Enhancing English to Amharic Machine Translation with Prior Knowledge Integration

Muluken Hussen Asebel

A Dissertation Submitted to the Department of Computer Science and Engineering,

School of Electrical Engineering and Computing

Presented in Partial Fulfillment of the Requirement for the Degree of Doctor of Philosophy in Computer Science and Engineering

Office of Graduate Studies

Adama Science and Technology University

December 2025

Adama, Ethiopia

Enhancing English to Amharic Machine Translation with Prior Knowledge Integration

Name of Candidate:

Muluken Hussen Asebel

Supervisors:

Shimelis Getu Assefa (PhD)

Mesfin Abebe Haile (PhD)

A Ph. D. Dissertation Submitted to the department of Computer Science and Engineering, School of Electrical Engineering and Computing.

Presented in Partial Fulfillment of the Requirement for the Degree of Doctor of Philosophy in Computer Science and Engineering

Office of Graduate Studies

Adama Science and Technology University

December 2025

Adama, Ethiopia

Declaration

I hereby declare that this dissertation entitled *"Enhancing English to Amharic Machine Translation with Prior Knowledge Integration"* is my original work. That is, it has not been submitted for the award of any academic degree, diploma or certificate in any other university. All sources of materials used for this dissertation have been duly acknowledged through appropriate citations.

Muluken Hussen Asebel	______________________	______________________
Name of Student	Signature	Date

Recommendation

I/we, the supervisor(s) of this dissertation, hereby certify that I/we have read and revised the dissertation entitled *"Enhancing English to Amharic Machine Translation with Prior Knowledge Integration"* prepared under my/our guidance by Muluken Hussen Asebel, submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Computer Science and Engineering. Therefore, I/we recommend the submission of the dissertation to the department for further review and defense.

Dr. Shimelis Getu Assefa ______________________ ______________________
Major Supervisor Signature Date

Dr. Mesfin Abebe Haile ______________________ ______________________
Co-supervisor Signature Date

Approval Sheet

I/We hereby certify that the recommendations and suggestions made by the board of examiners are appropriately incorporated into the final version of the dissertation entitled *"Enhancing English to Amharic Machine Translation with Prior Knowledge Integration"* by Muluken Hussen Asebel.

Dr. Shimelis Getu Assefa Major Supervisor	____________________ Signature	____________________ Date
Dr. Mesfin Abebe Haile Co-supervisor	____________________ Signature	____________________ Date

We, the undersigned, members of the Board of Examiners of the dissertation open defense by Muluken Hussen Asebel have read and evaluated the dissertation entitled *"Enhancing English to Amharic Machine Translation with Prior Knowledge Integration"*. This is therefore to certify that the dissertation is accepted for partial fulfillment of the requirement for the degree of Doctor of Philosophy in Computer Science and Engineering.

____________________ Chairperson	____________________ Signature	____________________ Date
____________________ Internal Examiner	____________________ Signature	____________________ Date
____________________ External Examiner 1	____________________ Signature	____________________ Date
____________________ External Examiner 2	____________________ Signature	____________________ Date

Finally, approval and acceptance of the dissertation is contingent upon submission of its final version to the Office of Postgraduate Studies (OPGS) through the candidate's Department Graduate Council (DGC) and School Graduate Committee (SGC).

____________________ School Dean	____________________ Signature	____________________ Date
____________________ Office of Postgraduate Studies, Dean	____________________ Signature	____________________ Date

Acknowledgment

First and foremost, I would like to express my deepest gratitude to the Almighty Allah for granting me the strength, health, and perseverance to complete this research work. Without His guidance and blessings, this accomplishment would not have been possible.

I sincerely thank my advisors, Dr. Shimelis Getu Assefa and Dr. Mesfin Abebe Haile, for their continuous support, insightful guidance, and constructive feedback throughout this research. Their expertise and encouragement were invaluable at every stage of my academic journey. I am also deeply grateful to my family for their unwavering love, prayers, and moral support. Their belief in me kept me motivated, especially during the challenging moments.

Finally, I would like to thank my friends and colleagues for their companionship, helpful discussions, and all the moments of laughter and encouragement that made this journey enjoyable.

Thank you all.

Table of Contents

List of Figures

List of Tables

List of Acronyms and Abbreviations

BLEU	Bilingual Evaluation Understudy
CBMT	Context-Based Machine Translation
FDRE	Federal Democratic Republic of Ethiopia
GAT	Graph Attention Network
GCN	Graph Convolutional Network
GGNN	Gated Graph Neural Network
GNN	Graph Neural Network
Graph2seq	Graph to Sequence model
GRU	Gated Recurrent Unit
LSTM	Long Short-Term Memory
M2M100	Facebook AI's Many-to-Many multilingual translation model
METEOR	Metric for Evaluation of Translation with Explicit ORdering
ML	Machine Learning
MT	Machine Translation
NLP	Natural Language Processing
NMT	Neural Machine Translation
PBMT	Phrase-Based Machine Translation
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RBMT	Rule-Based Machine Translation
RNN	Recurrent Neural Network
SDT	Syntax Dependency Tree
SMT	Statistical Machine Translation
SOTA	State-of-the-art
SOV	Subject-Object-Verb
SVO	Subject-Verb-Object

Abstract

English and Amharic serve as widely used languages in Ethiopia and belong to distinct linguistic families: English is an Indo-European language, whereas Amharic is a Semitic language within the Afro-Asiatic family. This linguistic divergence poses substantial challenges for machine translation, particularly due to Amharic's rich morphology, non-Latin script, and subject-object-verb (SOV) syntactic structure. Existing neural machine translation (NMT) systems often struggle to model these characteristics effectively, resulting in inadequate alignment, word order errors, and reduced translation fluency. This study addresses these challenges by integrating prior syntactic knowledge into English–Amharic machine translation through a Graph-to-Sequence (Graph2Seq) framework. Specifically, the proposed model incorporates syntactic dependency trees of the source language to enhance the representation of grammatical relationships and long-distance dependencies. To evaluate this approach, the study utilizes a large-scale parallel corpus comprising over 1.14 million English-Amharic sentence pairs, divided into training (70%), validation (10%), and testing (20%) sets. The proposed Graph2Seq model is evaluated against a standard Transformer model and the pretrained M2M100 multilingual model. Experimental results demonstrate substantial improvements in translation quality: the Graph2Seq model achieves a BLEU score of 37.30, significantly outperforming the Transformer model (13.06) and surpassing the M2M100 model (32.74). Qualitative and quantitative analyses indicate that incorporating syntactic dependency structures reduces alignment errors, improves word ordering, and enhances the handling of long-distance dependencies. Overall, the findings confirm that embedding syntactic prior knowledge through Graph Neural Networks markedly improves English-Amharic machine translation performance. This work highlights the effectiveness of graph-based approaches for morphologically rich and low-resource languages and provides a foundation for future research. Potential extensions include integrating semantic role labeling, expanding and refining parallel corpora, and developing computationally efficient models suitable for resource-constrained environments. By addressing linguistic structure explicitly, this study advances the development of more accurate, fluent, and linguistically informed graph neural machine translation systems.

Keywords: Machine Translation, English–Amharic Translation, Linguistic Prior Knowledge, Syntactic Dependency Trees, Graph Neural Networks, Low-Resource Languages.

