INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

ORIGINAL ARTICLE

OPEN ACCESS

Received: 25/03/2025

Accepted: 12/05/2025

Published: 22/08/2025

Integrating Syntactic Structures for Enhanced English-Amharic Machine Translation: A Graph2Seq-Based Approach

Muluken Hussen Asebel[1]*, Shimelis Getu Assefa[2], Mesfin Abebe Haile[1], Ramasamy Srinivasagan[3]

1 Department of Computer Science and Engineering, School of Electrical Engineering and Computing, Adama Science and Technology University, Adama, Ethiopia
2 Department of Research Methods and Information Science, Morgridge College of Education, University of Denver, Denver, CO, United States
3 King Faisal University, Computer Engineering, CCSIT, Al Hufuf, 31982, Saudi Arabia

Citation: Asebel MH, Assefa SG, Haile MA, Srinivasagan R (2025) Integrating Syntactic Structures for Enhanced English-Amharic Machine Translation: A Graph2Seq-Based Approach. Indian Journal of Science and Technology 18(32): 2626-2638. https://doi.org/10.17485/IJST/v18i32.558

* **Corresponding author**.

muluken2@gmail.com

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Abstract

Background/Objective: The shift from classical rule-based and statistical systems of Machine Translation (MT) to more advanced Neural Machine Translation (NMT) systems employing transformers is quite astonishing. Still, translating language pairs such as English and Amharic, which are structurally different, poses a considerable challenge. This study explores the impact of incorporating known syntactic information on the performance of English-Amharic translation in Graph2Seq model to the precision and fluency of translations. **Methods:** We aimed to enhance translation performance with the design of an attention-based Graph2Seq model that utilizes a shaped syntactic dependency tree and implements Graph Neural Networks (GNN) to capture the hierarchy of language. The model overcomes the constraints posed by conventional sequence-to-sequence (Seq2Seq) models by adding syntactic information into the encoder. The dataset includes roughly 1.14 million parallel sentences in English and Amharic. **Findings:** There is a positive, actionable gap, as the proposed model was able to exceed the benchmarks set by the transformer models and pretrained methodologies, attaining BLEU 37.3, and demonstrating marked improvement in translation quality. Furthermore, these findings have defied the argument stating that diversity on the morpho-syntactic level impedes translation, instead illustrating the potency of rich superset syntax modeling. This score beats baseline figures by 4.56 points and 24.24 points compared to M2M100 and the standard transformer models, respectively. **Novelty:** All these points towards the fact that incorporating deep syntactic structures to enrich Graph-NMT systems on low-resource languages enhances their capabilities, paving avenues for further investigations in the field of Machine Translation. This study can be a starting point towards more research on the enhancement of accuracy and fluency of translations using a multidisciplinary approach based on language.

Keywords: Graph neural networks; English-Amharic machine translation; Syntactic of source language; BLEU score; Low-resource language

1 Introduction

Machine Translation (MT) has undergone significant advancements, evolving from rule-based (RBMT) and statistical (SMT) approaches to neural machine translation (NMT)[1]. Transformer-based models have further improved NMT performance, demonstrating superior translation quality across many language pairs[2]. Despite advancements in Neural Machine Translation (NMT) and English-Amharic translation, the fundamental differences in the languages present a challenge. Currently, NMT systems use a sequentially organized encoder-decoder model which does not make any use of the languages' grammatical structures[3],[4],[5]. Incorporating language-related information, especially grammatical structures, into NMT has the capability of enhancing translation accuracy and coherence to a great degree[6],[7]. The purpose of this research is to incorporate grammatical structures or syntactical structure in performing a syntactic English to Amharic translation.

English and Amharic belong to different language families; while English is a Germanic language, Amharic is a Semitic language[8]. They also greatly differ in the order of their constituents or grammatical structures. English follows Subject-verb-object (SVO) order while Amharic is Subject-Object-Verb (SOV) based[9]. These differences pose difficulties to NMT models that have been trained exclusively using the data-driven approach because word order and grammatical agreement are particularly problematic to resolve. Prior research has suggested that translation quality improves with the addition of syntactic information. Eriguchi, Hashimoto[10] proved that syntax-aware neural models are more precise because they exploit hierarchical structures of sentences, and syntactic processes such as dependency parsing[11] make the resulting text more fluent and coherent. In light of these findings, this study aims to develop a syntax-aware graph-NMT model that utilizes such features in translation automation.

A challenge for Amharic neural machine translation (NMT) systems is the rich morphology and complex linguistics of the language. Given that a single word in Amharic contains information regarding the subject, object, gender, tense, mood, and aspect, this results in highly inflected words. Other ailments due to the abundance of vocabulary are also problematic for NMT systems. Moreover, Amharic verbs' non-concatenative morphology makes it difficult to generalize standard word-based models. Character-level word models employing morphological analysis have shown promise. These models better capture subword information, as shown with transformer-based models by Asefa and Assabie[12], which integrates character-level embeddings alongside advanced regularization techniques. Computational efficiency is also achievable with subword tokenization; Gezmu, Nürnberger[13] demonstrate this by mitigating the inflectional complexity of the Amharic language. Supporting these models, NMT systems still struggle to construct accurate representations of the language's deeply embedded morphology and syntax. Further developing model sensitivity to deeper morphological relativity aligned with syntax will enhance translation accuracy.

Still unresolved difficulties which the transformer-based approach has to face with are grammar errors, inconsistency in expected and produced output sequences, and non-contextualized meaning for English to Amharic translations. Existing NMT systems function based on data-driven learning techniques that do not incorporate features of syntax, which results in poorly formed translations, particularly in low-resource languages like Amharic[14]. To fill this gap, this paper seeks to incorporate techniques like dependency parsing tree, and syntactic knowledge to enhance accuracy and fluency in translations.

The attention mechanisms in the NMT systems were developed by Bastings, Titov[4] and Bahdanau, Chorowski[3]. The models used latent feature vectors as encoders for the source sentences. While these feature representations help with translation generation, they lack explicit syntactic structural information. This study attempts to improve the translations by adding dependency syntax trees to the encoder using Graph Neural Networks (GNNs). Figure 1 illustrates how a dependency syntax tree depicts the syntactic relationships between words, phrases, and clauses. In the clause "I booked a ticket to Adama," the subject "I" fulfills the action "booked" and has an object "ticket". These syntactic structures assist machine translation models to capture dependencies, therefore improving accuracy and efficiency in the creation of better translation models. Understanding these dependencies helps the models align elements of the source and target languages more accurately, particularly when dealing with languages that have fully different orders and grammatical systems. Translation models can incorporate issue of preserving meaning, managing long-range dependencies, resolving ambiguities, and increasingly fluent translations through the context by incorporating more syntax information into their computations.

