Multilingual Neural Machine Translation for Low Resourced Languages: Ometo-English

1st Mesay Gemeda Yigezu
dept. Information Technology
Wachemo University
Hosanna, Ethiopia
messay.gemeda@gmail.com

2rd Michael Melese Woldeyohannis
School of Information Science
Addis Ababa University
Addis Ababa, Ethiopian
michael.melese@aau.edu.et

3nd Atnafu Lambebo Tonja
dept. Information Technology
Wolaita Sodo Universty
Wolaita Sodo, Ethiopia
atnafulambebo1@gmail.com

***Abstract*—Unlike the technologically favored languages, under-resourced languages highly suffer from the lack of language resources for machine translation. In this paper, to overcome the problem of language resources, we have used four Ometo (Wolaita, Gamo, Gofa, and Dawuro) languages by automatically extracting parallel sentence pairs from religious domain available on the Internet. The corpora are used for conducting neural machine translation experiments are from Ometo to English. The BLEU score of the machine translation system shows a promising result despite the language differences, the morphological richness, and complexity of the Ometo languages which has an impact on the performance of the Ometo-English machine translation. We are now working towards developing a translation system that significantly reduces the effect of morphological richness and complexity of the Ometo languages using linguistic component.**

***Index Terms*—Ometo-English, Neural Machine Translation, Multilingual Translation , Low resourced languages, Ethiopian Languages**

I. Introduction

Language is a tool fashioned by a man, It is the only gift that identifies human beings from the rest of life [1]. Language is a means of communication in our day to day activities to do various things, like giving commands, asking questions and expressing feelings, but we use it specially to communicate information about world. Natural Language Processing is the computerized approach to analyzing text that is based on both a set of theories and a set of technologies [2]. One of the NLP application widely used to facilitate communication between human with different language is Machine Translation [3]. It is difficult to learn and speak all languages spoken in this world, for this reason not all peoples are communicated with each other. Usually, this communication gap is solved by using a human interpreter. However, the use of human interpreters is expensive and inconvenient. Many researches are being done to resolve this problem using machine translation techniques. Machine translation is an automatic translation of a source language to a target language. In today's Information age, a lot of written documents, brochures, text books, magazines, advertisements and other information in the web are being produced in technological supported and resourced languages such as English, European (French, Germany, Italy) and Asian languages (Indian, Chinese, Japanese) [4]. Several studies and applications have been done for foreign languages using different methodologies and approaches. Most of the machine translation works have been done on language pair of English and other languages, such as Arabic [5], Japanese [6], India [7], Malayalam [8], Bangla [9] among others. However, research in the area of MT for Ethiopian languages, which are under-resourced as well as economically and technologically disadvantaged, has started very recently [3].Some of research done for Ethiopian languages with English pair are English-Amharic language [10], [11], English-Afan Oromo machine translation [12], English to Wolaita [13] and English-Tigrigna [14]. On the other hand, Ometo language which have around four million speakers are facing many problems due to unavailability of language resources for different NLP application [15]. This hinders the communication between the people who speaks Ometo languages and people who speaks technologically favoured and resourceful language like English. To facilitate the communication between technologically favoured and resourceful languages like English and also to use the documents and information produced in local languages like Ometo,the documents need to be translated. Machine translation is one of NLP application that facilitates communication between languages like Ometo which are resource deficient language and technologically supported languages like English.

II. Motivation of this Paper

There is communication gap between local language speakers and speakers with technologically supported and resourceful language which is one of the problem especially between Ometo speakers and English-speaking community. However, many of the local language speakers especially Ometo(Wolaita, Gamo, Gofa and Dawuro) speakers have limited language resources due to this, it is difficult to communicate with technologically favoured and resourced languages like English if not translated. Ometo languages are very low resourced languages as there is no sufficient data, linguistic resource for different NLP application. Among these applications, machine translation is one that is used to facilitate communication from one language to another by translating documents, this helps local languages like Ometo(Wolaita, Gamo, Gofa and Dawuro) to facilitate communication with technologically favoured and resourced languages like English.