Chapter One: Introduction

1.1Background of the Study

Human beings have vast variety of languages as a communication tool to convey information and arguments to others. Languages are specific to the countries or nations; hence, they need to have a mechanism for translating one language to another for the betterment of communication. Translation facilities provide the ability for two parties to exchange and communicate thoughts from different countries or nations without language barriers. It supports relationships between individuals, improving professional associations and expanding their social network. Translation can be human or machine translation (Hutchins, 2001). Human translation is the most accurate and can understand the context of that text or speech, but machine translation still needs a lot of work to perform like human translation (Hutchins, 2001). Machine translation decomposes source texts into linguistic units to generate accurate target-language representations, thereby enabling effective and consistent communication across different formats. In terms of performance, availability, and cost machine translation is better than human translation. That is why we strive to have the best machine translation agent.

The application of computers to the translation of natural languages was first proposed in the 1940s, laying the foundation for what later became the field of machine translation. Since the initiation of systematic research in the 1950s, machine translation has undergone substantial development, often observed by human translators with varying degrees of doubt and worry (Hutchins, 2001). Despite notable progress driven by advances in statistical and neural approaches, these developments have largely benefited well-resourced languages. Ethiopian languages in general, and Amharic in particular, remain under-resourced in the context of machine translation. As a result, further in-depth investigation is required to reduce this disparity and to advance translation quality to a level comparable with that achieved for languages such as English (S. T. Abate et al., 2018).

Natural language processing (NLP) studies how computers represent, process, and understand human languages to enable effective human-computer communication (Banea et al., 2008). One major research area within NLP is machine translation, which focuses on developing computational systems capable of automatically translating text from a source language into a target language. To produce accurate translations, a machine translation system must correctly analyze the source-language text and generate a semantically equivalent and contextually appropriate representation in the target language. Translation therefore involves mapping both meaning and linguistic structure from the source language to the target language after the source text has been computationally processed. This process requires capturing lexical, syntactic, and semantic relationships, as well as modeling how words and structures in one language correspond to those in another. Accordingly, effective translation depends on deep linguistic knowledge of both the source and target languages. Researchers have implemented machine translation systems using a range of approaches, including rule-based, example-based, statistical, machine learning–based, and hybrid methods (S. T. Abate et al., 2018).

In Ethiopia, English and Amharic coexist as widely used languages, representing substantial linguistic diversity that poses significant challenges for machine translation. English, a member of the Indo-European language family, functions as a foreign language and serves as the medium of instruction in secondary and higher education, as well as a primary language of communication in governmental, non-governmental, regional, and international organizations. In contrast, Amharic belongs to the Semitic branch of the Afro-Asiatic language family and is widely spoken across the country. It also serves as the official working language of the Federal Democratic Republic of Ethiopia. Consequently, effective translation systems and human translators proficient in at least one of these languages are essential to support communication across educational, administrative, and institutional domains (S. T. Abate et al., 2018).

Machine translation has evolved through several interrelated paradigms that reflect advances in computational linguistics and artificial intelligence. Researchers have developed major approaches such as Rule-Based Machine Translation (RBMT), Statistical Machine Translation (SMT), Hybrid Machine Translation (HMT), and Neural Machine Translation (NMT), each introducing distinct modeling assumptions and learning mechanisms. More recently, graph-based neural approaches have extended NMT by incorporating structured linguistic representations. In particular, Graph Neural Networks (GNNs) model translation using graph-theoretic structures rather than purely sequential or grid-based representations, enabling more effective encoding of syntactic and relational information.

RBMT emerged as the earliest machine translation paradigm, relying on manually constructed linguistic rules and bilingual dictionaries to translate text through syntactic, semantic, and morphological analysis of source and target languages. While RBMT can generate grammatically coherent translations, it demands substantial linguistic expertise, extensive rule engineering, and continual manual refinement, which limits its scalability and adaptability across diverse language pairs (Dorr et al., 1999).

SMT uses statistical models to analyze large corpora of bilingual texts, identifying correspondences between words in the source and target languages. By framing translation as a machine-learning problem, SMT algorithms learn to translate by analyzing numerous examples of human-produced translations. Over the past two decades, SMT has achieved substantial progress (Z. Tan et al., 2020). For instance, M. G. Teshome et al. (2015a) applied SMT algorithms to Amharic-English translation, improving translation quality through phonemic transcription on the target side. In contrast, our research emphasizes the syntactic structure of the English language, integrating Graph Neural Networks (GNN) into our methodologies.

HMT integrates components of both Rule-Based Machine Translation (RBMT) and Statistical Machine Translation (SMT) to leverage their complementary strengths. By incorporating resources such as translation memories, HMT systems often achieve improved translation quality. However, HMT remains limited by its dependence on extensive manual editing and human intervention. Researchers have implemented HMT using various strategies, including multi-engine architecture, statistical rule induction, multi-pass processing, and confidence-based approaches (Dorr et al., 1999).

NMT employs neural network architectures to model translation in an end-to-end manner. Unlike earlier approaches, NMT integrates the representation and generation of both source and target texts within a single unified framework, reducing reliance on modular components commonly used in Statistical Machine Translation (SMT) systems (Dorr et al., 1999). Advances in graph-based machine learning have further extended the capabilities of NMT, enabling it to address complex linguistic phenomena more effectively. These developments are particularly promising for structurally complex and low-resource languages such as Amharic, which pose significant challenges for traditional machine learning approaches. In this context, Graph Neural Networks

(GNNs) have demonstrated strong performance across a range of natural language processing tasks that require structured reasoning and the incorporation of prior knowledge (Tu et al., 2020).

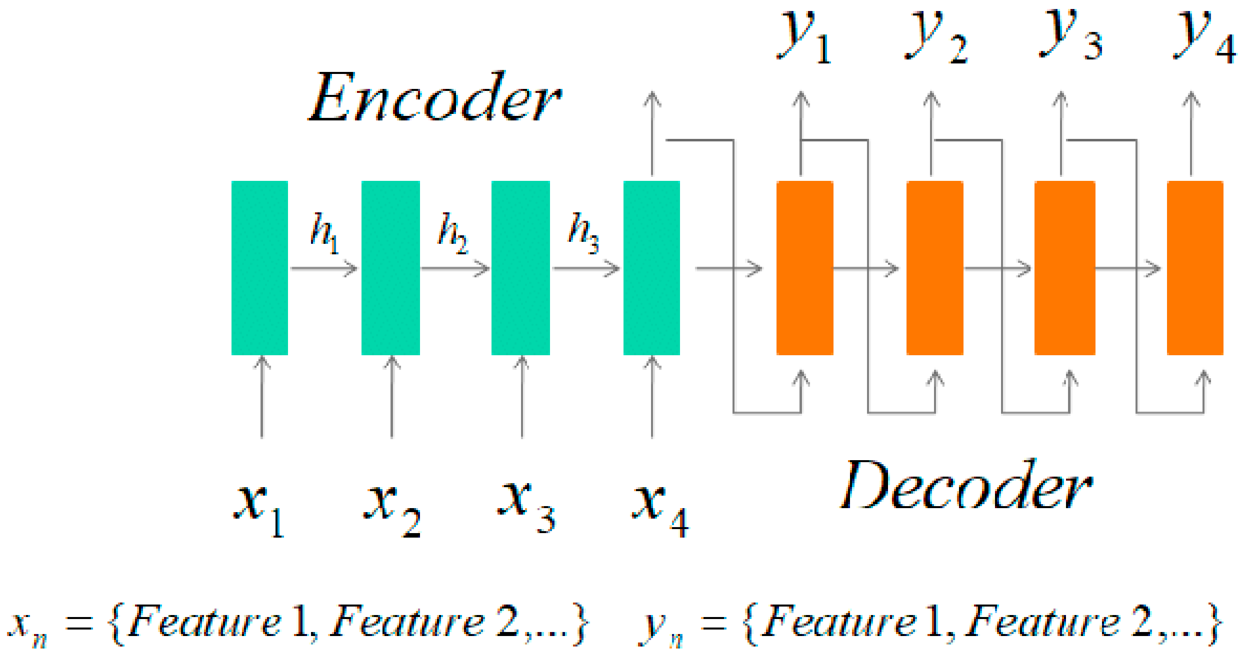


Figure 1.1: Sequence to sequence model of NMT.