To build as well as capture NMT representations that are informed by syntax, this research uses GNNs. GNNs improve the comprehension of word interdependencies and relations by synthesizing contextual feature representations with syntactic graph structures, making the understanding of word dependencies and relationships easier to capture[4]. More accurate machine translation is achieved through GNNs by capturing intricate syntactic structures.

The study seeks to analyze the impact of embedding syntactic structure into English to Amharic graph NMT, implement a hybrid Neural Machine Translator (NMT) that uses syntax aware models like dependency trees, and evaluate performance using automatic metrics such as the BLEU score. In addition, the goal of the study is to perform an evaluation of a syntax aware metamorphosed model against baseline transformer models. To achieve these goals, preliminary research had to be done in

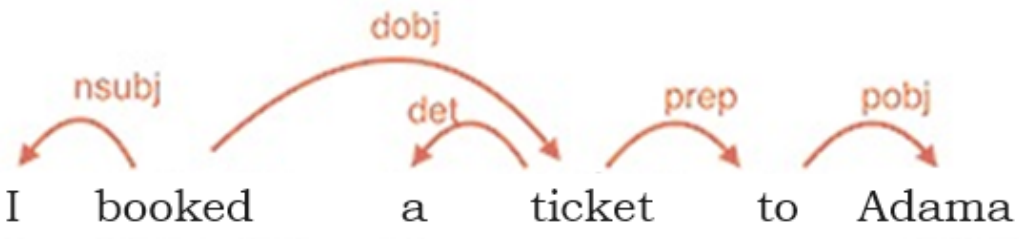

Fig 1. Syntax dependency tree for the example sentence: "I booked a ticket to Adama"

order to determine how an evaluation of multi-modality deep learning performed using graphs could be done, thus leading to the eventual formation of a model.

With the intention of developing a prototype, a hybrid syntax-aware model employing graph attention was designed [4]. First, a baseline transformer NMT model [15] is trained in order to set a performance baseline. Later stages in training and optimization are: hyperparameter tuning, and attention to the syntactic alignment of constituents. Models in evaluation are compared based on their BLEU scores [16].

By incorporating syntactic information within English-to-Amharic graph NMT, the objectives of this research are to improve proper alignment, and grammatical word order issues. The contribution of the study is towards the improvement of machine translation for under-represented languages by showing the impacts that linguistically informed graph NMT models have on the quality of translation and contextual understanding.

2 Methodology

2.1 Parallel Dataset Preparation

For training and evaluating machine translation between English and Amharic, [17] constructed a parallel corpus. As compared to its predecessors, this corpus is considerably larger and freely available for research purposes. This corpus was used to train neural machine translation models.

The corpus is comparatively bigger than its predecessors and is published without restrictions for research use. Neural machine translation models were trained using this corpus. Notably, the translation models performed most effectively with subword units, achieving the highest BLEU scores among the tested models [13].

Table 1 illustrates that Biadgligne and Smaili [18] compiled the most extensive parallel corpus for English to Amharic language pairs. In addition to the datasets outlined in Table 1, Belay, Tonja [17] also played a role in MT research by developing a novel parallel corpus. This corpus consists of 33,955 sentence pairs sourced from various news platforms, including the Ethiopian Press Agency, Fana Broadcasting Corporate, and Walta Information Center. As the data are drawn from diverse sources, they encompass a wide range of domains, such as religious texts, politics, economics, sports, and news.

The dataset consists of approximately 1.1 million parallel sentences, with approximately 888,837 being distinct. This variation occurs because of duplicate sentences in the source materials. Notably, this collection of unique parallel sentences is the most extensive compilation achieved to date [17]. We have extensively used this dataset for our experimental endeavors.

Additional normalization procedures were implemented to improve the consistency and quality of the corpus. These processes included character encoding with Unicode normalization (NFC) and removal of non-printable characters. Also includes duplication removal of sentence pairs for neutrality during the training process. Normalization at the token level was also completed, which included lowering case if needed, modifying punctuation, such as changing curly quotes to straight quotes, and deletion of redundant whitespace. Amharic (Morphologically rich language) underwent stemming or lemmatization language-specific preprocessing, alongside script unification. Moreover, tools for automatic language identification were employed to eliminate disconnected or incorrectly labeled sentence pairs. All of these measures enhanced dataset after the initial alignment, language logic, dataset linguistic coherency, calibration precision along with dataset agility and model training adaptability.

Table 1. Available Amharic and English parallel data

Data source	# Sentence pairs	Accessible
Am-En ELRA-W0074	13,347	yes
Biadgligne and Smaili (19)	225,304	yes
Horn MT	2,030	yes
Am-En MT corpus	53,312	yes
Gezmu, Nürnberger (13)	145,364	yes
Abate, Melese (20)	40,726	yes
Lison, Tiedemann (21)	562,141	yes
Tracey and Strassel (22)	60,884	no
Admasethiopia	153	yes
MT Evaluation Dataset	2,914	yes
Belay, Tonja (18)	33,955	yes
Total	**1,140,130**	**yes**
Unique sentence pairs	**888,837**	**yes**

2.2 Graph Neural Networks

GNNs are a more contemporary artificial intelligence construct built around the idea of neural networks performing machine learning tasks on data structures known as graphs[19]. Unlike traditional neural networks that work in vector space, GNNs grasp the complex interrelations and interdependencies among elements that are portrayed as nodes and edges on a graph.

GNNs generate node embeddings that are full of information which can be processed and encoded in the form of walls, relations of words, or anything that makes up the graph's syntactic features. For instance, in machine translation, these embeddings are capable of encoding the structure of the sentence and its dependencies of relations among its words. GNNs offer for each node a contextual embedding with consideration of the whole graph structure which helps capture long-range dependencies and complex relationships. Such contextual embeddings are important for contextually accurate sequence generation during the decoding phase[20].

The node embeddings produced by GNNs contain informative parts which can be processed and encoded in the form of walls, word relations, or any feature which constitutes the syntax of the graph. In the case of machine translation, these embeddings could structurally encode the sentence and several types of relationships among words in that sentence. Capturing long range dependencies and complex interactions is aided by GNN's ability to create contextual embeddings for each node, which is done by analyzing the entire network topology. These contextual embeddings are what makes it possible to accurately contextually relevant sequences during decoding[21].

