreol Morisien improved until it reached the best possible translation. The English equivalent of the Kreol Morisien sentences are also provided. Sentences of varying lengths have been used to show the progress of the translation.

TABLE VII. ENGLISH TO KREOL MORISIEN TRANSLATION

Source sentence 1:	For living the good life.	
Target sentence 1:	Pou viv la bel vi.	
Translated texts:	Enn bon bon. Pou enn bon lavi. Pou viv bon lavi. Pou viv enn bon lavi.	A good good. For a good life. To live good life. To live a good life.
Source sentence 2:	Let us just keep this between ourselves.	
Target sentence 2:	Les nou gard sa zis ant nou.	
Translated texts:	Les nou nou nou nou. Les nou gard sa ant sa mem-la. Les nou fek gard sa ant nou mem. Les nou gard sa ant ou mem. Les nou zis gard sa ant nou mem.	Let us us us us. Let us keep this between this. Let us just keep this between ourselves. Let us keep this between ourselves. Let us just keep this between ourselves.

The examples in Table 7 show how the translation from English to Kreol Morisien improves using sentences of varying lengths until it reaches an acceptable translation.

TABLE VIII. KREOL MORISIEN TO ENGLISH TRANSLATION

Source sentence 1:	A kisann-la to apartenir?
Target sentence 1:	Who do you belong to?
Translated texts:	Who is your favorite? Who did you belong to? Who do you belong to me? Who do you belong to?
Source sentence 2:	Nou lavi kouma enn karne kot bann paz kouver ar tou bann moman.
Target sentence 2:	Our life is like a notebook of which pages are covered with all the moments.
Translated texts:	Our life is like a real of all the pages of everything. Our life is how to be a real pages of all the times. Our life is like a diary where the pages of all kinds. Our lives like a diary in the pages of all the moment. Our life is like a diary where the pages of all the moment.

From Table 8, we can see that there is an improvement in the quality of the translation from Kreol Morisien to English with increase in the number of training steps. The BLEU scores show that the quality of the translation from Kreol Morisien to English is slightly better compared to the translation from English to Kreol Morisien This is because there are only about 20,000 words in Kreol Morisien words while there are more than 200,000 English words. Therefore, translating English is more challenging. Table 9 shows six sentences which has been translated perfectly from English to Kreol Morisien while Table 10 shows another six sentences whose translations are not so good. Table 11 and 12 shows examples for KM to English.

TABLE IX. SIX PERFECTLY TRANSLATED SENTENCES FROM ENGLISH TO KREOL MORISIEN

Source sentence	**Target sentence**	**Translated sentence**
When do you run?	Kan to galoupe?	Kan to galoupe?
I gave my brother a dictionary.	Mo finn donn mo frer enn	Mo finn donn mo frer enn diksioner.
Tom never seems to know what to	Tom zame paret kone ki pou dir.	Zame Tom paret kone ki pou dir.
I am attached to her.	Mo atase avek li.	Mo atase ar li.
He is always changing his mind.	Touletan li sanz so lespri.	Li touletan sanz so lespri.
The mother is certain.	Mama-la sertin.	Mama-la sir.

TABLE X. SIX NOT PERFECTLY TRANSLATED SENTENCES FROM ENGLISH TO KREOL MORISIEN

Source sentence	**Target sentence**	**Translated sentence**
If you pull too hard.	Si to ris tro for.	Si to tro dir.
The park is open to everybody.	Park-la ouver pou tou dimounn.	Park la ouver tou dimounn.
Where swimming is concerned.	Kot naze konserne.	Kot naze.
First cousins are too close for marriage.	Bann premie kouzin tro pros pou marye.	Dabor de kouzin pou maryaz.
Father bought me a new bicycle.	Papa ti aste enn nouvo bisiklet pou	Papa ti aste enn nouvo bisiklet.
Mother was busy cooking, in the meantime.	Mama ti okipe kwi, antretan.	Mama ti okipe nek kwi..lave enn violasion parke.

