A RESEARCH ON VIETNAMESE-K'HO LANGUAGE TRANSLATION SYSTEM USING NEURAL MACHINE TRANSLATION

Nguyen Thi Luong*, La Quoc Thang, Tran Nhat Quang, Duong Bao Ninh, Nguyen Huu Khanh, Phan Thi Thanh Nga, Tran Ngo Nhu Khanh, Tran Thong
Dalat University

ARTICLE INFO

Received: 29/10/2022
Revised: 31/3/2023
Published: 07/4/2023

KEYWORDS

K'Ho language
Bilingual Corpus
Automatic translation
RNN
OpenNMT

ABSTRACT

The K'Ho language is used by the K'Ho ethnic group, who live in the South Central Highlands, especially the districts of Don Duong, Duc Trong, Di Linh, Da Huoai, and Lac Duong in Lam Dong province. Currently, the provincial People's Committee and the Ethnic Minority Committee of Lam Dong province are encouraging cadres and officials in the province to learn the K'Ho language to contact and propagate the guidelines, lines, policies, and laws of the Party and government to the K'Ho people. In this paper, we utilize the K'Ho language resources and support from many K'Ho language experts to build a Vietnamese - K'Ho bilingual corpus to contribute the promotion and preservation of the K'Ho language. The corpus includes more than 16,000 Vietnamese-K'Ho bilingual sentence pairs, which are not easy to collect due to the limitation of K'Ho language resource. Moreover, we use the OpenNMT framework to build an automatic translation system based on the collected bilingual data. The result can reach to an accuracy of 56.54%, which is an acceptable result in the automatic translation field.

NGHIÊN CỨU HỆ THỐNG DỊCH NGÔN NGỮ TIẾNG VIỆT-K'HO SỬ DỤNG DỊCH MÁY BẰNG NƠRON

Nguyễn Thị Lương*, La Quốc Thắng, Trần Nhật Quang, Dương Bảo Ninh, Nguyễn Hữu Khánh, Phan Thị Thanh Nga, Trần Ngô Như Khánh, Trần Thống
Trường Đại học Đà Lạt

THÔNG TIN BÀI BÁO

Ngày nhận bài: 29/10/2022
Ngày hoàn thiện: 31/3/2023
Ngày đăng: 07/4/2023

TỪ KHÓA

Ngôn ngữ K'Ho
Ngữ liệu song ngữ
Dịch tự động
RNN
OpenMNT

TÓM TẮT

Ngôn ngữ K'Ho là ngôn ngữ được sử dụng bởi dân tộc K'Ho, sinh sống ở vùng Nam Tây Nguyên, đặc biệt là các huyện Đơn Dương, Đức Trọng, Di Linh, Đạ Huoai, Lạc Dương thuộc tỉnh Lâm Đồng. Hiện nay, ủy ban nhân dân tỉnh và ban dân tộc tỉnh Lâm Đồng đang khuyến khích cán bộ và viên chức trong tỉnh biết tiếng K'Ho để tiếp xúc và tuyên truyền các chủ trương, đường lối, chính sách, pháp luật của Đảng và Nhà nước tới người K'Ho. Trong bài báo này, chúng tôi sử dụng nguồn tài nguyên tiếng K'Ho và sự hỗ trợ từ các chuyên gia tiếng K'Ho để xây dựng bộ song ngữ Việt – K'Ho nhằm góp phần vào việc quảng bá và bảo tồn ngôn ngữ K'Ho. Bộ ngữ liệu bao gồm hơn 16.000 cặp câu song ngữ Việt-K'Ho, vốn không dễ dàng thu thập do giới hạn về nguồn tài liệu liên quan tới ngôn ngữ tiếng K'Ho. Chúng tôi sử dụng bộ mã nguồn OpenNMT để xây dựng hệ thống dịch tự động dựa trên bộ dữ liệu song ngữ. Kết quả dịch có thể đạt được độ chính xác lên tới 56,54%, là một kết quả có thể chấp nhận được trong lĩnh vực dịch tự động.