It is possible to learn from graphs in a ways knowledge graphs, by using graph attention networks (GATs). An attention mechanism is used by GATs to assign different weights to nodes and edges of a graph based on their relevance and importance[22]. Our proposed machine translation model use GAT, which to improve translation quality and capture syntactical information source language. Considering syntax dependency tree, we have some set of Node Features in the input layer $layer\ input\ h = h_1,\ h_2,\ \ldots, h_n \quad h_i \epsilon R^d$, where, h : input to the layer , n : number of nodes, d : number of features for each node. $layer\ output\ h' = {h'}_1,\ {h'}_2,\ \ldots, {h'}_N \quad {h'}_i \epsilon R^{d'}$ where, h' : output of the layer, N : Number of Nodes, d' : new representation of feature matrix for each node. For efficient learning, we use *multi-head attention* in all the intermediate layers for each node.

$$h'_j = \int_{k=1}^{k} \sigma \left(\sum_{v_{j \in N(v_i)}} \alpha_{ij}^k w^k \, h_j \right)$$

2.3 Syntactic encoder

Our method integrates GNNs with the translation system to incorporate source language syntax into the system. These GNNs are used in predicting the syntactic dependency trees of the source sentences. These syntactic neighborhoods of words need to impose the sensitivity on the word representation[21]. It suggests that when a sentence is represented as a subordinating dependency tree, the word relationships within the sentence are taken into account as defined by the tree. In this work, the

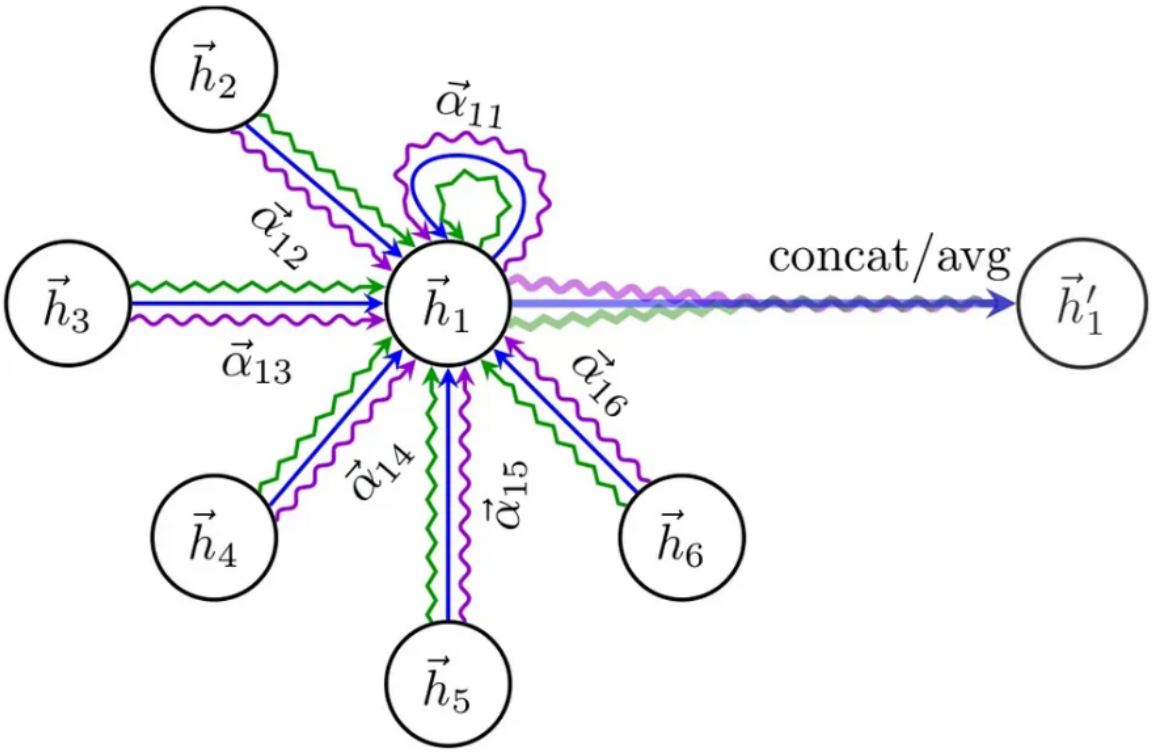

Fig 2. An illustration of multihead attention (with K = 3 heads) by node 1 on its neighborhood. Different arrow styles and colors denote independent attention computations. The aggregated features from each head are concatenated or averaged to obtain $\overrightarrow{h'_1}$ (Source Veličković, Cucurull[23])

GNNs are said to operate on word representations as inputs and outputs. They may be included as new layers on top of the standard encoders such as bidirectional RNNs or convolutional neural networks[4]. This type of encoder has access to rich syntactic information which makes the syntax sensitive part helpful in the machine translation problem and in ignoring the non-syntax sensitive parts without strict control over so many interacting structures. Such processes are flexible in terms of what can be altered in relation to the specific language structures[24].

To summarize, exploiting GNNs for including syntactic information into the encoder profoundly improves the translation processes by utilizing structural relations of words in the source language sentences. This precise strategy is expected to improve the translation output by having a better comprehension of the linguistic context.

3 The Proposed Graph2Seq Machine Translation Model

The structure of the sequence-to-sequence model (encoder-decoder), which is widely used, is depicted in Figure 2. The goal of this study is to add prior knowledge on the encoder side. Seq2Seq models have many flexible and expressive renderings, but they suffer from a serious limitation, which is their use is only constrained to tasks where the input data is represented exclusively as sequences. However, sequences are only the simplest form that structured data can take. Numerous other fundamental problems require far more complex structures. As an example, capturing complex pairwise relationships, which is a crucial aspect of many difficult problems, can be done using graphs[19]. Hence, a GNN is employed in the developed approach, allowing for the incorporation of prior knowledge, such as the syntax of the source language, to be utilized.