TABLE XI. SIX PERFECTLY TRANSLATED SENTENCES FROM KREOL MORISIEN TO ENGLISH

Source sentence	**Target sentence**	**Translated sentence**
A kisann-la to apartenir?	Who do you belong to?	Who do you belong to?
Mo mama pa koz angle.	My mother does not speak English.	My mother does not speak English.
Mo pou telefonn li dime.	I will call him tomorrow.	I will call him tomorrow.
Mo kapav esey li?	Can i try it on?	May I try it on?
Konpare ar li.	In comparison to him.	Compared with him.
Lafin so lavi.	At the end of his life.	The end of his life.

TABLE XII. SIX NOT PERFECTLY TRANSLATED SENTENCES FROM KREOL MORISIEN TO ENGLISH

Source sentence	**Target sentence**	**Translated sentence**
Si to ris tro for.	If you pull too hard.	If you are too rich.
Kot naze konserne.	Where swimming is concerned.	Where is concerned.
Sa garson-la pa manze.	That boy does not eat.	This boy is not food.
Li ti deza kontan li.	She was already in love with him.	It was already to him.
Bann vas sakre pou bann indou.	Cows are sacred to hindus.	The sacred ritual took action subsequent.
Mo espere mo ti kone kouma mo kapav konvink tom.	I wish i could figure out how to convince tom.	I wish I could figure out how to convince us.

To evaluate the quality of the translated texts by potential users, two Google Forms were created, one for English to Kreol Morisien translation and another one for Kreol Morisien to English translation. The surveys consist of 25 questions each. Each question was a pair of sentences which consist of a sentence in the source language and its translation in the target language. The translation was performed by our proposed system. The respondents were familiar with both languages. Responses were gathered from different age groups, from rural and urban regions and from both students and professionals. The sentences chosen were of varying lengths and from perfect translation to very poor translation. The survey was kept open for one week where 92 respondents filled the English to Kreol Morisien survey form while 103 respondents filled the other form. Figure 11 and 12 show the result of a random question from each survey.

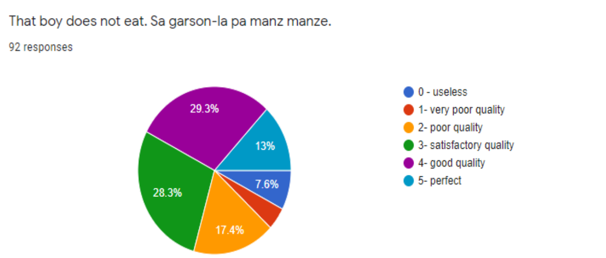

Figure 11. Result of a question in the English to Kreol Morisien survey

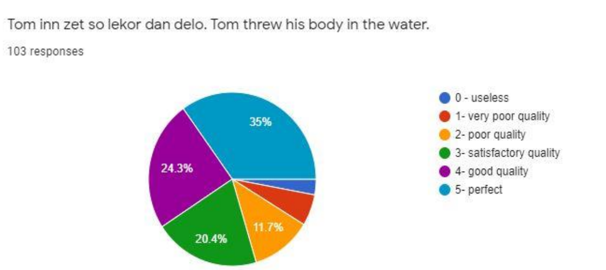

Figure 12. Result of a question in the Kreol Morisien to English survey

The average score for the English to Kreol Morisien survey was 3.8/5 and that of the Kreol Morisien to English survey was 3.9/5. While some have rated the translation based on how much meaning is preserved in the source sentence and the target sentence, others have rated it based on criteria like grammar, spelling and choice of words. Therefore, assessing the quality of translation is not an easy task. Similar to the results from the BLEU scores, the quality of the Kreol Morisien to English translated was rated as slightly better than that of the English to Kreol Morisien translations.

Compared to the first English to Kreol Morisien machine translation presented by Pudaruth et al. (2013) which had the ability to perform translation only with short sentences, our system further improves on this by its ability to translate much longer sentences with a higher degree of readability [9].

Anou Tradir, a statistical machine translation system developed by Sukhoo et al. (2014) used a dataset consisting of 13,795 parallel sentences which is much less than the one used in this work [31]. The final BLEU score obtained in Anou Tradir is 0.08 for the English to Kreol Morisien and 0.12 for the other way, which is much less than the ones obtained in our system. Furthermore, the dataset used was not standardized and was not up to date with respect to the standard way of writing Kreol Morisien. Sukhoo et al. (2014) developed another statistical machine translation system using the same dataset but this time using generic translation models developed by IBM but the BLEU scores were found to be within the same ranges as Anou Tradir [7]. This indicates that our system is an improvement over the former systems, with a larger dataset and using the standard Kreol Morisien language.