As illustrated in Figure 1.1, Neural Machine Translation (NMT) typically represents input text as a sequence of tokens. However, many natural language phenomena exhibit rich structural relationships that sequence-to-sequence models cannot adequately capture. Graph-based representations offer a more expressive modeling framework for such problems (Wu et al., 2023). Accordingly, Graph Neural Networks (GNNs) enable machine translation systems to encode these structures effectively and to integrate prior linguistic knowledge into the encoder component of NMT models.

Amharic uses a adaptation of the Geʿez script, which consists of 33 consonants, each with seven distinct forms based on the vowel pronounced in the syllable, resulting in over 231 unique characters. This orthographic complexity, coupled with the language's unique syntactic and morphological structures, makes Amharic more challenging for machine translation compared to English. However, researchers identify significant potential for improving machine translation models through the application of graph learning algorithms, which can better capture the intricate relationships within the language's structure (Wu et al., 2023).

The irregularity of sentence construction in Amharic poses another challenge for translation models. For example, the sentences "አበበ ከበደን መታው።" and "ከበደን አበበ መታው።" have the same meaning ”Abebe hit Kebede” but differ in words order. In languages like English, changing the order of words typically alters the meaning of a sentence. But, In Amharic word ordering may not change the meaning of sentence like the example shown above.

Previous attempts to improve translation models used methods other than graph neural networks, which resulted in low performance. Therefore, researchers believe that the GNNs with prior knowledge integration approach can fill this gap. To the best of our knowledge, no previous research projects have utilized GNNs for English-Amharic machine learning to extract information from prior knowledge graphs to enhance translation models.

1.1.1 Amharic Language and the Imperative for Machine Translation Research

Amharic is spoken primarily in Ethiopia and serves as the Ethiopia federal government’s working language (Ababa, 2011). Classified as a South Semitic language within the Afroasiatic language family, Amharic has approximately 32 million native speakers and a substantial number of second-language users (M. G. Teshome & Besacier, 2012a), making it the second most widely spoken Semitic language globally after Arabic. The language plays a central role in Ethiopia’s cultural, social, administrative, and educational systems, functioning as the national language, the medium of instruction in schools, the language of government, and the primary medium for literature and cultural expression.

The complexity of Amharic increases due to its use of the Geʿez script, a syllabic writing system that differs fundamentally from the Latin alphabet (M. G. Teshome & Besacier, 2012a). Geʿez characters consist of base consonantal forms that vary systematically to encode syllabic information, which poses significant challenges for computational representation and processing. In addition, Amharic language exhibits rich morphological structure characterized by extensive inflection and derivation, combining both agglutinative and fusional features.

The engagement of researchers and practitioners in English-to-Amharic MT has intensified owing to the need to bridge communication and information dissemination gaps resulting from globalization (Melese et al., 2017). Machine translation, unlike human translation, does not require a translator and acts as a sophisticated technology that bridges the gap between languages. This modern-day aid enhances communication across cultures, enables knowledge transfer, and fosters inclusivity around the world (Shadiev et al., 2019). Additionally, constructing efficient MT systems for the Ethiopian languages helps in the preservation and wider access of Ethiopian cultural and intellectual property, hence greatly contributing to enriching multicultural resources available online.

Developing machine translation for Amharic poses both challenges and opportunities from the standpoint of computational linguistics. The more challenging aspects of the morphologically rich language inflection and syntax make it impossible for models designed for more analytically appealing languages to be directly used. But overcoming these difficulties can lead to innovation in the methods used to capture processes in NLP, especially with languages that have few resources (Nigusie, 2024). Progress in this field not only has advantages for speakers of the Amharic language but also provides strategies for the advancement of language processing technologies to be deployed among other marginalized communities of languages.

The modern history of machine translation research indicates a dominant shift towards neural machine translation for heavily resourced languages and language pairs due to its proven effectiveness. Many, including Amharic, remain less represented due to data scarcity and linguistic diversity. Lack of quality parallel corpora and the complex underlying linguistic features of these languages, such as dialectal variation and morphosyntactic ancestry compound challenges (Iyer et al., 2024; Nzeyimana, 2024). These constraints render traditional NLP pipelines ineffective, creating the need for new architecture-driven solutions and more novel datasets.

To address these challenges, researchers have introduced effective strategies such as back-translation, synthetic data generation, fine-tuning of pretrained multilingual models, and other transfer learning techniques. Empirical evidence demonstrates that these approaches yield substantial performance gains for low-resource languages, including notable improvements in BLEU scores (Prasada & Rao, 2024; Tran, 2024). These results underscore the importance of developing strategically efficient and cost-effective training frameworks that leverage existing linguistic resources while incorporating novel modeling concepts.

Achieving equitable access to language technologies demands proactive attention to low-resource NLP. Closing the innovation gap for these languages requires the development of reliable resources, such as automated corpus generation, semi-supervised learning, and cross-lingual knowledge transfer. Moreover, collaboration among linguists, computer scientists, and regional populations is important in the construction of vocabulary and annotated resources that capture rich detail (Joshi et al., 2024). With these actions in place, NLP as well as MT systems can be developed to suit sociocultural, geographic, and multilingual realities across the world.

Ultimately, improving machine translation for English-Amharic presents not only a linguistic and technical problem but a cultural and moral one as well. Tackling the issues related to lack of information, complexity of the language, and adaptation of the model could help create a more equitably accessible digital world, one that truly represents and values all communities and their respective cultures.

1.1.2 Significance of Machine Translation

Machine translation's core aim is to foster interaction among people who use different languages. Work in this area seeks to automate translation. This advancement significantly helps with under-sourced languages such as Amharic by forming corresponding collections for each language. Amharic, the official working language of the Ethiopian federal government and the Amhara Regional State, can use machine translation to transform several educational documents. This advancement aids government institutions, such as higher education institutions, in language research.

Moreover, many mainstream media outlets broadcast content in both Amharic and other languages, which increases the need for machine translation systems capable of accurately translating news and information across languages without exclusive reliance on human translators. This study contributes valuable data resources and establishes a foundation for further research, thereby supporting scholars working in Amharic machine translation. In addition, the study introduces an approach that has not been previously explored in Amharic machine translation, thereby advancing the existing body of knowledge. By leveraging prior linguistic knowledge, the proposed approach aims to improve machine translation performance. More broadly, language translation technologies facilitate communication across geographical boundaries and enable effective use of information and communication technologies. Consequently, translation agencies, including governmental institutions, increasingly support and adopt machine translation tools across a wide range of practical applications.

This study unfolds in three stages. The first stage involves a comprehensive literature review to understand machine translation using graph neural networks (GNNs), including an exploration of state-of-the-art machine translation and graph knowledge representation. The second stage outlines the processes of machine translation, specifically encoding and decoding. The third stage focuses on designing a machine translation model and evaluating its performance. We implemented several stages of the machine translation process using GNNs and identify prior knowledge to enhance the translator's effectiveness.