In our study, the encoder we employ utilizes a syntactic dependency tree of the source language, which undergoes processing via graph neural networks (GNNs). The construction of our encoder involves a series of steps aimed at maximizing the utilization of this graph-based representation. First, we represent the input sentence as a graph structure. We then incorporate two layers to facilitate the learning of node representations, leveraging this graph representation. These node representations

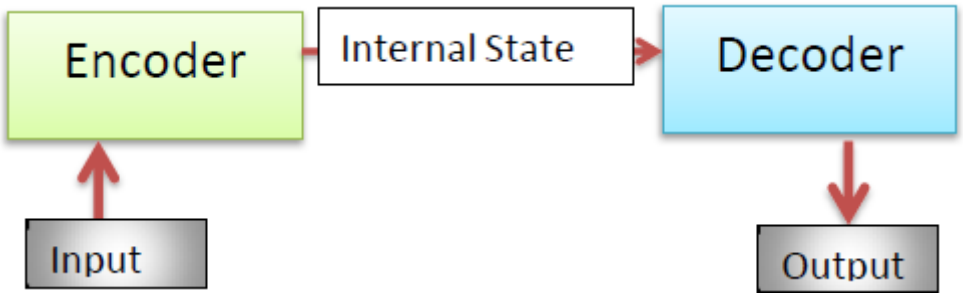

Fig 3. Encoder-decoder model

serve as the basis for generating the attention-based context vector, which is then passed to the decoder. Notably, our architecture employs the standard transformer decoder. By focusing exclusively on the states of textual nodes, our approach empowers the decoder to dynamically leverage contextual information, thereby enhancing its translation capabilities.

Figure 2 shows the main workflow of our study. The architecture consists of two main components: the encoder and the decoder. The encoder is responsible for processing the source language, which in this case is English, through multiple layers before passing it to the decoder. Conversely, the decoder receives input from the encoder and generates the target language, which in our context is Amharic. The preprocessed graph-based data undergo processing in an embedding layer before being fed into the stacked fusion layers. The first layer in the stack of fusion layers is the multihead self-attention layer. In this layer, self-attention mechanisms are used to generate contextual representations for each node, combining messages from neighboring nodes.

Formally, the contextual representations $C_x^{(l)}$ of all textual nodes are calculated as follows:

$$C_x^{(l)} = MultiHead\left(H_x^{l-1}, H_x^{l-1}, H_x^{l-1} \right), \tag{1}$$

where MultiHead(Q, K, V) is a multihead self-attention function that takes a query matrix Q, a key matrix K, and a value matrix V as inputs.

We also adopt positionwise feedforward forward networks $FFN\left(M_x^l \right)$ to generate textual node states $H_x^{(l)}$

$$H_x^{(l)} = FFN\left(M_x^l \right), \tag{2}$$

where $M_x^{(l)} = \left[M_{xi}^{(x)} \right]$ denotes the above updated representations of all textual nodes.

On the side of the decoder, we employ a layer similar to the transformer decoder layer. In the method of (Vaswani et al., 2017), L_d identical layers are stacked, where each layer l is made up of three sublayers, to create target-side concealed states. To integrate the target and source-side contexts, the first two sublayers are masked self-attention and encoder-decoder attention:

$$E^l = MultiHead\left(S^{l-1}, S^{l-1}, S^{l-1} \right), \tag{3}$$

$$T^l = MultiHead\left(E^l, H_x^{(Le)}, H_x^{(Le)} \right), \tag{4}$$

where S^{l-1} denotes the target-side hidden states in the l-1-th layer. In particular, $S^{(0)}$ are the embeddings of the input target words. Then, a positionwise fully connected forward neural network is used to produce $S^{(l)}$ as follows:

$$S^{(l)} = FFN(S^{(l)}) \tag{5}$$

Finally, the probability distribution of generating the target sentence is defined by using a Softmax layer, which takes the hidden states in the top layer as input:

$$P(Y|X) = \prod_t Softmax(WS_t^{Ld} + b) \tag{6}$$

where X is the input sentence, Y is the target sentence (i.e., the Amharic sentence in our case), and W and b are the parameters of the Softmax layer.