A. Related Works

Neural machine translation is new machine translation approach and current state-of-the-art that has been shown to be more effective in translation tasks compared to other machine translation approaches [16]. Many researchers carried out research in the area of neural machine translation and multilingual machine translation. Here are some of related works done by different researchers in the area of low resource and multilingual machine translation. In paper [17] conducted research on multilingual neural machine translation for low resource language for three languages (English, Italian, Romanian) covering six translation directions. In this work they showed how so-called multilingual NMT can help to tackle the challenges associated with low resource language translation. Finally, the study achieves competitive results also for language pairs not seen at training time using a pivoting (x-step) translation. The study showed that a single multilingual system achieves comparable performances with the bilingual baselines while avoiding the need to train several single language pair models. Then, they showed how a multilingual model can be used for zero-shot translation by using a pivot language for achieving slightly lower results than a bilingual model trained on that language pair. In paper [18] conducted research on Multilingual Neural Machine Translation for Indian Languages (Sindhi, Bhojpuri, Magahi).They proposed a data augmentation technique to improves the model. The technique helps achieve a jump of more than 15 points in BLEU score from the Multilingual NMT Model. A BLEU score of 36.2 was achieved for Sindhi–English translation, which is higher than any score on the leader board of the LoResMT SharedTask at MT Summit 2019, which provided the data for the experiments. In paper [19] conducted research on NMT between English and five African Low Resource Language pairs (Swahili, Amharic, Tigrigna, Oromo, Somali).For the five languages aligned to English, the researchers collected all available parallel data from JW300, Bible, Tanzil, and Ted talks. The study used BLEU score to measure systems' performance. All the models are trained using the OpenNMT implementation of Transformer. In this work, the study analyzed that the state of NMT approaches using five low-resource languages. The study shows that the baseline single-pair model can be significantly improved by the more robust semi-supervised, transfer-learning, and multilingual modeling approaches. In paper [20] applied neural machine translation to the task of English Bangla translation in both directions and compare it against a standard phrasebased statistical machine translation system.The study focused on two objectives: Firstly, to present the result on the English-Bangla translation using neural machine translation. Secondly, to present the result on the low- resource English-Bangla neural machine translation using sub word segmentation. They used Moses to build a standard phrase-based SMT, GIZA++ for Word alignment and Nematus to train NMT. Multilingual machine translation addresses the task of translating between multiple source and target languages [21] but many multilingual machine translation researches are done by using zero shoot and pivot translation methods which needs large data set [22], [18], but this is not work for low resourced language like Ometo languages. So to overcome the problem of low resource in multilingual machine translation we used a new approach for languages which have low resource and shares vocabulary count.

III. Ometo Languages

Ethiopia has more than 83 different languages with up to 200 different spoken dialects [23]. The Ethiopian languages are divided in to four major language family groups [15]. These are Semitic, Cushitic, Omotic and Nilo-Saharan. Omotic is one of the six language families within the Afro-asiatic phylum predominantly spoken between the lakes of southern rift valley and Omo River. Ometo represents a large group of languages within Omotic language branch in Afro-asiatic family which are classified together as a genetic unit because or their phonology, grammar and lexicon are quite close to each others [24]. Ometo languages are sub grouped into North, South, East and West Ometo [25]. The Northern Ometo group includes languages which have traditionally been known as the Wolaita dialect cluster, notably Wolaita, Gamo, Gofa, Dawuro and Dorze. From these Ometo groups; Wolaita has the largest speaker followed by Gamo, Gofa and Dawuro in order they appear.

A. Wolaita Language

Wolaita refers to people, language and area in southern part of Ethiopia which is located in Wolaita zone with around 2.48 million speakers of the language [26], [27]. The language is given as medium of instruction at primary school level and taught as a subject in secondary and high school. Currently, the language is offered as a subject in Bachelor Degree at Wolaita Sodo University. Meanwhile the language is serving as working language and means of communication in government offices in Wolaita Zone.

B. Gofa Language

Gofa refers to language, people and area in southern part of Ethiopia, which is spoken by the people of Gofa Zone as well as the different communities living in Gamo Zone and other border areas with 362,000 speakers [27], [28]. The language is given as medium of instruction at primary school level and taught as a subject in secondary, high school and serving as working language and means of communication in government offices in Gofa Zone.