DOI: https://doi.org/10.34238/tnu-jst.6818

* Corresponding author. *Email: luongnt@dlu.edu.vn*

1. Introduction

In recent years, deep learning-based natural language processing, especially in machine translation studies, has achieved outstanding results. Until now, there exist many researches about automatic translation of text from Vietnamese to English and vice versa [1]. Machine translation from English to Vietnamese is based on a method with low efficiency [2]. In [3], [4] the authors translated English - Vietnamese documents with IWSLT data [5] which used more data and combined complex models to improve translation efficiency. Phan-Vu et al. [6] created a dataset for the English-Vietnamese bilingual translation and used the seq2seq method to achieve a high BLEU index. Le and Nguyen [1] used the Transformer architecture with a BLEU index of 32.1.

Currently, several domestic studies devoted to bilingualism between Vietnamese and ethnic minority languages to help preserve of languages and improve the understanding of ethnic groups about other cultures as well as regulations and policies of the government. The K'Ho language, which is widely used in Lam Dong province, was studied by Dinh et al. in [7], [8]. The authors built a bilingual dataset based on weather forecasts on Lam Dong radio in 2014 which had over 1,300 pairs of bilingual sentences. They applied statistical machine translation [7] and sample-based translation [8] to translate from Vietnamese into the K'Ho language. However, the study only used data on a narrow range of weather forecasts, so it could not be applied to other documents. Therefore, in this paper, we enhance the scope of our automatic translation system to cover not a specific field but different ones of K'Ho group using the dictionary-based format.

1.1. Motivation

Lam Dong is a province in the South Central Highlands with a big natural area and many district-level administrative units. Among 47 ethnic groups, some minorities account for a high proportion such as K'Ho, Ma, Churu, Nung, Tay, Hoa, and M'Nong. Most of the ethnic groups in Lam Dong live side by side with the spirit of solidarity, equality, respect, and mutual development. To contribute to the study of languages of ethnic minorities using modern technologies and to provide an automatic translation system from Vietnamese to K'Ho and vice versa, in this paper, we aim to build a Vietnamese - K'Ho bilingual corpus, then we use the deep learning methods to build an automatic translation system, which improve the understanding between the two languages.

1.2. Literature Survey

1.2.1. Vietnamese – K'Ho translate

In [7], [8], Dinh et al. studied a Vietnamese – K'Ho translation system for weather forecasting. In [7], they applied the example-based machine translation method while they used a Statistics Machine Translation-based application in [8]. Experimental data involved 212 pairs of Vietnamese - K'Ho bilingual sentences extracted from the weather forecast bulletins from 2015 to 2017 of Lam Dong Radio and Television Station, Lam Dong newspaper, and Voice of Vietnam. Both studies only used small data about 212 Vietnamese-K'Ho bilingual sentences and did not evaluate the accuracy of the system. Moreover, the research was limited to the field of weather forecasting.

1.2.2. RNN in the translation system

Luong et al. [9] proposed two RNN attentional mechanisms for NMT, in which the global approach looked at all words, and the local one attended to a subset at once, to enable higher learning efficiency and higher translation quality. The global model was more extensive than the other and was impractical with long sequences, such as paragraphs and documents. Hence, the local attentional approach was considered a better alternative.

Convolutional Sequence to Sequence Learning (ConvS2S) [10] is a fully convolutional neural network for NMT. Gehring et al. applied CNN which is less popular for NMT due to the lack of capability for long time-series data. The authors incorporated gated linear units into the encoder to optimize the gradient propagation. They also equipped a separate attention module in each decoder layer. Thanks to these optimizations, ConvS2S outperformed GNMT [11], one of the best RNN models for NMT, on WMT'14 English-German and WMT'14 English-French datasets.

Radford et al. [12] proposed a transformer-based language model called GPT-2 with 1.5 billion parameters. Based on the ConvS2S architecture, the authors proposed a new network by modifying their previous work [13] with additional layer normalizations and modified residual paths. GPT-2 trained on specific domains which was not the same with training datasets, thus, it outperformed others method. It also showed a high capacity for several tasks without explicit supervision needs.