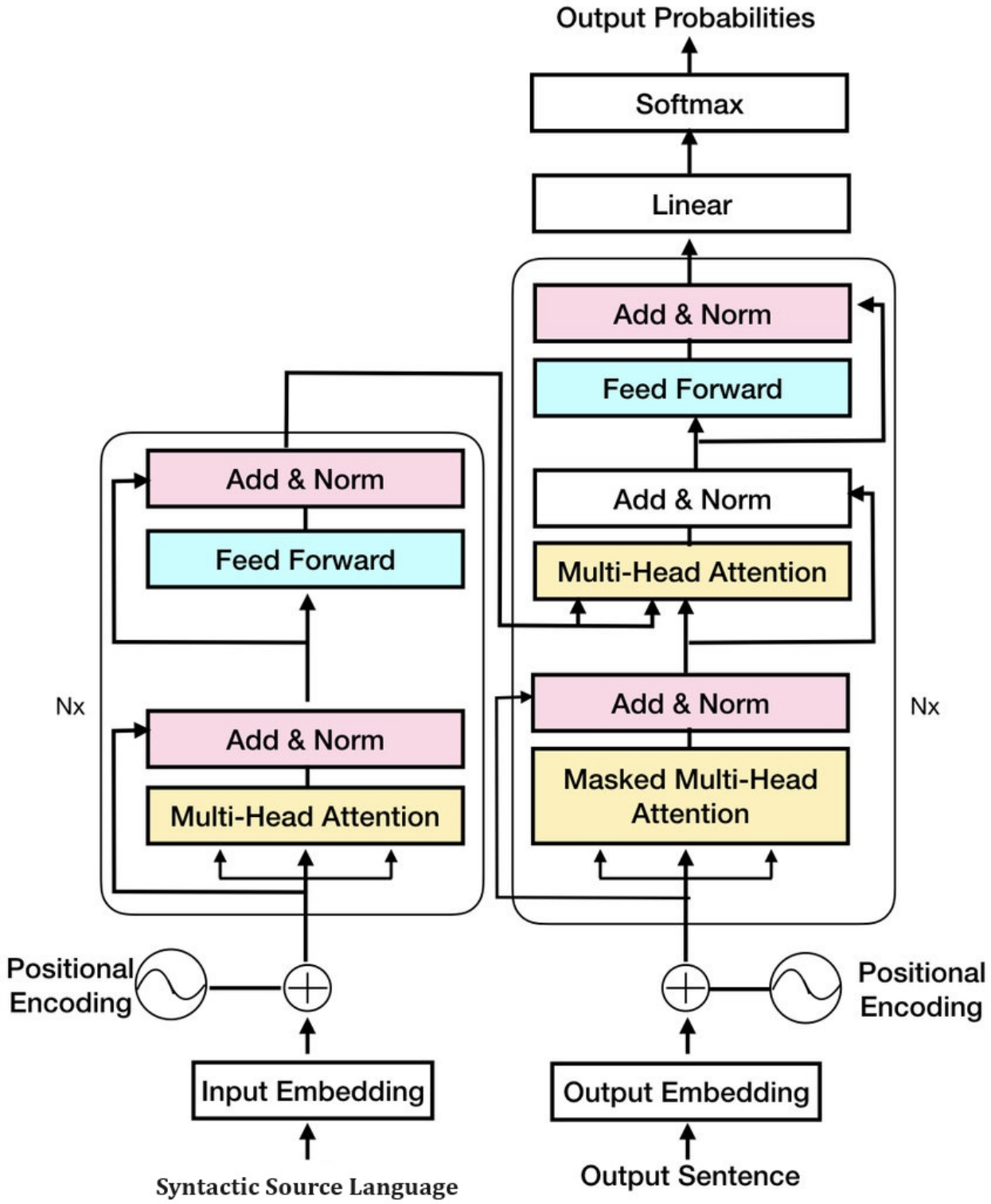

Fig 4. Transformer Encoder-Decoder architecture (source: Vaswani et al. (2017)); the proposed English-to-Amharic machine translation architecture

4 Result and discussion

In order to test the effectiveness of our system, we performed experiments using the attention-based Graph2Seq model [19]. With extensive experience capturing syntactic dependency trees within graph structures, this model was ideal for our evaluation. For training the model for Amharic to English translation, we used the Google Colab environment. We had to divide the parallel sentences for both English and Amharic into three parts: eighty percent for training, ten percent for validation, and the other ten percent for testing. The Adam optimizer [25] was utilized with a learning rate of 0.001 together with other parameters of batch size 32, hidden size 128, layer dropout probability of 0.1, and 50 training epochs. The validation for the model is done using the bilingual evaluation under study (BLEU) metric Papineni, Roukos [16]. It ranges from 0 to 1 and measures the similarity between the output of the translation and the given reference translated input. 1 indicates a perfect match while 0 stand for no matching words at all tested.

Transformer: we applied the OpenNMT framework for the TensorFlow deep learning environment for the implementation of the Transformer sequence-to-sequence NMT model from English to Amharic. The training process was performed from the ground-up [26]. We used Byte Pair Encoding which is a form of subword tokenization for text encoding. Byte Pair Encoding works on the principle of data compression where a pair of bytes that occur most frequently is replaced with a single byte that does not occur in the data. Different models were trained having 512 hidden units, 6 layers, learning rate of 0.0001, maximum step of 50K, batch size of 32 together with Adam optimization.

Pre-trained model: To develop our bi-directional English-to-Amharic NMT system, we utilized the multilingual Facebook M2M-100 pre-trained model with 418M parameter [17]. For fine-tuning, the training and validation were conducted with a maximum source and target length of 128 per device. We used a batch size of 4 and trained for 4 epochs.

4.1 Results

In our experimental setup, we utilized a methodically normalized dataset, ensuring consistency and reliability across our analyses. Leveraging the syntactic source language, we use Graph2seq transformer models, capitalizing on their efficacy in capturing complex syntactic dependency tree patterns. To establish a robust benchmark, we engaged in meticulous fine-tuning on M2M100 418 M, a state-of-the-art pretrained language model tailored for English to Amharic translation. This step provided a solid foundation (baseline) for our subsequent evaluations.

Furthermore, we use the attention-based GNN2Seq model and StanfordNLP, enriching the encoder with a syntactic dependency tree while maintaining the integrity of the decoder from the standard transformer architecture. This innovative approach allowed us to explore the nuanced influence of integrating syntax dependency into the graph neural networks of the encoder, revealing new insights into the interplay between syntactic structures and sequence generation. The following table shows the results.

Table 2. Experimental results for the BLEU score

Model	Direction	Result (BLEU Score)
Transformer	(English→Amharic)	13.06
M2M100 418 M	(English→Amharic)	32.74
GNN2Seq (with syntactic integration)	(English→Amharic)	37.3

Table 2 presents the results of three different translation models evaluated on the task of translating from English to Amharic, with the performance metric being the BLEU score, which is a common metric used to evaluate the quality of machine-translated text.

English→Amharic: The Seq2Seq (Transformer) model achieved a BLEU score of 13.06 while the pre-trained model achieved a BLEU score of 32.74, whereas the GNN2Seq model, with syntax integration, outperformed and achieved a higher BLEU score of 35.3. The BLEU score is a measure of how closely the generated translation matches human-generated reference translations. A higher BLEU score indicates better translation quality, with scores above 30 generally considered to be indicative of relatively good translation performance.

The improvement in the BLEU score from the Seq2Seq model to the GNN2Seq model suggests that incorporating syntax dependency tree into the graph neural networks of the encoder, as in the GNN2Seq model, leads to enhanced translation quality. This finding indicates that leveraging syntactic information during the translation process can improve the accuracy and fluency of the translated text. Additionally, the difference in BLEU scores between the two models provides valuable insight into the effectiveness of integrating syntactic information into neural machine translation models.

A qualitative error analysis reveals that GNN2Seq yields significant gains beyond the overall BLEU score improvement. First, syntactic alignment errors are notably reduced. In Amharic, the Transformer model often struggles with long-distance dependencies and complex clause structures. Incorporating syntactic graphs into GNN2Seq allows better capture of these dependencies, resulting in grammatically coherent outputs. Second, word order and agreement issues are addressed more effectively. Amharic's SOV (Subject-Object-Verb) structure contrasts with English's SVO. As compared to baseline models, the GNN2Seq model correctly reorders verbs and objects more frequently. Third, function word errors, particularly those involving prepositions and determiners, are reduced. Words with no direct English-Amharic correspondence often require contextual inference, which the syntactic model seems to be better at handling.

Example translations

Source (EN): "The teacher who spoke at the conference gave an inspiring talk."

- *Transformer:* "አስተማሪው በኮንፈረንሱ ተናግሯል እና ንግግሩ ደስ አሰኘ።" (Disjointed and unnatural ordering)
- M2M100: "አስተማሪው በኮንፈረንሱ ተናግሬ አነቃቂ ንግግር ሰጠ።" (Better, but still missing grammatical cohesion)
- GNN2Seq: "በኮንፈረንሱ የተናገረው አስተማሪ አነቃቂ ንግግር ሰጠ።" (Correct relative clause structure and natural word order)

This example illustrates how GNN2Seq preserves the relative clause and correctly restructures the sentence for natural Amharic expression, which standard models fail to do consistently.