The significance of machine translation encompasses various dimensions. MT represents a technological milestone in the advancement of language and linguistics theories (Wilks, 2008). Over the decades, MT has influenced computational linguistics, knowledge representation, and information theory (Nirenburg et al., 2003). In academic settings, it supports bilingual or multilingual education while enhancing access to scholarly resources (G. H. Lin & Chien, 2009). Furthermore, MT addresses the growing global demand for translation driven by cross-regional communication, serving as a substitute for human translators when manual translation is not scalable (Chéragui, 2012).

In multilingual societies like Ethiopia, MT plays a crucial role in facilitating communication and improving governmental processes (Dwivedi & Sukhadeve, 2010). In the business sector, MT enhances translation speed, efficiency, and brand consistency, empowering global corporations to improve internal knowledge sharing and customer experiences (Sakre, 2019). Nonetheless, challenges such as overfitting training data and the inability to replace fully human translators remain key areas for development (Wibawa, 2019). Overall, MT serves as a valuable tool for overcoming language barriers and driving efficiency, innovation, and accessibility across diverse fields.

Machine translation plays a critical role in addressing the challenges associated with low-resource languages such as Amharic while expanding access to digital content. By enabling automatic translation, machine translation systems significantly improve access to information for millions of Amharic speakers. For example, systems such as Lesan demonstrate the potential of machine translation to democratize access to web-based content and promote digital inclusion among Amharic-speaking communities (Hadgu et al., 2022).

The development of linguistic resources, particularly Amharic–English parallel corpora, constitutes a critical factor in advancing Amharic machine translation. Such resources enable researchers to train more robust models and achieve measurable improvements in translation quality, as demonstrated by the Extended Parallel Corpus for Amharic–English Machine Translation (A. M. Gezmu et al., 2021). At the same time, Amharic's rich morphological structure and non-Latin script introduce substantial challenges for machine translation. Neural approaches that incorporate subword modeling and transliteration techniques have proven effective in addressing these challenges, leading to significant performance gains. For instance, studies on neural machine translation for Amharic–English have shown that subword-based models successfully capture Amharic's inflectional morphology and improve translation accuracy (A. M. Gezmu et al., 2021).

Recent advances in translation techniques have substantially improved English–Amharic machine translation. Approaches such as corpus augmentation and Transformer-based architectures have played a central role in achieving state-of-the-art performance for English–Amharic translation (Biadgligne et al., 2023). These results highlight the effectiveness of modern machine translation methods in addressing the challenges of low-resource languages. In addition, the normalization of Amharic homophonic characters has proven to be an effective strategy for improving translation quality in both translation directions. By reducing orthographic variation, normalization enables models to represent and process Amharic text more consistently, thereby enhancing overall translation performance (Destaw Belay et al., 2022).

Benchmarking studies emphasize the value of multilingual modeling in improving machine translation for Amharic. For instance, the Low Resource Neural Machine Translation study demonstrated significant gains by leveraging shared linguistic features across multiple languages, showcasing the potential of collaborative approaches to low-resource language translation (Lakew et al., 2020). Early work in statistical machine translation laid the foundation for current advances. Preliminary experiments on English-Amharic translation proved the feasibility of this approach, achieving baseline results that informed subsequent research. The study, M. G. Teshome & Besacier (2012a) highlighted the importance of techniques like morpheme segmentation in enhancing performance.

Finally, the evaluation of commercial machine translation systems and the development of datasets tailored for Amharic highlights the promising potential of this field. However, as the Hadgu et al. (2020a) study points out, Amharic MT systems still have room for improvement compared to high-resource language pairs, underscoring the need for continued research and innovation. In summary, machine translation for English-Amharic plays a crucial role in bridging language barriers, fostering digital inclusivity, and addressing linguistic complexities. The advancements in this field not only benefit the Amharic-speaking community but also contribute to the broader goal of improving machine translation technologies for low-resource languages worldwide.

1.1.3 Challenges of English-Amharic Machine Translation

Machine translation involves transforming text from one language to another while preserving meaning, syntactic structure, and semantic context, a task that remains inherently complex. One of the central challenges lies in modeling the relationships among words and phrases within a sentence. This difficulty intensifies when languages exhibit deep structural divergences and complex sentence constructions. Traditional approaches, including phrase-based statistical models and sequence-to-sequence (Seq2Seq) neural architectures (Figure 1.1), primarily process text as linear sequences and rely heavily on large-scale datasets or surface-level statistical patterns. As a result, these methods often fail to explicitly represent syntactic structure, which limits their ability to produce accurate and contextually coherent translations. This limitation is particularly evident in English-Amharic translation, where the two languages differ substantially in syntactic organization.

English follows a Subject-Verb-Object (SVO) word order, whereas Amharic exhibits a Subject-Object-Verb (SOV) structure. Although this distinction may appear minor, it significantly complicates translation, especially for syntactically complex and context-dependent sentences. One promising strategy to mitigate this challenge involves exploiting syntactic dependency structures of the source language, particularly English. Dependency trees explicitly encode hierarchical relationships and grammatical dependencies within a sentence, providing a structured representation that can guide more faithful translation. By incorporating such syntactic information into the translation process, models can better preserve sentence structure and meaning across typologically divergent languages.

Despite the potential benefits of syntax-aware modeling, existing research has not adequately explored the systematic integration and optimization of dependency tree structures within neural machine translation systems for English–Amharic translation. This gap limits the effectiveness of current neural machine translation approaches and motivates the development of new frameworks that embed syntactic dependency information directly into neural machine translation architectures. Such frameworks have the potential to improve word alignment, fluency, and translation accuracy in both translation directions, thereby advancing the overall quality of English-Amharic machine translation.

Beyond syntactic divergence, Amharic-English machine translation faces additional challenges rooted in the linguistic properties of Amharic and its status as a low-resource language. The limited availability of high-quality parallel corpora severely constrains data-driven approaches and reduces model performance. Sparse training data remains one of the most significant bottlenecks in developing effective English-Amharic translation systems. To address this issue, researchers have explored strategies such as corpus augmentation and optimization techniques, which have demonstrated measurable improvements in translation quality (Biadgligne & Smaïli, 2021). Nevertheless, the scarcity of large, diverse, and well-annotated datasets for Amharic continues to hinder progress (Hadgu et al., 2022).

In addition, Amharic language use of a non-Latin script and its rich morphological structure introduce further complexity. The language exhibits extensive inflection and derivation, which necessitate subword modeling and other open-vocabulary techniques to capture morphological variation effectively. While subword-based neural approaches have alleviated some of these challenges, they require further refinement and optimization to fully exploit Amharic’s linguistic characteristics (A. M. Gezmu et al., 2021). Moreover, the presence of homophonic characters introduces ambiguity at the orthographic level, complicating lexical disambiguation and negatively affecting translation quality (Destaw Belay et al., 2022).

The disparity between the state of machine translation for Amharic and that of well-resourced global languages further underscores existing research gaps. Although recent studies report incremental progress, English-Amharic machine translation systems still lag behind those developed for high-resource languages. This imbalance highlights the need for continued efforts in resource development, model innovation, and methodological adaptation. Furthermore, the limited generalizability of many Amharic-focused translation systems restricts their applicability across domains, reducing their broader impact.

Although multilingual and semi-supervised approaches have improved translation accuracy, their ability to generalize across domains and linguistic variations remains insufficiently explored. The high morpheme density characteristic of Amharic and other Ethiopian languages further affects translation precision, particularly when Amharic serves as the target language. Addressing these issues requires models that can effectively integrate morphological, syntactic, and semantic information.

In summary, English-Amharic machine translation faces multiple interconnected challenges, including limited resources, complex morphology, syntactic divergence, and insufficient modeling of linguistic structure. Overcoming these challenges demands innovative solutions that incorporate syntactic dependency structures, semantic representations, and efficient data utilization strategies. Such advancements are essential for developing robust, accurate, and scalable machine translation systems capable of handling the linguistic intricacies of Amharic.