2. Proposed methodology

In this paper, the proposed method includes two main parts: the first part is building a Vietnamese - K'Ho bilingual corpus and the second one is researching a deep learning-based approach to build an automatic translation system.

2.1. Viet – K'Ho bilingual corpus

2.1.1. Overview

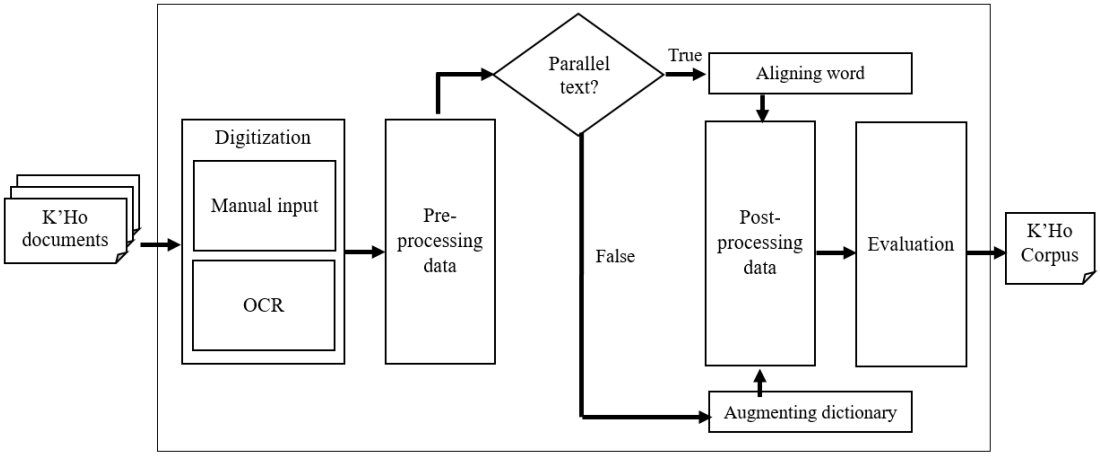

Figure 1. *The flow diagram for bilingual data processing*

To build a Vietnamese - K'Ho bilingual corpus, we proposed a procedure for acquiring, processing, and synthesizing a dictionary from Viet – K'Ho bilingual texts. Figure 1 shows the flow diagram for bilingual data processing. We collected bilingual texts from different K'Ho documents, including K'Ho teaching and learning materials, a dictionary, and issues from Ethnic and Mountainous Photojournalism.

2.1.2. Digitization and preprocessing

Language documents can be digitized using optical character recognition (OCR) or manually input. Using OCR will speed up the digitization of documents that are high-definition images and easy-to-see text. However, many documents are primarily old, so they must be manually input.

To type such special characters as in Figure 2, we need to use the software called TayNguyenKey [14]. TayNguyenKey is a new solution for typing characters for the ethnic minority in Central Highlands. Using this software, everyone can type words for many ethnic

minorities including Rhade, Jarai, Bana, Sedang, Mnong, and K'Ho in different applications such as Microsoft Word, Microsoft Excel, MS Access, etc.

ŏ Ŏ ĕ Ĕ ŭ Ŭ ŏ Ŏ ŭ Ŭ
ă Ă ĕ Ĕ ŏ Ŏ
ñ Ñ ĭ Ĭ ƀ Ƀ ƀ Ƀ č Č

Figure 2. *Some characters for the ethnic minority in the Central Highlands*

After data digitization, the input should be preprocessed to find and correct invalid data, such as spelling, editing, and encoding errors before being applied to the next steps. The characters in Figure 2 use composite Unicode encoding, so they must be converted to the relative precomposed Unicode to limit errors that occur when aligning words in the next step.

2.1.3. Dictionary augmentation and word alignment

It is possible to augment data by the two methods as follows:

Separated by commas: If a Vietnamese word in the dictionary has more than one meaning and a comma separates it, then it can be separated into a new pair of words and meanings. The examples are presented in Table 1 and Table 2.