Effect of sentence length: In Figure 4, sentence length, divided into four bins, is depicted on the x-axis, while the y-axis indicates the BLEU score, which measures accuracy for the three models evaluated: Transformer, M2M-100, and Graph2seq-SDT. The Transformer model (solid blue line) achieves around a BLEU score of 26 for shorter sentences, which indicates competitive performance. Such efficiency can be credited to self-attention mechanisms, which are especially good at capturing dependencies and structures—common in shorter sequences—within localized portions of the data. However, the model struggles with longer sentences. This observation implies that, while Transformers have a strong performance on short-range dependency models, they struggle with long-range sytnactic or semantic relationships without structural scaffolding.

M2M-100 (dashed red line) achieves the highest BLEU score for short sentences, which results from extensive multilingual pretrained representation learning. However, this model struggles with longer sentences. This indicates a steeper performance drop. M2M-100 seems to face challenges generalizing to complex structures in lower-resourced languages, such as English-Amharic, particularly when fine-tuning datasets lack diverse representations of complex data.

In comparison, Graph2seq-SDT (dotted green line) seems to lag behind on shorter sentences with a relatively low score. This may be because of the overhead it focuses on processing structural information, which might not be needed for simple inputs. Nevertheless, with longer sentences, Graph2seq-SDT exhibits lower decline in BLEU score and eventually surpasses M2M-100. This shows that its explicit integration of syntactic structure (e.g., through the use of dependency trees) is proficient in modeling long-range dependencies and complex grammatical constructs.

This portrays the relative robustness to length variation emphasizing the strength of syntactic modeling in MT low-resource contexts. Although every model faces the translation performance drop with additional sentence length, overall contracted, loose, straight, inconsecutive, or any other complex in nature Graph2seq-SDT proves more resilient. It enables more dependable MT for longer, more complicated sentences, such as those encountered in advanced academic writing.

Comparing with other English-Amharic studies: Table 3 summarizes various studies on machine translation between English and Amharic or related languages, comparing the datasets used, the methodologies applied, and the resulting BLEU scores. The methods range from traditional statistical machine translation (SMT) and phrase-based SMT to more advanced neural machine translation (NMT) techniques and the use of pre-trained models like M2M100. The table also highlights the size of the datasets used in each study, showing a wide variation from as few as 1,915 sentence pairs to over a million, as well as the BLEU scores that measure the quality of the translations produced.

The earlier works of Abate, Melese[27] and Teshome, Besacier[31] focused on statistical and phrase-based techniques which, although useful at the time, have problems with long-range contextual dependencies. These models suffer from poor generalization due to their reliance on exact phrase matching, which contributes to their rather modest BLEU scores of 13.31 and 35.32 respectively.

Incorporating context-based translation, as advanced by Ashengo, Aga[29], employed recurrent networks, but the results were hampered by the very small dataset of only 8,603 sentence pairs, which constrained the model's ability to learn more complex linguistic relationships, as reflected in the underwhelming BLEU score of 11.34. With an equally small dataset of 1,915 sentences, Hadgu, Beaudoin[30] took a black-box approach using Google Translate, yielding a BLEU score of 9.6, further persisting the notion that generic models are not sufficient for morphologically rich low-resource languages such as Amharic.

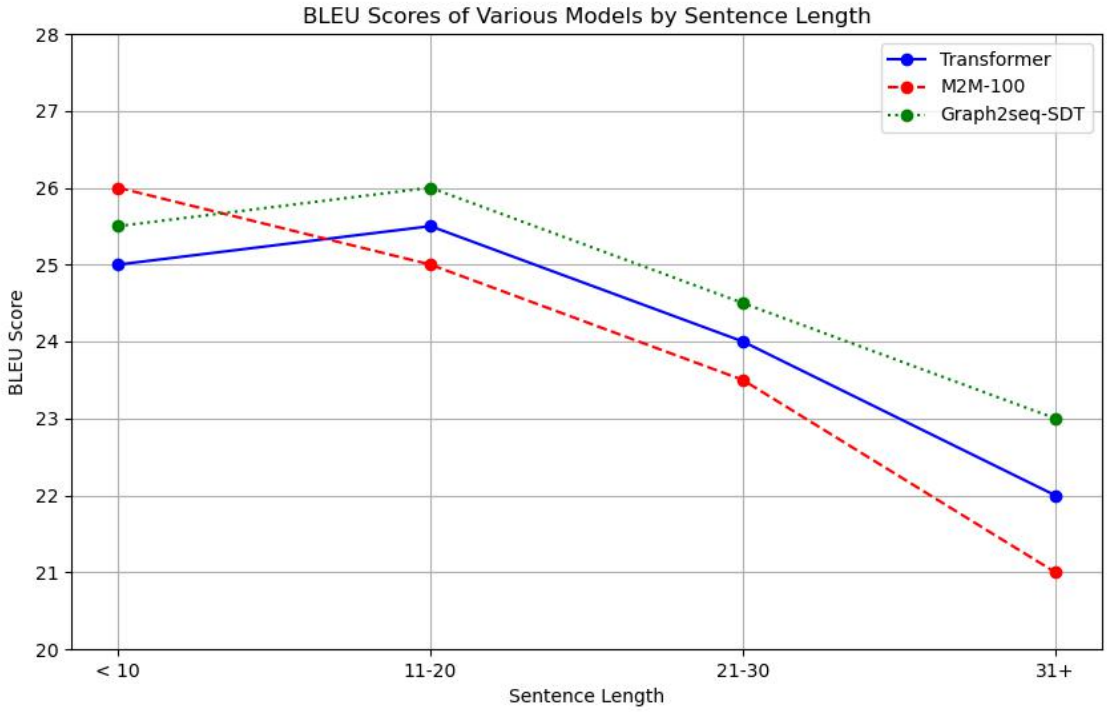

Fig 5. displays a line graph titled "BLEU Scores of Various Models by Sentence Length." It compares the performance of three models: Transformer, M2M-100, and Graph2seq-SDT, based on BLEU scores across different sentence lengths

Table 3. Previous studies on English-Amharic machine translation have been assessed in terms of dataset size, method(s) used, and the BLEU score achieved

Authors	# Dataset used	Method(s)	BLEU score
Biadgligne and Smaili [18]	225,304	Neural machine translation	32.44
Abate, Melese [27]	40,726	Statistical machine translation	13.31
Teshome and Besacier [28]	18,432	Phrase-based statistical machine translation	35.32
Ashengo, Aga [29]	8,603	Combination of context-based MT (CBMT) with RNN	11.34
Hadgu, Beaudoin [30]	1915	Google translate	9.6
Belay, Tonja [17]	1,140,130	M2M100 418 M fine-tuning pre-trained model	32.74
Our work	1,140,130	Attention-based Graph2seq	37.3

Biadgligne and Smaili [18]'s recent work utilized neural machine translation (NMT) technology on a moderately sized dataset of 225,304 pairs achieving a BLEU score of 32.44 which shows some potential, yet still remains constrained by the scale of data and the architecture of the model used. Belay, Tonja [17] made some progress by applying M2M100 418M multi lingual model with the corpus of over 1.14M sentences and attaining BLEU score of 32.74. Such models, however powerful, treat syntax in a non-conscious manner and are not built up for the pairs that have highly different grammatical arrangements.