C. Gamo Language

The name Gamo is widely used both as a name of the people and of the language cluster, a collective name to which all the Gamo dialects belong. In fact, the people call themselves Gamo and they refer to their language as Gamotstso, literally means "the Gamo language". The language Gamo is spoken by around 1.63 million [27]. Gamo is an Ometo language of the Omotic language family used as a language of instruction in the lowest grades in primary school and medium of communication in Gamo Zone and in border areas [29].

D. Dawuro Language

Dawro is one of the Ometo group language, which belongs to the Omotic language family spoken primarily in the Dawro zone of the Southern Nations, Nationalities, and Peoples' Region (SNNPR) in the Southwest of Ethiopia. Dawro people also refer to their language (locally) as Daurotsua or Dauro K'ala with an approximate speakers of 538,000 [27], [30]. Dawro is used in education in the Dawro Zone, and students receive native language instruction through all grades and now it is also possible to study Dawro in higher education to obtain a diploma [31]. The language is also used working language and means of communication in government offices in Dawuro Zone.

E. Writing System of Ometo Language

Writing System of Ometo Language The writing system of this language pair uses Latin alphabet.Ometo languages are a suffixing language in which words can be generated from root words recursively by adding suffixes only. Ometo nouns are inflected for number, gender and case whereas verbs are inflected for person, number, gender, aspect and mood [32]. Ometo languages follow Subject-Object-Verb (SOV) word order. Ometo languages(Wolaita, Gamo, Gofa and Dawuro) share the greatest majority of their consonant inventories [33]. Consonants in the four dialects can be categorized into six categories: stops, fricatives, affricates, nasals, approximants and semi-vowels. Gamo has twenty-six consonant phonemes, Wolaitta has twenty-four consonant phonemes while Dawuro and Gofa have twenty-five phonemes each. Gamo has one peculiar consonant,/dz/, which is absent from the others. Phonemic inventories of the four dialects also show variation with respect to the consonants /t'/ and /s'/. Wolaita has /t'/ while the other three have /s'/ instead. Considering the phonemic inventories of the four dialects, Wolaitta differs much from the other three because, firstly, it lacks the alveolar affricate consonant ts and secondly it has its own peculiar phoneme, /t'/ which is absent elsewhere [34], [35]. However, the / t'/ in Wolaitta regularly corresponds to /s'/. Gamo, Dawuro and Gofa have much phonological feature in common to each other than Wolaitta. Among the three, Gamo seems a bit divergent from the other two (Dawuro and Gofa). Gamo has /dz/ which is absent from the phonemic inventories of others [34]. Like other Latin languages Ometo language use five common vowels. This is similar for Wolaita, Gamo, Gofa and Dawuro. The Figure1 show vocabulary count with in Ometo languages.

In Figure 1 shows, highest rate of identical vocabulary correspondences is exhibited between the Wolaitta and Dawuro varieties. According to the research [34] the researcher concluded that majority of the vocabulary items in the four dialects are share cognates and only few vocabulary items are being peculiar to each speech variety. Based on the Figure1 Gamo shares more vocabulary items with Gofa and Wolaitta than with Dawuro.

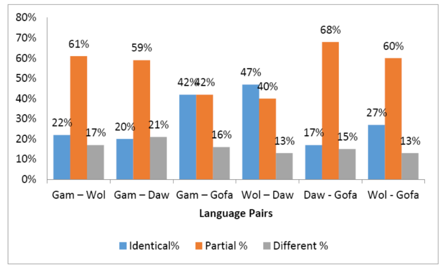


Fig. 1: Identical, Partial and Different Vocabulary correspondences in Ometo [34]

.

IV. Challenges of NMT

Neural machine translation is affected by amount of parallel data, morphological complexity and domain. Neural machine translation needs huge parallel dataset to train a model, this might be a challenge in neural machine translation for Ometo-English because there is no enough resource for Ometo languages. Ometo(Wolaita, Gamo, Gofa and Dawuro) languages are used in this study are morphological rich languages, this might challenge neural machine translation system because for Ometo languages their may be different meaning in English. Due to unavailability of resource for Ometo languages this study used only Bible domain. Using one domain in translation might challenge the translation system.