1.1.4 Benefit of Prior Knowledge and GNNs

Incorporating prior linguistic knowledge and Graph Neural Networks (GNNs) offers substantial advantages for machine translation, particularly in low-resource and domain-specific settings. By embedding syntactic and semantic structures into learning frameworks, these approaches reduce performance gaps in resource-scarce scenarios and provide structured representations that facilitate more effective and robust model learning.

Recent studies demonstrate the effectiveness of GNN-based methods in leveraging cross-lingual knowledge. For example, IBRAHIM et al. (2024) employ GNN models that integrate cross-lingual word embedding fusion with knowledge distillation to enhance multilingual performance. Their approach enables semantic alignment across languages and supports tasks such as social event detection in under-resourced languages, even in the absence of large annotated datasets.

Similarly, Annervaz et al. (2018) show that neural models augmented with knowledge graphs can access structured prior information during inference, thereby improving performance across a range of natural language processing tasks. This integration reduces reliance on extensive labeled data and increases model flexibility in multitask learning environments.

Beyond supervised settings, GNN-based techniques also facilitate knowledge transfer in unsupervised and weakly supervised scenarios. Imani et al. (2022) demonstrate that combining multilingual embeddings with graph-based label propagation enables effective part-of-speech tagging for low-resource languages without direct supervision. By exploiting relational structures within graphs, these methods transfer linguistic knowledge from high-resource to low-resource languages more efficiently.

Overall, integrating prior knowledge into GNN-enhanced neural architectures significantly strengthens a model's capacity to acquire, generalize, and adapt linguistic information. This capability is particularly valuable for machine translation tasks involving complex linguistic phenomena and unexplored domains, where traditional data-driven approaches struggle. Consequently, prior knowledge–aware GNN frameworks offer a promising direction for addressing the persistent challenges of low-resource machine translation.

1.2Motivation of the Study

Ethiopia's multilingual and multicultural society demands accurate and reliable translation systems to support effective communication at both national and international levels. English and Amharic play central roles in the country's education, commerce, governance, and international relations, where the timely and accurate exchange of information is essential. Despite this need, existing machine translation (MT) systems struggle to handle the linguistic intricacies and nuanced structures of these two languages, particularly Amharic. The limited linguistic awareness of current MT models often leads to inaccurate translations, which can hinder communication and the dissemination of information.

Recent advances in machine translation have significantly improved performance for well-resourced languages; however, substantial gaps remain for language pairs with pronounced structural divergence. English and Amharic differ markedly in their morphosyntactic organization, making accurate translation especially challenging. Current MT systems frequently fail to capture these differences, resulting in reduced fluency and semantic fidelity. These limitations highlight the need for translation models that move beyond surface-level patterns and incorporate deeper linguistic understanding.

This study is motivated by the premise that integrating prior linguistic knowledge specifically, the syntactic structure of the source language, can substantially enhance English-Amharic machine translation. By embedding syntactic dependency information into the translation process, the proposed approach aims to improve the model's ability to capture long-distance dependencies, preserve grammatical relations, and generate more accurate and fluent translations. Such integration enables the development of an MT system with a more sophisticated understanding of source-language syntax than that offered by conventional approaches.

In high-resource languages, advanced MT systems already facilitate seamless communication across linguistic boundaries. In contrast, Ethiopia’s low-resource language context continues to face significant challenges in achieving comparable translation quality. Reliable English–Amharic translation remains critical for improving information accessibility, supporting economic development, and fostering cultural exchange. However, current systems do not yet meet the required standards of accuracy and fluency for these purposes.

This research therefore seeks to address these challenges by leveraging prior knowledge, particularly syntactic dependency trees, within an English-Amharic MT framework. By integrating structured syntactic information into modern neural architectures, the study aims to enhance translation performance and reliability. Ultimately, this work aspires to contribute to the development of more inclusive, accurate, and effective machine translation systems that better serve individuals, institutions, and communities in Ethiopia, thereby supporting the country’s broader goals of informed participation and global engagement.

1.3Statement of the Problem

Despite substantial progress in Neural Machine Translation (NMT), current models often fail to effectively exploit pre-existing linguistic and translational knowledge. Unlike earlier Statistical Machine Translation (SMT) systems, which explicitly modeled word alignments and translation probabilities, NMT relies primarily on dense vector representations that implicitly encode linguistic information. As a result, much of the structured knowledge embedded in SMT models has not been systematically preserved or utilized within modern NMT frameworks (Wang et al., 2019).

Several studies have attempted to integrate prior knowledge into NMT models. For instance, Li et al. (2018). proposed a posterior regularization framework that incorporates translation constraints into NMT and reported performance improvements on Chinese-English translation tasks. However, the study did not clearly specify the source or structure of the prior knowledge and introduced additional computational overhead during training. Similarly, Chen et al. (2021) demonstrated improved translation fidelity by injecting prior translation knowledge to mitigate fluent but below expected outputs. Nevertheless, the origin, representation, and domain relevance of the incorporated knowledge remained insufficiently defined.

Other approaches have explored the transfer of knowledge from SMT to NMT. Wang et al. (2019), for example, integrated SMT-derived lexical knowledge into NMT by modifying word generation probabilities during both training and inference. This method improved the translation of rare words and increased source coverage. However, these efforts have largely focused on high-resource language pairs such as Chinese-English and English-German, leaving low-resource and morphologically rich languages, including English-Amharic, largely unexplored.

Furthermore, standard sequence-based NMT architectures struggle to model structured linguistic information, particularly syntactic dependency relationships, which naturally exhibit graph-like structures. Sequence-to-sequence models lack the representational capacity to directly encode such relational structures (Bastings et al., 2017). This limitation necessitates the adoption of alternative modeling paradigms capable of accommodating graph-structured data, such as Graph Neural Networks (GNNs).

Consequently, this study addresses the fragmented and limited exploration of prior knowledge integration within Graph-based Neural Machine Translation (GNMT), particularly for low-resource language pairs such as English–Amharic. Existing research either neglects explicit incorporation of syntactic and lexical prior knowledge or restricts integration into isolated components of the translation pipeline. This gap underscores the need for a comprehensive, linguistically motivated framework that systematically embeds structured prior knowledge, especially syntactic dependency information, into GNMT architectures. Such an approach has the potential to significantly enhance translation quality for morphologically complex and under-resourced languages.

1.4Research Questions

- RQ1: What are the current state-of-the-art machine translation models for English-Amharic language pairs?
- RQ2: What prior knowledge will improve the performance of the translator?
- RQ3: How can prior knowledge be effectively integrated into the design of a graph neural network machine translation model?

- RQ4: What is the impact of incorporating prior knowledge on the performance of the proposed machine translation model?
- RQ5: Comparative analysis of results with the SOTA?

Hypothesis: The integration of prior knowledge into attention-based machine translation model using GNN, will result in an improvement in the performance of English-to-Amharic text translation compared to existing machine translation NMT models.

1.5Objective of the Study

1.5.1 General Objective

The primary objective of this study is to investigate a graph neural network-based machine translation model for English-Amharic text translation. By incorporating prior knowledge, such as the syntactical structure of the source language, the study aims to enhance the performance of the transformer model.

1.5.2 Specific Objective

To meet the general objective of this study, the following specific objectives are established.