In our case, the proposed model tackles these limitations using the attention mechanism of Graph2Seq to embed syntactic information. The model can capture hierarchical and long-range dependencies in the source language more effectively, which is crucial for translating between English (SVO) and Amharic (SOV), which are structurally divergent languages. The same dataset described in (18) along with linguistic priors put the syntax-aware model at a BLEU score of 37.3, outperforming all previous attempts, showcasing the syntax-aware modeling advantage.

4.2 Discussion

The results of my research show great progress in English to Amharic Neural Machine Translation (NMT) after addition of syntactic dependency trees into the Graph2Seq (GNN2Seq) model. This method outperformed both the Transformer based models and the pretrained M2M100 baseline models. For translation improvement, the results substantiate the role of explicit syntactic modeling especially for rich morphology and syntactically distant languages like Amharic.

By examining the BLEU scores, a comparative analysis shows a distinct performance ranking amongst the models. The Transformer-based model (OpenNMT) scored 13.06 BLEU, while the pre-trained M2M100 (418M parameters) outscored it by a significant margin at 32.74. With further integration of the proposed GNN2Seq model, that adds syntactic dependencies, the translation quality increased to a BLEU score of 37.3. It is evident from these cases that large multilingual corpora pre-training remains beneficial as shown by the M2M100's overwhelming improvement over the Transformer model. However, GNN2Seq's enhancement shows that explicit syntactic modeling through dependency trees improves translation accuracy beyond pre-training.

Also, when compared with earlier works, the current model performs at the state-of-the-art level, beating Biadgligne and Smaili[18] NMT model (32.44 BLEU) and Teshome and Besacier[28]'s phrase-based SMT method (35.32 BLEU). This improvement demonstrates how effective hybrid models that comprise both graph-based encoders and attention-centric components are.

The increase in translation quality can be explained by the capability of syntactic integration to solve linguistic problems of Amharic language. Amharic is a morphologically complex language with an SOV (Subject-Object-Verb) structure and poses a lot of difficulty to standard Transformer models that use sequential encoders. The graph encoder GNN2Seq incorporates in the model preserves long distance dependency and reduces the loss of information during the word rearrangement process. This structural advantage is crucial for maintaining the grammatical correctness of translated sentences.

The analysis of sentence length and its effect on the BLEU score reveals other patterns. All models decline in performance with longer sentences, but the rate of decline differs among the models. The Transformer model achieves a high BLEU score (26) for shorter sentences, but with more complex sentences, that score declines sharply. M2M100 also seems to drop performance, with fixed-length attention bottlenecks making it less capable of dealing with longer sentences.

On the other hand, the GNN2seq model continues to show a relatively smooth decline of BLEU scores which indicates that having syntactic structures integrated in the encoder makes one more robust to long and complex sentences. Such models are particularly beneficial to low-resource languages because these languages struggle the most with training data. These models also remind us of the importance of pretrained models like M2M100 which need to implement changing attention mechanisms in order to reduce drop in performance as the length of the sentence increases.

A broader look at prior works, compare with Table 3 data, provides additional context for the context of these improvements. The dataset size plays a major part as the 1.14 million parallel sentence pairs in this work greatly surpass the majority of earlier studies which often had less than 50,000 parallel sentences. There is also a clear change in the evolution of methodologies; for example, Abate, Melese[27]'s Statistical Machine Translation (SMT) approach was incredibly low, at 13.31 BLEU due to the lack of adequate capturing context and syntactic features. More recent NMT models, like Biadgligne and Smaili[18], showed incredible improvement, at 32.44 BLEU, but did not still include the explicit use of syntax.

The GNN2Seq model has set a new record in the English-to-Amharic translation task along with its hybrid encoders by graphing in a sequence and attention modeling, scoring 37.3 in the BLEU score. In summary, the findings from this particular study assert the important influence that syntactic modeling has on NMT enhancement for diverse language pairs with different constructs. The addition of syntax-aware features into the model improves the translation accuracy and also increases the accuracy for robustness against sentence length. These give essential contributions towards further work in machine translation, especially for languages with low resources that for many complex linguistic patterns, data driven systems can be ineffective.

5 Conclusion

This study shows that the addition of explicit synactic knowledge into neural machine translation (NMT) systems greatly enhances the quality of English-to-Amharic translations. Our proposed model, Graph2Seq (GNN2Seq), achieves a BLEU score of 37.3 which is 4.56 points higher than the M2M100 pretrained model and 24.24 points higher than the baseline transformer model. These results reinforce the advantage of using linguistic structures in translating, especially in morphologically complex and syntactically diverse languages such as Amharic.

The primary contribution of this work is the syntactically-informed Graph2Seq architecture that employs a hybrid encoder-decoder design with dependency tree structures. The model incorporates graph neural networks (GNNs) in the encoder with a transformer based decoder to overcome the problem of word-order in translating from English (SVO) to Amharic (SOV).

This demonstrates that syntactic graphs, that capture word order, greatly improve the modeling of long range dependencies compared to more traditional approaches or pre-trained counterparts.

GNN2Seq is also robust to modifications in sentence length. GNN2Seq's performance is better than that of the transformer and M2M100 models when the complexity of sentences increases. This implies that the ability to capture syntactic dependencies aids in better translation for languages which have free word orders. All these point towards the fact that incorporating deep syntactic structures enrich Graph-NMT systems on low-resource languages enhances their capabilities, paving avenues for further investigations in the field of Machine Translation. This study can be a starting point towards more research on the enhancement of accuracy and fluency of translations using a multidisciplinary approach based on language.

Further work could be on the cheap and easy to access lightweight syntactic integration and human-in-the-loop refinement. Some other possible changes could be along the lines of syntax, such as merging semantic role labeling with dependency parsing to deepen the linguistic aspects.

6 Acknowledgement

A preprint has previously been published.