7. Conclusion

Kreol Morisien (KM) is a language that is predominantly spoken only in the Republic of Mauritius. It is an under-resourced language because the number of people speaking this language in the world is less than 1.5 million. Furthermore, although the Kreol Morisien language has been the lingua franca for centuries, it gained a national status only in the year 2012 when it was introduced as a formal language in the educational system. To our knowledge, there are currently no reliable automatic translation services for translating Kreol Morisien into English or vice-versa. Thus, in this work, we have implemented an automatic system for translating Kreol Morisien into English and English into Kreol Morisien using a neural machine translation approach based on the concepts of deep learning, attention and the transformer model. The English to Kreol Morisien translation model achieved a BLEU score of 0.20 while the Kreol Morisien to English model achieved a BLEU score of 0.23. The translation of KM to English is of slightly better quality because the vocabulary of KM is much less than that of English and therefore it is easier for the program to understand and translate a KM word than an English word. The user acceptance test carried out by means of online surveys showed that the quality of KM to English translation was rated at 78% while the score was 76% for English to KM translation. These numbers are relatively high showing that even with a moderate training set of 19,650 parallel sentences, our proposed models are able to produce reliable translations. In the future, we intend to increase the size of our dataset by a considerable amount as this is the primary means of improving the quality of the translated texts. These translation models can also be deployed online via a web portal to allow the general public to use them and provide their feedback.

REFERENCES

[1] Statista. (2019, Oct. 5). *Global digital population*. [Online]. Available: https://www.statista.com/statistics/617136/digital-population-worldwide/

[2] D. Lyons. (2019, Sep. 9). *Babbel Magazine: How Many People Speak English And Where Is It Spoken?* Available: https://www.babbel.com/en/magazine/how-many-people-speak-english-and-where-is-it-spoken/

[3] A. Ah-Vee. (2019, Sep. 6). *The state of the mother tongue in Mauritius today*. [Online]. Available: https://www.lalitmauritius.org/en/newsarticle/2192/the-state-of-the-mother-tongue-in-mauritius-today/

[4] S. Quirin, 2012. Kreol Morisien au Primaire (Kreol Morisien in Primary). *Le Mauricien*, 15 Jan, p17.

[5] Mauritius Examinations Syndicate. (2019, Oct. 5). *PSAC Statistics*. [Online]. Available: http://mes.intnet.mu/English/Statistics

[6] Montray Kréyol. (2009, July. 14). Le Moteur de Recherche Google en Kreol Morisien (The Google Search Engine in Kreol Morisien). [Online]. Available: https://www.montraykreyol.org/article/le-moteur-de-recherche-google-en-kreol-morisien

[7] A. Sukhoo, P. Bhattacharyya, and M. Soobron, "Translation between English and Mauritian Creole: A statistical machine translation approach," in: Proc. *IST-Africa Conference Proceedings*, Mauritius, May 6-9, 2014, pp. 1-10. doi: 10.1109/ISTAFRICA.2014.6880635

[8] G.H.C. Lin and P.S.C. Chien, "Machine Translation for Academic Purposes," in: *Proceedings of the International Conference on TESOL and Translation, Taiwan*, December 2009, pp. 133-148.

[9] S. Pudaruth, L. Sookun, and A.K. Ruchpaul, "English to Kreol and Kreol to English Rule Based Machine Translation System," International Journal of Advanced Computer Science and Applications (IJACSA), vol. 4, no. 8, pp. 25-29, 2013.

[10] A. Geitgey, (2019, Sep. 21). *Machine Learning is Fun Part 5: Language Translation with Deep Learning and the Magic of Sequences.* [Online]. Available: https://medium.com/@ageitgey/machine-learning-is-fun-part-5-language-translation-with-deep-learning-and-the-magic-of-sequences-2ace0acca0aa

[11] T. Mitamura, E. Nyberg, and J.G. Carbonell, "Automated Corpus Analysis and the Acquisition of Large, Multi-Lingual Knowledge Bases for MT," in*: Proceedings of TMI-93, Kyoto, Japan*, July 14-16, 1993.