V. Corpus preparation

Neural machine translation is a new breed of corpus-based machine translation which is also called data-driven or, less often, corpus-driven machine translation [36]. Data-driven machine translation is trained on large corpora of pairs of source language segments usually sentences and their translations, that is, basically from huge translation memories containing hundreds of thousands or even millions of translation units. To develop machine translation model, it needs parallel corpus which has source and target dataset in order to translate one (source) language to another (target) language [13]. Compared to technologically favored language like English, European (like French and Spanish) and Asian languages (Chinese and Japanese), resource for the Ometo languages are difficult to access as most of the data used in these languages are available printed format. Due to this problem, a parallel corpus for Gamo, Gofa and Dawro languages paired with English collected from Ebible[1]. The parallel corpora is collected from Holy Bible that support many languages including Ometo. While for Wolaita language, the researcher used dataset from [3] found in GitHub[2]. To extract this religious resource from websites, a web crawler is used for each article after identifying the structure of web documents (html) including the

[1] http://ebible.org

[2] https://github.com/AAUThematic4LT/Parallel-Corpora-for-Ethiopian-Languages

page, book and phrases. Accordingly, Python libraries such as requests, regular expression (RE) and BeautifulSoup (BS) were used to analyze the structure of the websites and extract the content of the article for a given unified resource locator (URL). Table I presents the details corpus distribution of Ometo-English in terms of sentence, token and types.

TABLE I: Sentence distribution of Wolaita, Gamo, Gofa and Dawuro languages

Language	Sentence	Token	Type	Average Sentence length
English	26,943	703,122	12,131	26
Wolaita		469,851	42,049	17
English	7,866	177,410	11,078	23
Gamo		125,509	23,589	16
English	7,928	175,727	8,769	22
Gofa		119,289	25,301	15
English	7,804	207,954	4,368	27
Dawuro		126,734	17,392	16

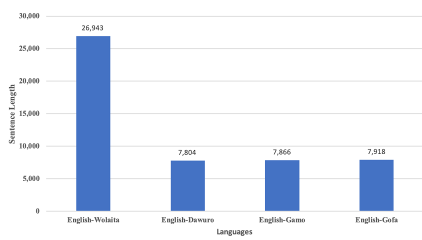


Fig. 2: Sentence distribution of Ometo-English language

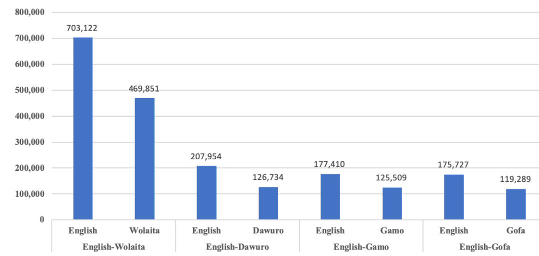
.

In Figure 2 Gofa, Dawuro, and Gamo has less amount of parallel dataset compared to Wolaita. This is because data collected for Gofa, Dawuro, and Gamo languages are only from the bible version of the "New Testament" while the Wolaita dataset contains both "Old Testaments" and "New Testament" version of Bible and the size of used dataset for Wolaita is around three times larger than that of Gamo, Dawuro and Gofa. Similarly, Figure 3 shows sentence, token and type distribution of Ometo English pairs in collected corpus

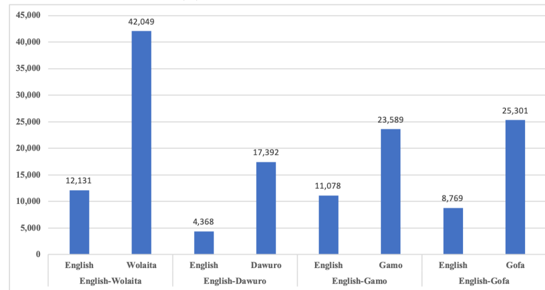
VI. Experimental Setup

To conduct the NMT experimental, OpenNMT [37] is used to train, develop and evaluate the machine translation model. The experiment is conducted using open source neural machine translation toolkit called OpenNMT which is freely available in GitHub with implementation of LSTM type of RNN architecture with global attention using encoder-decoder language model. Along with that we used Python programming language to pre-process our parallel data before feeding it into colab notebook. We used online Google collaborator environment with web browser-based notebook which is used to write programs in python programming language and it also support GPU run-time to train and test our mode faster than using CPU of computer. Bilingual Evaluation Under Study (BLEU) is used for automatic scoring. In this study, a total of three different experiments conducted without any attempt to split the dataset. In each experiment, a parallel data consist of Wolaita English combined with two other languages from Gamo, Gofa and Dawuro for the training while the third language is used for evaluating the translation model. Bilingual Evaluation Under Study (BLEU) is used for automatic scoring. Table II presents the training, development and testing data used in the experiment of Ometo-English machine translation.