- To analyze existing machine translation models for Amharic and English language pairs.
- To collect parallel corpus of English-Amharic dataset.
- To identify the prior knowledge that improves the performance of a translator.
- To design the knowledge graph representation (syntax dependency tree of source language)
- of the prior knowledge of graph neural network machine translation (GNMT).
- To evaluate the performance of the proposed model on a parallel corpus of English-Amharic text.
- To compare the results of the proposed model with state-of-the-art machine translation models (NMT).
- To demonstrate the effectiveness of integrating prior knowledge in improving the accuracy of the machine translation model.

1.6Scope and Limitation

1.6.1 Scope of the Study

The current investigation sits at the crossroads of computational linguistics, natural language processing (NLP), and low-resource machine translation (MT), concentrating on enhancing

English-Amharic translation using knowledge incorporation techniques. This research tackles the numerous problems associated with neural machine translation (NMT) systems for morphologically complex, sparsely resourced, and highly structurally complex languages, with a focus on Amharic and English languages. In terms of practical implementation, the scope of the study attempts to address both the theoretical and practical aspects of machine translation with the intent to:

- **Model Prior Knowledge Integration:** This study investigates the problem of embedding clear-cut formal and syntactic knowledge into GNMT systems, which involves capturing source-side syntactic representations such as dependency trees (part-of-speech tagging) fragments of speech, and words order for steering translation from English into Amharic to assist with grammar-based structural alignment of subordinate clauses. Integration is achieved using pre-processing, structural alterations (like the addition of syntax-centric encoders).
- **Address Linguistic Divergence:** The understanding of the structural divergences of English (an analytic, SVO language) and Amharic (a morphologically rich SOV language) was used to explain why translation errors, occur in standard MT systems. The goal of studying is to bridge the gap syntactically and morphologically, which will enhance the output fidelity and semantic satisfaction in translations.
- **Enhance Performance in Low-Resource Settings:** In the case of English to Amharic translation, there are few high-quality parallel corpora, which the researchers sought to address through prior knowledge. This entails handling the effects of integrating knowledge on translation performance with limited training data, as well as the application of data augmentation approaches (like synthetic data and back-translation) used together with prior knowledge.
- **Experimental Design and Evaluation:** The research proposes design and assessment of baseline and enhanced NMT models based on the available English–Amharic parallel corpora. Evaluation of the models is quantitative and qualitative. Quantitative evaluation is incorporate standard MT measures like BLEU, and METEOR.
- **Toolkits and Frameworks:** We implement and test extensions of linguistic features using popular NMT systems like OpenNMT, Graph2seq, and M2M100 pretrained model. To

retrieve knowledge from the English source texts, a linguistic annotation tool and parser like StanfordNLP is employed.

- **Language-Specific Considerations:** The focus of the analysis is English-Amharic languages pair, but it aims to add insights applicable to languages that are morphologically rich but under-researched. The approaches developed could aid prior knowledge integration in MT systems designed for a wider scope of neglected languages.

This study aims to foster discussion on the contributions of linguistically directed advances in deep learning technologies for NLP, broadening the singular focus towards English-Amharic MT at the implementation level. With this aim, the study seeks to critically address under-resourced languages and enable equitable approaches in language technology alongside deep-level inclusive approaches.

1.6.2 Limitations of the Study

While this study aims to contribute to the advancement of English-Amharic graph neural machine translation through the integration of prior linguistic knowledge, the following limitations are acknowledged:

- **Limited Parallel Data:** The study is constrained by the scarcity and limited quality of English-Amharic parallel corpora. In addition, data augmentation techniques such as back translation and synthetic data generation are not employed. The lack of large-scale, domain diverse, high-quality parallel data remains a significant challenge to model robustness and generalizability.
- **Focus on Syntactic and Morphological Knowledge:** The research specifically targets syntactic of the source language as a prior knowledge. Other forms of prior knowledge, such as semantic role labeling, sentence structure, or pragmatic context, are outside the scope of this study.
- **Language-Specific Generalizability:** While the findings aim to be informative for other morphologically rich, low-resource languages, the methods and results are empirically evaluated in the English-Amharic language pair. Therefore, direct generalizability to other language pairs with different structural characteristics may require additional adaptation and validation.

- **Computational Constraints:** Due to resource limitations, the scope of experimentation is restricted in terms of model size, training duration, and hyperparameter optimization. Larger-scale models and longer training cycles could potentially yield higher translation performance but fall beyond the feasible computational capacity for this study.
- **Evaluation Metrics:** The study primarily employs automatic evaluation metrics such as BLEU and METEOR. Although these are widely used in MT research, they capture nuances of meaning, fluency, or adequacy, particularly for morphologically complex target languages like Amharic. Human evaluation is employed for a small subset of the test data to see the effectiveness of machine translation model using qualitative.
- **Tooling and Parser Availability:** The extraction of linguistic features relies heavily on existing syntactic and morphological annotation tools, specifically StanfordNLP. It is optimized high-resource languages like English. We only consider the source language of machine translation.

These limitations do not diminish the contribution of the study but rather outline areas for further exploration and development. Future research is encouraged to address these constraints by expanding the linguistic scope, increasing computational experimentation, and improving resource availability for underrepresented languages like Amharic.

1.7Significance of Study

This research holds significant importance for the fields of computational linguistics, natural language processing (NLP), and machine translation (MT), particularly in the context of under-resourced languages such as Amharic. The study contributes technical, linguistic, social, and practical value by addressing longstanding challenges in English-Amharic machine translation through the integration of prior linguistic knowledge and advanced neural architectures.

At the linguistic level, the study directly confronts the structural and morphological disparities between English and Amharic, which constitute one of the primary obstacles to effective translation. Amharic, a Semitic language, exhibits rich morphological characteristics, including root-and-pattern morphology, extensive verbal inflection, and a subject-object-verb (SOV) word order, in contrast to English's subject-verb-object (SVO) structure. These fundamental differences complicate direct translation and frequently lead to syntactic misalignment and grammatical incoherence in existing MT systems. Many contemporary models circumvent these challenges due to data sparsity and insufficient integration of syntactic and morphological knowledge. By explicitly incorporating syntactic prior knowledge, this study enables the translation model to capture grammatical dependencies and structural relations more effectively, thereby improving translation accuracy and fluency for English-Amharic language pairs.

From a methodological perspective, the study advances machine translation techniques by integrating Graph Neural Networks (GNNs) with syntactic dependency representations. While conventional neural machine translation models have demonstrated strong performance, they remain limited in their ability to encode syntactic relationships and long-distance dependencies due to their sequential nature. The proposed GNN-based framework addresses these limitations by modeling linguistic structures as graphs, allowing the system to represent non-linear dependencies and complex contextual interactions that are often overlooked by sequence-based models. This approach bridges theoretical insights from linguistics with recent advances in deep learning, offering a principled and interpretable framework for syntactically informed translation.

The study also makes a substantial contribution to low-resource language processing. Amharic, like many under-resourced languages, suffers from a lack of large-scale parallel corpora required for training high-performing NMT systems. This research demonstrates that integrating structured linguistic prior knowledge can partially compensate for data scarcity and lead to meaningful improvements in translation quality. The findings are therefore applicable beyond Amharic, providing a transferable framework for other low-resource and morphologically complex languages facing similar constraints. By improving translation accessibility and usability, the study promotes digital inclusion and enhances information dissemination within low-resource linguistic communities.

In addition, this research establishes a solid foundation for future investigations into hybrid NLP models that combine linguistic knowledge with neural architecture. The proposed framework and empirical findings support further exploration of syntactic and semantic integration in machine translation and underscore the importance of developing richer and more balanced datasets for under-resourced languages. Moreover, the study opens avenues for extending similar methodologies to other languages with complex grammatical systems, thereby advancing research in multilingual and cross-lingual NLP.