Table 1. *Original dictionary*

Vietnamese word	K'Ho meaning	Vietnamese example	K'Ho example
Bắc	Bá, bá yung	Bắc cây qua suối	Bá yung tiá dà
Bần bật	Rơm, mprơm	Run bần bật vì rét	Mprơm nuát kòp
Bật	Tơrglơs, gơ tơrglơs	Bật lên	Tơrglơs guh

Table 2. *Dictionary after separating by commas*

Vietnamese word	K'Ho meaning	Vietnamese example	K'Ho example
Bắc	Bá		
Bắc	Bá yung	Bắc cây qua suối	Bá yung tiá dà
Bần bật	Mprơm	Run bần bật vì rét	Mprơm nuát kòp
Bật	Tơrglơs	Bật lên	Tơrglơs guh

Table 3. *Dictionary after combining Vietnamese synonyms*

Vietnamese word	K'Ho meaning	Vietnamese example	K'Ho example
Bẩn thỉu	Bớ bơl	Chân tay bẩn thỉu	Jơng tê bớ bơl
Bê bết	Bớ bơl	Quần áo bê bết đất	Ồi àu bớ bơl ù
Nhem nhuốc	Bớ bơl		

Table 4. *Dictionary after combining Vietnamese synonyms*

Vietnamese word	K'Ho meaning	Vietnamese example	K'Ho example
Bẩn thỉu	Bớ bơl	Chân tay bẩn thỉu	Jơng tê bớ bơl
Dơ dáy	Bớ bơl		
Rếch rác	Bớ bol		
Bê bết	Bớ bơl	Quần áo bê bết đất	Ồi àu bớ bơl ù
Be bét	Bớ bơl		
Nhem nhuốc	Bớ bơl		

Combining Vietnamese synonyms: If a Vietnamese word in the Viet-K'Ho dictionary has synonyms in the Vietnamese synonyms dictionary [15], each corresponding synonym will be alternated in the dictionary to form a new pair of word-meaning. Table 3 and Table 4 show examples of combining Vietnamese synonyms. However, after augmenting the dictionary, there

is no need to include examples in new word pairs because the example is not suitable for the context of the new word. The data obtained from the bilingual texts cannot be used in training natural language processing models and must be aligned with the corresponding translation. Bilingual texts from Ethnic and Mountainous Photojournalism have almost been aligned at the paragraph level and can be aligned at the sentence level and word level as presented in Figure 3.

Figure 3. *Alignment levels in bilingual text*

Word tokenization is a process of splitting a sentence into compound words or phrases, also known as tokens. This process is built in many Vietnamese natural language processing libraries such as Python Vietnamese Toolkit (PyVi), VnCoreNLP, etc. In K'Ho texts, the Vietnamese words that do not have corresponding words in K'Ho will keep the same words. K'Ho words will be splitted by Unicode character changes (i.e. spaces). A word in the source language can be associated with one or more words in the target language. In the word aligners, words are separated based on spaces by default, so combining the Vietnamese word tokenization with the word aligner is necessary for more efficient execution and accurate results. Word aligners can be classified by different approaches, including IBM model-based [16], word class [17], and neural networks [18].

2.2. Deep learning-based approach for the automatic translation system

Neural machine learning translation (NMT) is a method for machine translation that uses an artificial neural network to learn and translate via the likelihood of a word sequence prediction. NMT has been the widely-adopted machine translation approach since 2016. Generally, NMT models the entire machine translation process via a deep neural network consisting of an encoder and a decoder (Figure 4). Here, words are transcribed into vectors in the encoding and decoding process, in which each vector has a unique magnitude and direction. Specifically, the translator analyzes input text and then encodes it into vectors. The decoder takes these vectors and predicts the likely correct translation. This approach helps NMT derive more natural and more human-liked results than other legacy forms of machine translation, such as statistical machine translation.