As depicted in Table II, In the first experiment data, we used a total 42,614 parallel sentences for training which consists of Wolaita, Dawuro and Gamo language paired with English. Whereas for development and testing a total of 1,584 and 6,336 parallel sentences used from Gamo language, respectively. While in the second experiment data, we used a total of 42,727 parallel sentences from Wolaita, Gofa and Gamo language for purpose of training while 1,561 sentences for development and



(a) Token distribution



(b) Type distribution

Fig. 3: Sentence,Token and Type distribution of Ometo-English Language pairs

TABLE II: Train, Validation and Test set split for the experiments

	Language Pair	Train	Dev	Test
Experiment 1	Wolaita-English	26,943	-	-
	Dawuro-English	7,804	-	-
	Gamo-English	7,866	-	-
	Gofa-English	-	1,584	6,336
Experiment	Wolaita-English	26,943	-	-
	Dawuro-English	-	1,561	6,244
	Gamo-English	7,866	-	-
	Gofa-English	7,918	-	-
Experiment 3	Wolaita-English	26,943	-	-
	Dawuro-English	7,804	-	-
	Gamo-English	-	1,574	6,266
	Gofa-English	7,918	-	-

6,244 parallel sentences for testing from Dawuro language. In all the experiment, the Wolaita-English sentence are used as a training beside the two languages from Dawuro, Gamo and Gofa languages.

VII. Experiment Results and Discussion

To train and test NMT model in neural network we have used default OpenNMT parameters [37] which are described in separate YAML files in OpenNMT. They define data files, optimization settings, dynamic model parameters, and options related to training and inference.Some of paramters used in study are Optimizer-Adam, Learning rate-1.0, Dropout -0.3, Rnn type-LSTM, Embedding size-500, Batch size-64, Training steps-20000 steps, Evaluation batch size-32, Save checkpoints-500 steps, Encoder and decoder num-layers-2. Figure 4 shows

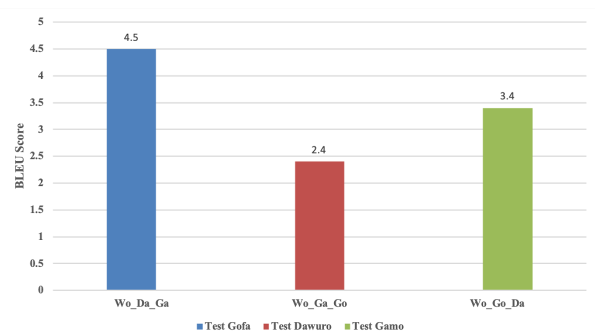


Fig. 4: Experiment results of Ometo – English translation

results of three experiments,in Experiment I, using Wolaita, Dawuro and Gamo parallel sentence combination for training and Gofa parallel sentence for testing and validation gives BLEU score of 4.5, which is greater than two other combinations. This is because Gofa language shares more vocabulary with Wolaita and Gamo so the model gives good result on predicting Gofa words. In Experiment II, using the combination of Wolaita, Gamo and Gofa parallel data set for training and Dawuro dataset for testing and Validation showed BLEU score of 2.4 which is the lowest from the three experiments, this is because Dawuro language shares less vocabulary with Gamo and Gofa so this affects the probability of predicting Dawuro sentence. In Experiment III,using the combination of Wolaita, Gofa and Dawuro parallel data set for training and Gamo dataset for testing and Validation showed BLEU score of 3.4 which is the greater than the score of Experiment II and less than the score of Experiment I, this is because Gamo shares more vocabulary items with Gofa and Wolaitta than with Dawuro as result of this the Experiment III shows higher score than Experiment II. For experiment I, II and III we used three languages for training and one language for testing by changing language in training and testing. We used Wolaita data in all three experiments for training because Wolaita has large parallel dataset compered with the three(Gamo, Gofa and Dawuro) languages. In addition, Wolaita language shares large vocabulary items with the three languages compared to other languages. So, we trained all three experiment using Wolaita dataset combining with three Ometo(Gamo, Gofa and Dawuro) languages by interchanging Gamo, Gofa and Dawuro for testing and validation.