Beyond its theoretical contributions, the study offers significant practical implications. An improved English-Amharic machine translation system can support a wide range of real-world applications. In education, it facilitates the translation of instructional materials and academic resources, enhancing knowledge dissemination. In government and public services, it enables accurate translation of official documents, public communications, and digital platforms, ensuring equitable access to information. In media and communication, it improves content localization and cross-linguistic engagement. Collectively, these applications strengthen cross-cultural understanding, information access, and social cohesion in Ethiopia’s multilingual context.

Finally, this study contributes to the broader theoretical and methodological discourse in computational linguistics by challenging the assumption that increased data alone suffices for high-quality translation. The findings demonstrate that informed system design and the deliberate integration of linguistic structures remain essential for developing robust and context-sensitive MT systems. By examining the interaction between linguistic knowledge and neural representations, this research provides valuable insights into how structured prior knowledge can enhance model performance and interpretability.

Overall, this study is significant both theoretically and practically. It advances English-Amharic machine translation through the integration of syntactic prior knowledge and graph-based neural architectures, addresses critical challenges faced by low-resource languages, and contributes to the development of linguistically grounded and socially impactful machine translation systems.

1.8Organization of the Dissertation

The organization of this dissertation is structured into six main parts. Chapter II focuses on literature review and related work, covering foundational concepts of machine translation, including Statistical Machine Translation (SMT), Neural Machine Translation (NMT), and state-of-the-art techniques, as well as related work and evaluation metrics. It also examines the integration of prior knowledge in MT, discussing its types, methods, and associated challenges, alongside an exploration of Graph Neural Networks (GNNs) and their applications in NLP. Chapter III presents the research methodology, detailing the dataset and preprocessing steps for the English-Amharic parallel corpus, followed by the proposed model architecture featuring GNN-based encoder and decoder components and a prior knowledge integration module. Chapter IV is dedicated to experiments setup, which describes the experimental setup of baseline models and proposed one. Chapter V presents results and discussion which analyzes quantitative results, including baseline models. Finally, Chapter VI concludes with a summary of contributions, limitations, and implications, and proposes directions for future research and potential applications of the developed techniques.

Chapter Two: Literature Review and Related Works

2.1Machine Translation

Machine translation (MT), defined as the automated conversion of text from one human language to another using computational methods, has been a central research focus in natural language processing for several decades. Recent years have witnessed substantial improvements in translation quality and efficiency, driven primarily by advances in neural machine translation (NMT). Despite these developments, current MT systems remain far from achieving human-level translation performance, particularly in capturing deep linguistic structure, semantic nuance, and contextual meaning.

Early and contemporary research has consistently highlighted the importance of incorporating prior linguistic knowledge into translation systems. Several studies argue that explicit knowledge of syntactic and structural properties of languages can substantially enhance NMT performance (Chen et al., 2021; Yang et al., 2017). However, most modern NMT systems continue to rely heavily on sequential encoder–decoder architectures, which model language as a linear sequence of tokens. While effective for many tasks, these architectures inadequately represent hierarchical linguistic phenomena such as syntax and long-distance dependencies (Bahdanau et al., 2014; Bastings et al., 2017; Sutskever et al., 2014).

The limited integration of syntactic information in NMT largely stems from the difficulty of encoding structured linguistic knowledge within traditional neural architectures, particularly recurrent neural networks (RNNs). Structured representations, such as parse trees and dependency graphs, do not naturally align with the sequential processing mechanisms of RNN-based encoders, leading many systems to rely instead on latent representations that implicitly encode linguistic features without explicit structural modeling (Bastings et al., 2017).

Long-standing concerns regarding the gap between machine and human translation quality further motivate the integration of linguistic knowledge into MT systems. Simard & Isabelle (2009) emphasize that achieving human-like translation remains a persistent challenge, as machines continue to struggle with deep semantic interpretation and contextual understanding. In response, knowledge-based approaches to MT have been proposed as a means of improving translation robustness and interpretability (Mahesh & Nirenburg, 1996; Nirenburg et al., 2003). Nirenburg et al. (2003) advocate the incorporation of domain-specific and linguistic knowledge into translation systems to enhance accuracy and reliability, while Mahesh & Nirenburg (1996) stress the role of linguistic theory, world knowledge, and semantic representation in capturing meaning beyond surface-level text processing. Collectively, these works underscore the critical role of prior knowledge in advancing MT systems toward more human-like performance.

Incorporating syntactic information from the source language into neural encoders offers a principled approach to improving both comprehension and generation in MT systems. Explicit modeling of syntactic relationships enables the system to better preserve grammatical structure and semantic roles during translation, thereby reducing meaning loss and structural distortion. This approach not only enhances translation quality but also supports the broader objective of aligning machine translation behavior with human linguistic reasoning.

Attention-based NMT models, introduced by Bahdanau et al. (2014) and further developed by Luong et al. (2015) and (Bastings et al., 2017), represent source sentences as latent feature vectors produced by the encoder and selectively attended to during decoding. While these representations effectively capture contextual information, they do not explicitly encode syntactic relations between words. To address this limitation, this study aims to integrate syntactic dependency information directly into the attention-based graph neural machine translation (GNMT) encoder.

Dependency syntax trees provide a natural and linguistically motivated structure for modeling grammatical relationships between words in a sentence. These trees explicitly represent head–dependent relations, such as subject-predicate and object-verb associations, thereby capturing the hierarchical organization of language. As illustrated in Figure 2.1, the word “I” functions as the subject of the verb “booked,” while “ticket” serves as its object. By leveraging dependency trees within a graph-based neural framework, the proposed model encodes syntactic relationships alongside semantic representations, enabling more informed and structurally faithful translation. Through this integration, the study seeks to overcome the limitations of purely sequential models and advance machine translation toward more linguistically grounded and accurate systems.

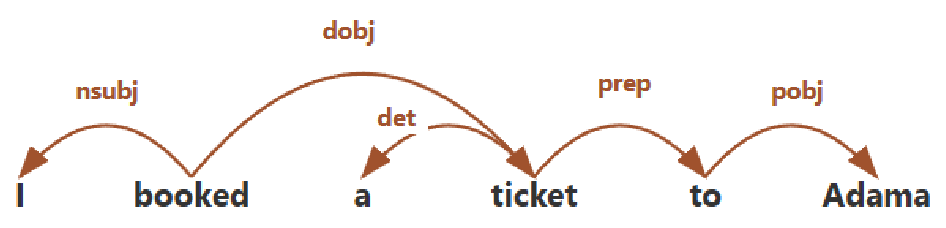


Figure 2.1: A syntax dependency tree for the example sentence "I booked a ticket to Adama."

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 issues of preserving meaning, managing long-range dependencies, resolving ambiguities, and increasingly fluent translations through the context by incorporating more syntax information into their computations.

Figure 2.1 shows a syntax dependency tree for the example sentence I booked a ticket to Adama, representing all its constituents and their grammatical relations. From the principal verb "booked", which is the head of the syntax tree, the dependents are also connected. The subject "I" is connected to the verb with the nsubj relation, and the direct object "ticket" is connected with a relation dobj ("verb object") "a". In addition, it contains a det (determiner) relation. There is a prepositional phrase, "to Adama" that also stems from the verb. For this phrase, "to" functions as a verb and head and is marked with prep and "Adama" as the pobj. Both constituents are marking what the action is directed to. The arrows illustrate these components depict the grammatical relations of not only sentence constituents but also words.