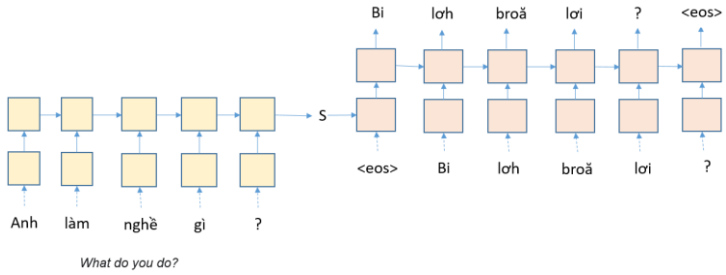

Figure 4. *Illustration of a general NMT Vietnamese – K'Ho*

OpenNMT [19] is an open-source of NMT which was introduced in 2016 by Harvard University and SYSTRAN. This framework aims to provide open-source code for the core translation task, prioritize training and test efficiency, maintain model modularity and readability, and support significant research extensibility. Unlike some NMT algorithms which are closed source, many other NMT systems, such as GroundHog, TensorFlow-seq2seq, and seq2seq-attn, are released as research code that provide minimal support to users. OpenNMT tackles this limitation by helping academic and industrial communities to use the existing NMT implementations easier.

OpenNMT provides implementations on two popular deep-learning frameworks, PyTorch (OpenNMT-py) and TensorFlow (OpenNMT-tf). Benefitting from the used libraries, the

OpenNMT-py is user-friendly and multimodal, and the OpenNMT-tf is modular and stable. Despite the differences in implementation, advantages, and drawbacks, both open-source implementations are highly configurable model architectures and training procedures and scalable to other tasks such as text generation, image-to-text, and speech-to-text. They also provide efficient model-serving capabilities for use in real-world applications. These open-source projects provide many state-of-the-art NMT algorithm implementations, which are presented in Table 5 as well as support many related features in model configuration, training, and decoding.

Table 5. *The list of the OpenNMT supported models*

	OpenNMT-py	OpenNMT-tf
ConvS2S [10]	✓	
DeepSpeech2 [20]	✓ (v1 only)	
GPT-2 [12]	✓	✓
Im2Text [21]	✓ (v1 only)	
Listen, Attend and Spell [22]		✓
RNN with attention [9]	✓	✓
Transformer [23]	✓	✓

3. Experiments

In this paper, we used ABBYY FineReader to digitize K'Ho documents into document files and saved them in spreadsheet files, PyVi to tokenize Vietnamese words, and Eflomal [16] and AWESOME [18] to align bilingual sentences. The result of the bilingual corpus is shown as follows: The total number of words in the dictionary is 8,723; the total number of words after augmenting by the dictionary is 19,864. The total number of bilingual sentences is 16,217. Table 6 shows the results of using Eflomal and AWESOME to align bilingual sentences. The rate is only for the bilingual corpus's evaluation and does not reflect that the others in the bilingual corpus are wrong. More specifically, the result of aligned words in Eflomal is pointed out in Table 7. Figure 5 displays the correlation of bilingual corpus.

Table 6. *Bilingual corpus result*

	Eflomal	**AWESOME**
Number of aligned words	38,200	78.218
Number of aligned words which are similar to words in the dictionary	1,080	674
Rate	2.83%	0.86%

Table 7. *The result of aligned words in Eflomal*

Vietnamese word	**K'Ho word**	**Occurrences**
lễ hội	lèh chờ	203
bản	ɓòn	202
địa phương	anih ơm	199
khó khăn	kal ke	198
chăm	prum	197

3.1. Experiment setup

In this experiment, we create a bilingual dataset consisting of Vietnamese and K'Ho pairs of sentences and use it for training and validation. The training set takes 12,290 pairs of sentences, and the validation test takes 3,927. Our experiment uses the OpenNMT framework for implementation and follows its configuration for building vocals, training, and translating. For training, we set the maximum truncate source sequence length to 400, the truncate target sequence length to 100, the word embedding size to 128, and the size of RNN hidden states to

512. We apply similar configurations for validation, except for the maximum truncate source sequence length and truncate target sequence length, which are set to 1200 and 300, respectively.