VIII. Conclusion and Feature Work

In this paper, we have introduced new approach in order to solve low resource problem in machine translation task. The paper includes collecting parallel corpus for four language pairs and three experiments were conducted by using collected dataset to investigate the performance of the model. The first experiment carried out by combining Wolaita, Dawuro and Gamo for training and using Gofa for testing shows BLEU score of 4.5,The second experiment carried out by combining Wolaita, Gofa and Gamo for training and using Dawuro for testing showns BLEU score of 2.4. The third experiment carried out by combining Wolaita, Dawuro and Gofa for training and using Gamo for testing and shows BLEU score of 3.4, From test result of conducted experiments in this research we obtained promising result for low resourced Ometo language pairs. To increase the performance of the model, using different domain, using linguistic resource like dictionaries, using Byte pair encoding (BPE) level of corpus segmentation and increasing the size of data set could be explored.

References

[1] M. Patchell, H. F. Pommer, and W. M. Sale.: The Use of Language, vol. 8, no. 1. 1948.

[2] E. D. Liddy.: Natural Language Processing. In Encyclopedia of Library and Information Science, Marcel Decker, Inc., pp. 1–15, 2001.

[3] S. T. Abate et al.: Parallel Corpora for bi-Directional Statistical Machine Translation for Seven Ethiopian Language Pairs, October, vol. 0, no. 1, pp. 153–156, 2018

[4] S. T. Abate et al.:"English-Ethiopian Languages Statistical Machine Translation," Proc. 2019 Work. Widening NLP, pp. 27–30, 2019.

[5] K. SHAALAN, A. RAFEA, A. A. MONEIM, and H. BARAKA, :"Machine Translation of English Noun Phrases into Arabic," Int. J. Comput. Process. Lang., vol. 17, no. 02, pp.121–134, 2004, https://doi.org/:10.1142/s021942790400105x.

[6] M. Murata and M. Nagao, :"Determination of referential property and number of nouns in Japanese sentences for machine translation into English," pp. 1–8, 1994, [Online]. Available: http://arxiv.org/abs/cmp-lg/9405019.

[7] S. Saini and V. Sahula,:"Neural Machine Translation for English to Hindi," Proc. - 2018 4th Int. Conf. Inf. Retr. Knowl. Manag. Diving into Data Sci. CAMP 2018, pp. 25–30, 2018, doi: 10.1109/INFRKM.2018.8464781.

[8] R. Rajan, R. Sivan, R. Ravindran, and K. P. Soman: "Rule based machine translation from english to Malayalam," ACT 2009 - Int. Conf. Adv. Comput. Control Telecommun. Technol., pp. 439–441, 2009, https://doi.org/:10.1109/ACT.2009.113.

[9] M. A. Al Mumin, M. H. Seddiqui, M. Z. Iqbal, and M. J. Islam: "Neural Machine Translation for Low-resource English-Bangla," J. Comput. Sci., vol. 15, no. 11, pp. 1627–1637, 2019, https://doi.org/:10.3844/jcssp.2019.1627.1637.

[10] L. B. Mulu Gebreegziabher Teshome:"PRELIMINARY EXPERIMENTS ON ENGLISH-AMHARIC STATISTICAL MACHINE TRANSLATION," 2011.

[11] M. G. Teshome, L. Besacier, G. Taye, and D. Teferi: "Phoneme-based English Amharic Statistical Machine Translation," IEEE AFRICON Conf., vol. 2015Novem, pp. 6–10, 2015, https://doi.org/:10.1109/AFRCON.2015.7331921.

[12] S. Adugna and A. Eisele: "English - Oromo Machine translation: An experiment using a statistical approach," in Proceedings of the 7th International Conference on Language Resources and Evaluation, LREC 2010, 2010.