References

1) Jooste W, Haque R, Way A. Philipp Koehn: Neural Machine Translation. *Machine Translation*. 2021;35(2):289–299. https://doi.org/10.1017/9781108608480.
2) Wan Y, Yang B, Wong DF, Chao LS, Yao L, Zhang H, et al. Challenges of neural machine translation for short texts. *Computational Linguistics*. 2022;48(2):321–342. Available from: https://aclanthology.org/2022.cl-2.3/.
3) Bahdanau D, Chorowski J, Serdyuk D, Brakel P, Bengio Y. End-to-end attention-based large vocabulary speech recognition. IEEE. 2016. Available from: https://arxiv.org/abs/1508.04395.
4) Bastings J, Titov I, Aziz W, Marcheggiani D, Sima'an K. Graph convolutional encoders for syntax-aware neural machine translation. *arXiv preprint arXiv:170404675*. 2017. Available from: https://arxiv.org/abs/1704.04675.
5) Sutskever I, Vinyals O, Le QV. Sequence to sequence learning with neural networks. *Advances in Neural Information Processing Systems*. 2014;27. Available from: https://arxiv.org/abs/1409.3215.
6) Chen K, Wang R, Utiyama M, Sumita E. Integrating prior translation knowledge into neural machine translation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. 2021;30:330–339. https://doi.org/10.1109/TASLP.2021.3138714.
7) Yang S, Wang Y, Chu X. A survey of deep learning techniques for neural machine translation. *arXiv preprint arXiv:200207526*. 2020.
8) Gerencheal B, Mishra D. Foreign Languages in Ethiopia: History and Current Status. 2019. Available from: https://eric.ed.gov/?id=ED592409.
9) Aklilu YY. Exploring Neural Word Embeddings for Amharic Language. *NEAR EAST UNIVERSITY*. 2019;p. 1–20. Available from: https://docs.neu.edu.tr/library/6747057286.pdf.
10) Eriguchi A, Hashimoto K, Tsuruoka Y. Incorporating source-side phrase structures into neural machine translation. *Computational Linguistics*. 2019;45(2):267–292. https://doi.org/10.1162/coli_a_00348.
11) Stahlberg F. Neural machine translation: A review. *Journal of Artificial Intelligence Research*. 2020;69:343–418. https://doi.org/10.1613/jair.1.12007.
12) Asefa SH, Assabie Y. Transformer-Based Amharic-to-English Machine Translation with Character Embedding and Combined Regularization Techniques. *IEEE Access*. 2024;13(99):1090–1105. http://dx.doi.org/10.1109/ACCESS.2024.3521985.
13) Gezmu AM, Nürnberger A, Bati TB. Extended parallel corpus for Amharic-English machine translation. *arXiv preprint arXiv:210403543*. 2021. Available from: https://arxiv.org/abs/2104.03543.
14) Lakew SM, Negri M, Turchi M. Low resource neural machine translation: A benchmark for five african languages. *arXiv preprint arXiv:200314402*. 2020.
15) Wang X, Tu Z, Wang L, Shi S. Exploiting sentential context for neural machine translation. *arXiv preprint arXiv:190601268*. 2019. Available from: https://arxiv.org/abs/1906.01268.
16) Papineni K, Roukos S, Ward T, Zhu WJ. Bleu: a method for automatic evaluation of machine translation. *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*. 2002. Available from: https://aclanthology.org/P02-1040/.
17) Belay TD, Tonja AL, Kolesnikova O, Yimam SM, Ayele AA, Haile SB, et al. The effect of normalization for bi-directional amharic-english neural machine translation. IEEE. 2022. Available from: http://dx.doi.org/10.48550/arXiv.2210.15224.
18) Biadgligne Y, Smaili K. Boosting English-Amharic Machine Translation Using Corpus Augmentation and Transformer. *Interciencia*. 2024. http://dx.doi.org/10.59671/MBULJ.
19) Bignoli A. Graph attentional neural network in multi-agent pickup and delivery problems. 2021.
20) Lin J, Sun X, Ren X, Li M, Su Q. Learning when to concentrate or divert attention: Self-adaptive attention temperature for neural machine translation. *arXiv preprint arXiv:180807374*. 2018. Available from: https://arxiv.org/abs/1808.07374.
21) Tan X, Ren Y, He D, Qin T, Zhao Z, Liu TY. Multilingual neural machine translation with knowledge distillation. *arXiv preprint arXiv:190210461*. 2019. Available from: https://arxiv.org/abs/1902.10461.
22) Nguyen LH, Pham V, Dinh D. Integrating AMR to neural machine translation using graph attention networks. IEEE. 2020. http://dx.doi.org/10.1109/NICS51282.2020.9335896.
23) Veličković P, Cucurull G, Casanova A, Romero A, Lio P, Bengio Y. Graph attention networks. *arXiv preprint arXiv:171010903*. 2017.
24) Gong Z, Zhou K, Zhao WX, Sha J, Wang S, Wen JR. Continual pre-training of language models for math problem understanding with syntax-aware memory network. *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2022. http://dx.doi.org/10.18653/v1/2022.acl-long.408.
25) Tato A, Nkambou R. Improving adam optimizer. 2018. http://dx.doi.org/10.13140/RG.2.2.21344.43528.

26) Abrishami M, Rashti MJ, Naderan M. Machine translation using improved attention-based transformer with hybrid input. IEEE. 2020. https://doi.org/10.1109/ICWR49608.2020.9122317.
27) Abate ST, Melese M, Tachbelie MY, Meshesha M, Atinafu S, Mulugeta W, et al. Parallel corpora for Bi-Lingual English-Ethiopian languages statistical machine translation. *Proceedings of the 27th International Conference on Computational Linguistics*. 2018. Available from: https://aclanthology.org/C18-1262/.
28) Teshome MG, Besacier L. Preliminary experiments on English-Amharic statistical machine translation. *SLTU*. 2012.
29) Ashengo YA, Aga RT, Abebe SL. Context based machine translation with recurrent neural network for English–Amharic translation. *Machine translation*. 2021;35(1):19–36. https://doi.org/10.1007/s10590-021-09262-4.
30) Hadgu AT, Beaudoin A, Aregawi A. Evaluating amharic machine translation. *arXiv preprint arXiv:200314386*. 2020.
31) Teshome MG, Besacier L, Taye G, Teferi D. Phoneme-based English-Amharic statistical machine translation. IEEE. 2015. http://dx.doi.org/10.1109/AFRCON.2015.7331921.