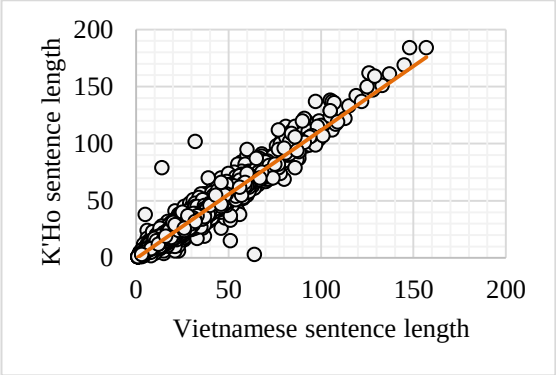

Figure 5. *The correlation of bilingual corpus*

3.2. Deep learning-based methods comparison

This experiment compares five NMT approaches, combining different encoders, decoders, and attention models. The encoder set includes a bidirectional encoder (brnn), a simple RNN encoder (rnn), and a convolutional encoder (cnn) [10]. Here, the brnn comprises two independent encoders, in which one encodes the normal sequence and the other processes the reversed one. The decoder's set includes an RNN decoder (rnn_dec) [9] and a convolutional decoder (cnn_dec) [10]. The attention model set includes two options, from [9] and [24]. Table 8 describes the comparison of several NMT methods. Method (2), which is the combination between the bidirectional encoder and the decoder proposed by Luong et al. with the number of layers in the encoder and decoder of one, achieves the lowest perplexity with 20.34, the highest accuracy with 56.54%, and the highest BLEU with 41.45 among the state-of-the-art on our bilingual datasets. Besides, by increasing the depth of the network as (3), the translation accuracy reduces by 8%. With the small size of our bilingual dataset, the bidirectional approach with the depth of one shows fine capability, although it is not practical enough.

Table 8. *Comparison between various methods, in which the best values are highlighted as bold*

Methods	(1)	(2)	(3)	(4)	(5)
Encoder	brnn	brnn	brnn	rnn	cnn
Decoder	rnn_dec	rnn_dec	rnn_dec	rnn_dec	cnn_dec
#Layers	1	1	2	1	1
Global attention	[24]	[9]	[9]	[9]	[9]
Perplexity	24.31	**20.34**	31.79	20.67	81.06
Accurate (%)	54.98	**56.54**	48.06	56.12	40.74

As in Table 8, the current deep learning-based methods have not produced high translation performance. In our observation, this drawback comes from the limitation of training data, although the introduced extensive pre-processing process and data embedding have been applied. For enhancing translation efficiency, we consider increasing the core of the Vietnamese-K'Ho bi-lingual dataset in both diversity and quantity.

4. Conclusion

In this paper, firstly, we built a Vietnamese - K'Ho bilingual corpus that includes more than 16,000 Vietnamese-K'Ho bilingual sentence pairs. Secondly, we use the OpenNMT framework for implementation with the collected bilingual data to build an automatic translation system, which achieves the highest accuracy with 56.54%. However, this is only the initial study for the

Vietnamese - K'Ho automatic translation system because the bilingual corpus is limited. In the future, we will continue to collect more data for the Vietnamese - K'Ho bilingual repository and research new methods to improve the accuracy of the translation system.

REFERENCES

[1] D. C. Le and T. T. T. Nguyen, "Vietnamese-English Translation with Transformer and Back Translation in VLSP 2020 Machine Translation Shared Task," in *Proceedings of the 7th International Workshop on Vietnamese Language and Speech Processing, Hanoi, Vietnam*, Association for Computational Linguistics, 2020, pp. 64–70.

[2] H. H. P. Vu, V. T. Tran, V. N. Nguyen, H. V. Dang, and P. T. Do, "Machine Translation between Vietnamese and English: an Empirical Study," *Journal of Computer Science and Cybernetics*, vol. 35, no. 2, pp. 147-166, 2019.

[3] N. Q. Phuoc, Y. Quan, and C.-Y. Ock, "Building a bidirectional english-vietnamese statistical machine translation system by using moses," *International Journal of Computer and Electrical Engineering*, vol. 8, no. 2, pp. 161-168, 2016.

[4] P. Huang, C. Wang, D. Zhou, and L. Deng, "Neural phrase-based machine translation," in *CoRR*, 2017.