[12] R. Miikkulainen, J. Liang, E. Meyerson, A. Rawal, D. Fink, O. Francon, B. Raju, H. Shahrzad, A. Navruzyan, N. Duffy and B. Hodjat, "Evolving deep neural networks," in *Artificial Intelligence in the Age of Neural Networks and Brain Computing*. Massachusetts: Academic Press, 2018, ch. 15, pp. 293-312.

[13] H. Jabbar and R.Z. Khan, "Methods to avoid over-fitting and under-fitting in supervised machine learning (comparative study)," in *Computer Science, Communication and Instrumentation Devices*, 2014, pp. 163-172. doi: 10.3850/978-981-09-5247-1_017

[14] R. Sennrich, and B. Haddow, "Linguistic Input Features Improve Neural Machine Translation," in: *Proceedings of the First Conference on Machine Translation*, Berlin, Germany, August 2016. pp. 83–91.

[15] P. Sharma, (2020, Feb. 10). *Must-Read Tutorial to Learn Sequence Modeling*. [Online]. Available: https://www.analyticsvidhya.com/blog/2019/01/sequence-models-deeplearning/

[16] S. Abeywardana, (2020, Jan. 25). *Sequence to sequence tutorial*. [Online]. Available: https://towardsdatascience.com/sequence-to-sequence-tutorial-4fde3ee798d8

[17] S.A. Jadhav, "Marathi To English Neural Machine Translation with Near Perfect Corpus and Transformers," *arXiv preprint arXiv:2002.11643*, Feb. 2020.

[18] R.M.K. Sinha, and A. Jain, 2003. AnglaHindi: An English to Hindi machine-aided translation system," in: *Proceedings of the 9th MT Summit, MTArchive, New Orleans*, *USA*, Sept 23-27, 2003, pp. 494-497.

[19] Y. Salem, A. Hensman, and B. Nolan, "Implementing Arabic-to-English Machine Translation using the Role and Reference Grammar linguistic model," in: *Proceedings of the 8th Annual Conference on Information Technology and Telecommunication*, *Galway, Ireland*, 23-24 October, 2008, pp. 103-110.

[20] A. Almahairi, K. Cho, N. Habash, and A. Courville, "First result on Arabic neural machine translation," *arXiv preprint arXiv:1606.02680*, 2016.

[21] M.R. Costa-Jussà, and R.E. Banchs, 2011. "The BM-I2R Haitian-Créole-to-English translation system description for the WMT 2011 evaluation campaign," in: *Proceedings of the Sixth Workshop on Statistical Machine Translation, Edinburgh, Scotland, UK,* July 30–31, 2011, pp. 452-456.

[22] E. Greenstein, and D. Penner, "Japanese-to-English Machine Translation using Recurrent Neural Networks," Stanford University, June 2015.

[23] L. Martinus, and J.Z. Abbott, "A Focus on Neural Machine Translation for African Languages," *arXiv preprint arXiv:1906.05685*, Jun. 2019.

[24] Sa-v, (2020, March 8). *Xhosa language*. [Online]. Available from: https://www.sa-venues.com/language-xhosa.htm

[25] A. Mueller and Y.K. Lal, "Sentence-Level Adaptation for Low-Resource Neural Machine Translation," In: *Proceedings of the 2nd Workshop on Technologies for MT of Low Resource Languages*. *Dublin, Ireland,* August 19-23, 2019, pp, 39-47.

[26] M. Price. (2020, March. 8). *How to Learn Swahili: An In-Depth Guide*. [Online]. Available from: https://www.fluentin3months.com/swahili/

[27] D. P. Guy, P.W. Wagacha, and D.S. Gilles-Maurice, "Exploring the SAWA corpus: collection and deployment of a parallel corpus English-Swahili," *Language Resources and Evaluation*, vol. 45, no. 3, pp. 331-344, 2011.