[13] M. Mara: "ENGLISH-WOLAYTTA MACHINE TRANSLATION," St. Mary's University, 2018.
[14] A. Gebremariam: "Amharic-to-Tigrigna Machine Translation Using Hybrid Approach," 2017.
[15] A. Amha: The Omotic Language Family, no. February. 2017.
[16] X. Wang, Z. Lu, Z. Tu, H. Li, D. Xiong, and M. Zhang: Neural machine translation advised by statistical machine translation, 31st AAAI Conf. Artif. Intell. AAAI 2017, pp. 3330–3336, 2017.
[17] S. M. Lakew, M. A. Di Gangi, and M. Federico: Multilingual neural machine translation for low resource languages, CEUR Workshop Proc., vol. 2006, 2017, https://doi.org/: 10.4000/books.aaccademia.2419.
[18] Madaan, P. and Sadat, F.: Multilingual neural machine translation involving Indian languages, In Proceedings of the WILDRE5–5th Workshop on Indian Language Data: Resources and Evaluation (pp. 29-32).
[19] S. M. Lakew, M. Negri, and M. Turchi: Low Resource Neural Machine Translation: A Benchmark for Five African Languages, ICLR 2020, pp. 1–10, 2020, [Online]. Available: http://arxiv.org/abs/2003.14402.
[20] M. A. Al Mumin, M. H. Seddiqui, M. Z. Iqbal, and M. J. Islam: Neural Machine Translation for Low-resource English-Bangla:J. Comput. Sci., vol. 15, no. 11, pp. 1627–1637, 2019, https://doi.org/: 10.3844/jcssp.2019.1627.1637.
[21] Blackwood, G., Ballesteros, M., Ward, T. (2018): Multilingual neural machine translation with task-specific attention.arXiv:1806.03280.
[22] Johnson, M., Schuster, M., Le, Q. V., Krikun, M., Wu, Y., Chen, Z., ... Dean, J. (2017): Google's multilingual neural machine translation system: Enabling zeroshot translation. Transactions of the Association for Computational Linguistics, 5, 339-351.
[23] H. Woldemariam: The challenges of mother-tongue education in Ethiopia: The case of North Omo area, Lang. Matters, vol. 38, no. 2, pp. 210–235, 2007, https://doi.org/: 10.1080/10228190701794608.
[24] A. Azeb:Aspects of the verb in Ometo: Ethiopia. J. Lang. Lit., vol. 6, pp. 43–64, 1996.
[25] Henok Wondimu:The Grammaticalization of Copula Markers in the Ometo Subgroup, Addis Ababa University, 2010.
[26] M. Wakasa,:A Sketch Grammar of Wolaytta, 2014
[27] Eberhard, David M., Gary F. Simons, and Charles D. Fennig. (2021). Ethnologue: Languages of the World. Twenty-fourth edition. Dallas, Texas: SIL International. Online version: http://www.ethnologue.com.
[28] S. C. Hirboro:Documentation and grammatical description of Gofa, Addis Ababa Univeristy, 2015
[29] A. L. Thomassen: Contributions to the description of the phonology of the Bonke variety of Gamo, University of Oslo, 2015
[30] T. NEGESE: Aspect of Dawro Phonology, Addis Ababa University, 2010
[31] S. Hanserud: Dawro verb morphology and syntax, Universtitas Osloensis, 2018.
[32] Motomichi Wakasa,:A descriptive study of the modern Wolaytta language. Unpublished PhD thesis, University of Tokyo,2008.
[33] W. Hirut, "Writing both difference and similarity: towards a more unifying and adequate 78 — P a g e orthography for the newly written languages of Ethiopia: the case of Wolaitta, Gamo, Gofa and Dawuro," J. Lang. Cult., vol. 5, no. 3, pp. 44–53, 2014, https://doi.org/: 10.5897/jlc2013.0235.
[34] W. Hirut, "Notes on the North Ometo dialects:mutual intelligbility tests and structural variations."
[35] M. L. Bender, Comparative morphology of the Omotic languages, vol. 19. 2000.
[36] Forcada, M. L. (2017). Making sense of neural machine translation. Translation spaces, 6(2), 291-309.
[37] Klein, G., Kim, Y., Deng, Y., Senellart, J.,Rush, A. M. (2017) : Opennmt: Opensource toolkit for neural machine translation:arXiv preprint arXiv:1701.02810.