[5] M. Cettolo, J. Niehues, S. Stuker, L. Bentivogli, R. Cattoni, and M. Federico, "The iwslt 2015 evaluation campaign," in *Proceeding of the 12th International Workshop on Spoken Language Translation*, 2015, pp 2-14,.

[6] H. P. Vu, V. Nguyen, V. Tran, and P. Do, "Towards state-of-the-art english-vietnamese neural machine translation," in *Proceedings of the Eighth International Symposium on Information and Communication Technology*, Nha Trang, 2017, pp 120-126,.

[7] H. Nguyen, L. Nguyen, P. Le, H. Nguyen, and T. Dinh, "Vietnamese-K'Ho automatic translation using statistical-based methods," *Science Journal of Da Lat University*, vol. 8, no. 3, pp.135-148, 2018.

[8] T. Nguyen and T. Dinh, "Vietnamese-K'Ho automatic translation using an example-based method," *Science Journal of Da Lat University*, vol. 6, pp.160 - 173, 2016.

[9] M-T. Luong, H. Pham, and C. D. Manning, "Effective Approaches to Attention-based Neural Machine Translation," *arXiv prePrint arXiv:1508.04025*, 2015.

[10] J. Gehring, M. Auli, D. Grangier, D. Yarats, Y. N. Dauphin, "Convolutional Sequence to Sequence Learning," *arXiv prePrint arXiv:1705.03122*, 2017.

[11] Y. Wu, M. Schuster, Z. Chen, Q. V. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey, et al., "Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation," *arXiv prePrint arXiv:1609.08144*, 2016.

[12] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, "Language Models are Unsupervised Multitask Learners," *Technical report*, OpenAi, 2019.

[13] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, "Improving Language Understanding by Generative Pre-Training," *Computer Science*, 2018. [Online]. Available: https://www.semanticscholar.org/paper/Improving-Language-Understanding-by-Generative-Radford-Narasimhan/cd18800a0fe0b668 a1cc19f2ec95b5003d0a5035. [Accessed October 05, 2022].

[14] Y. G. Nie, V. N. Hiep, and T. C. Lam, "Research and perfect the program to support handwriting processing of some ethnic minorities in the Central Highlands by TayNguyenKey software," Science and technology research topics, DakLak, 2010.

[15] F. J. Och and H. Ney, "Improved Statistical Alignment Models," in *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics*, Hong Kong, 2000, pp. 440-447,.

[16] R. Östling and J. Tiedemann, "Efficient Word Alignment with Markov Chain Monte Carlo," *The Prague Bulletin of Mathematical Linguistics*, vol. 106, pp. 125-146, 2016.

[17] S. Ker and J. Chang, "A Class-based Approach to Word Alignment," *Computational Linguistics*, vol. 23, pp. 313-343, 2002.

[18] Z.-Y. Dou and G. Neubig, "Word Alignment by Fine-tuning Embeddings on Parallel Corpora," *CoRR*, vol. abs/2101.08231, pp. 2112-2128, 2021.

[19] G. Klein, Y. Kim, Y. Deng, J. Senellart, and A.M. Rush, "OpenNMT: Open-Source Toolkit for Neural Machine Translation," *arXiv prePrint arXiv:1701.02810*, 2017.

[20] D. Amodei, S. Ananthanarayanan, R. Anubhai, J. Bai, E. Battenberg, et al. "Deep Speech 2: End-to-End Speech Recognition in English and Mandarin," *arXiv prePrint arXiv:1512.02595*, 2015.

[21] Y. Deng, A. Kanervisto, J. Ling, and A. M. Rush, "Image-to-Markup Generation with Coarse-to-Fine Attention," *arXiv prePrint arXiv:1609.04938*, 2016.

[22] W. Chan, N. Jaitly, Q. V. Le, and O. Vinyals, "Listen, Attend and Spell," *arXiv prePrint arXiv:1508.01211*, 2015.

[23] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention Is All You Need," *arXiv prePrint arXiv:1706.03762*, 2017.

[24] D. Bahdanau, K. Cho, and Y. Bengio, "Neural Machine Translation by Jointly Learning to Align and Translate," *arXiv prePrint arXiv:1409.0473*, 2014.