[28] G. Borlaza, C. Hernandez, M. Cullinane, (2020, March. 8). *Philippines.* [Online]. Available from: https://www.britannica.com/place/Philippines

[29] J. Ang, M.R. Chan, J.P. Genato, J. Uy and J. Ilao, "Development of a Filipino-to-English Bidirectional Statistical Machine Translation System that dynamically updates via user feedback," In: *Proceedings of the 12th International Workshop on Spoken Language Translation, Da Nang, Vietnam*, December 3-4, 2015.

[30] T.M. Oo, Y.K. Thu, and S, K. M, "Neural Machine Translation between Myanmar (Burmese) and Rakhine (Arakanese)," in: *Proceedings of the Sixth Workshop on NLP for Similar Languages, Varieties and Dialects, Minneapolis, USA,* June 7, 2019, pp. 80-88.

[31] A. Sukhoo, R. Dabre, and P. Bhattacharyya, "Anou Tradir: Experiences in Building Statistical Machine Translation Systems For Mauritian Languages–Kreol, English, French," in: *Proceedings of the 11th International Conference on Natural Language Processing, Goa, India,* December 18-21, 2014, pp. 82-88.

[32] B. Sankaran, H Mi, Y. Al-Onaizan, and A. Ittycheriah, "Temporal attention model for neural machine translation," *arXiv preprint arXiv:1608.02927*, Aug. 2016.

[33] A. Bhandare, V. Sripathi, D. Karkada, V. Menon, S. Choi, K. Datta, and V. Saletore, "Efficient 8-bit quantization of transformer neural machine language translation model," *arXiv preprint arXiv:1906.00532*, Jun. 2019.

[34] J. Alammar, (2020, Jan. 25). *The Illustrated Transformer*. [Online]. Available: http://jalammar.github.io/illustrated-transformer/

[35] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser and I. Polosukhin, "Attention is all you need," in: *Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA*, December 4-9, 2017, pp. 6000–6010.

[36] K. Papineni, S. Roukos, T. Ward, and W.J. Zhu, "BLEU: a method for automatic evaluation of machine translation," in: *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, Philadelphia, Pennsylvania, USA*, July 7-12, 2002, pp. 311-318.

Zaheenah Boodeea is currently a final year student studying the BSc (Hons) Computer Science programme at the Department of Information and Communication Technologies, Faculty of Information, Communication and Digital Technologies, University of Mauritius. As part of the compulsory Industrial Training module, she had the opportunity to work as an intern at Ceridian Mauritius Ltd for period of 10 weeks from May 2019 – August 2019, where she refined her skills as a programmer as well as her communication & social skills. Her research interests include Mobile Application Development, Web Technologies, Database Management Systems, Machine Learning, Machine Translation, Deep Learning, and Artificial Intelligence. She is delighted to see her work being published in the International Journal of Computing and Digital Systems, which is published by the University of Bahrain and is indexed by Scopus. Currently, IJCDS has a CiteScore of 0.32. She is looking forward to graduate soon from the University in the coming months and she intends to pursue a Master's course in the field of Artificial Intelligence or Data Science.

Sameerchand Pudaruth is a Senior Lecturer and Head of the ICT Department at the University of Mauritius. He is a member of IEEE, founding member of the IEEE Mauritius Subsection and the current Vice-Chair of the IEEE Mauritius Section. He is also a member of the Association for Computing Machinery (ACM). His research interests are Artificial Intelligence, Machine Learning, Data Science, Machine Translation, Computer Vision, Robotics, Mobile Applications, Web Technologies, Multimedia, Blockchain and Information Technology Law. He has written more than 50+ papers for national & international journals and conferences. He has been in the organising committee of many successful international conferences such as Africon 2013, IST Africa 2014, Africon 2015, ICCCS 2015, BigData 2015, DIPECC 2015, Africhi 2016, Emergitech 2016, Nextcomp 2017, ISCMI 2017, Mauritian Academic Conference 2018, icABCD 2018, icABCD2019, icABCD2020, Mauricon ICONIC 2018 and Mauricon ICONIC 2020. He has also written a book entitled, 'Python in One Week'. He is a reviewer for IEEE Access and Environment, Systems and Decisions (Springer) journals, amongst many others.