We create syntax-aware features for words using graph neural networks (GNNs). This utilizes the relationships defined within syntactic graphs as the building blocks to better understand the dependencies and other relationships between words, therefore capturing more meaningful features (Bastings et al., 2017). The utilization of GNNs allows the capturing of complex syntactic features of the given language that improves the performance of a machine translation model.

Different studies show the improvement in the performance of Natural Language Processing (NLP) systems, especially in machine translation, because of incorporating the source language's syntax into the framework of the Graph Neural Network (Bahdanau et al., 2014; Bastings et al., 2017; Sutskever et al., 2014). This study works together to point out the dominant impact that adding source language syntax as prior information into GNNs could have on the performance of these systems in machine translation tasks.

The inquiry aims at developing a model of GNNs for machine translation from English to Amharic that embeds prior knowledge in the form of the source language of syntactic structure. The objective of incorporating syntactic structures into the translation process is to improve the precision and fluency of the text, thereby improving the system performance on English to Amharic translation. Our starting point was the study by Destaw Belay et al. (2022), who employed a transformer model and then improved it by fine-tuning it with the pre-trained M2M100 48M model. This model will serve as the baseline translation system, and thus, we can measure the performance of our model against this benchmark. Using their methodology, we can evaluate the effectiveness of the proposed improvements in an organized manner.

2.2Evolution of Machine Translation

Earlier technologies improved translation with new technology, but the evolution of machine translation has greatly enhanced the accessibility of translation technology. Each leap forward, from rule-based systems to statistical and neural models, has brought us closer to achieving human-like translation capabilities. While virtually all languages have been automated, there is room for improvement in lesser-known languages, precision in specialized fields, and infallible accuracy. There is ongoing development to enhance the effectiveness and boundaries of machine translation. The use of machine translation to solve the issue of evolving multilingual translation systems has the potential to break boundaries of freedom, encourage understanding, and enhance communication around the globe. Neural Machine Translation (NMT) began a new era in the world of translating. Around 2014 NMT was introduced (Ragni & Nunes Vieira, 2022).

NMT has received a lot of attention due to its promise of more fluent, accurate, and contextually aware translations compared to older, traditional techniques. NMT models built by Google Brain went a step further with deep learning techniques and attention mechanisms. Named the sequence-to-sequence (seq2seq) model, these techniques process entire sentences as a single unit. These models are far more capable than their predecessors since they capture subtler dependencies and meaning from further back in the text (Sutskever et al., 2014). The emergence of neural machine translation has advanced the sophistication of machine translation, making it critical in fields like academia, global business communications, browsing services, or language learning.

Like many modern AI systems, NMT has specific data and resource requirements. To achieve optimal results, it must be trained using parallel text data and computational power. However, after training, generalization becomes considerably easier to accomplish as the system can tackle fresh content and produce more human-like translations. NMT has also benefited from implementing transformer architectures, such as OpenAI's GPT (Radford et al., 2018) and Google's BERT (Gu et al., 2018). These systems have advanced the capability of context-aware translations significantly.

The much-anticipated introduction of machine translation (MT) has drastically improved the AI and computational linguistics landscape. The gradual shift from rule-based frameworks to neural networks has significantly enhanced the ease of communication across different languages. With the pace of growing technology, it embodies how much the world needs MT; it holds key potential to unite people globally. It must be clear that the journey has only just begun, and we can only anticipate what possibilities unfold in the future.

2.2.1 Rule-based Machine Translation

RBMT systems are based on explicit linguistic knowledge and rules derived from linguistic analysis. These systems typically rely on the following components:

Table 2.1: Components and Descriptions of RBMT

Component	**Description**
Lexicon/Dictionary	Stores vocabulary, including morphological, syntactic, and semantic information.
Grammar	Rules Encodes the syntax and morphology of both source and target languages.
Transfer	Rules Define how syntactic/semantic structures in the source language are mapped to the target
Parser/Analyzer	Analyzes input text to generate intermediate linguistic representations.

Generator	Converts intermediate representation into target language output.

Recent advancements in machine translation (MT) have increasingly underscored the value of integrating rule-based and statistical methods, particularly in low-resource language scenarios. Torregrosa et al. (2020) investigated the potential of enhancing neural machine translation systems with linguistic knowledge derived from rule-based models, specifically morphological features, named entities, and specialized terminology. Although the overall quantitative improvements were limited, qualitative evaluations revealed meaningful enhancements in certain outputs, such as the retention of passive voice constructions. Similarly, Terumasa (2007) demonstrated that combining rule-based MT with statistical post-editing improved the accuracy of Japanese-to-English patent translations, though the resulting text sometimes contained unnatural expressions. Expanding on this hybrid approach, Simard et al. (2007) showed that applying statistical phrase-based post-editing to rule-based outputs led to significant increases in translation quality, as indicated by BLEU scores. In a comparative analysis, (Costa-Jussà et al., 2012) found that rule-based systems outperformed statistical models in orthographic and morphological accuracy, while statistical approaches were more effective at minimizing semantic errors. Together, these studies illustrate the complementary strengths of rule-based and statistical methods and support their integration in hybrid machine translation architectures, especially in linguistically complex or under-resourced contexts.

2.2.2 Statistical Machine Translation

Statistical Machine Translation (SMT) is where natural language processing begins, and it deals with the task of translation as a probabilistic learning problem within the context of an aligned bilingual corpus. Lopez (2008) offers an introductory-style summary of SMT, dividing it into three main components: translation modeling, parameter estimation, and decoding. He also suggests a taxonomy of different approaches to SMT. In a subsequent and more cited work, Lopez (2008) explains that SMT treats translation as a learning problem, which is refined over successive iterations using human-provided translations. They describe the evolution of rule-based (RBMT) and example-based (EBMT) methodologies to statistical and neural methods, highlighting the modular nature of SMT system architecture and their flexibility.

Hassan & Darwish (2014) likewise define SMT as an approach that operates on the assumption that knowledge is derived from bilingual corpora without providing specific results or methods since. Like them, Specia (2010) surveys both classical and contemporary methods of SMT but does not contribute original research results or methods. In his foundational work, Koehn et al. (2007) analyzes the reasoning behind the phrase-based SMT which was the most used technique at the peak of research on SMT.

Using a parallel corpus of children's stories, they executed sentence-level alignments and employed SRILM for language modeling, GIZA++ for word alignment, and Moses for decoding. Their approach demonstrates a full SMT pipeline and highlights the possibilities of applying SMT to lower-resourced language pairs. A similar application was demonstrated by Singh et al. (2014) who tackled English to Assamese translation. In this case, a phrase-based SMT model was combined with a small corpus of five thousand sentence pairs, which led to reasonable outcomes, although the authors recognized the inadequate size of the data.

Collins (2011) describes IBM Models 1 and 2, which serve as the initial foundation for the mathematical structure of SMT systems. These models begin with a step that utilizes large parallel corpora like Hansards and Europarl to determine word alignment and translation probability of the words within the defined corpus. Brown et al. (1993) propose new methods for estimating parameters in SMT, suggesting five statistical models of translation and alignment. They furnish algorithms that calculate the most likely word-to-word alignments but caution that such algorithms do not guarantee optimal results.

Regarding SMT progress from 1996 to 2005, Popovic & Ney (2006) praised the development of model combination techniques, the advancement of training algorithms, and the availability of computing resources. He points out the complex language pairs SMT was able to competitively perform on and how it achieved industry standards. In a similar fashion, Koehn & Knowles (2017) report results from the WMT-07 shared task, which assessed European language SMT systems and tackled issues like domain adaptation. These workshops were milestones marking the development in the area.

Other than the suggested uses of SMT, Riezler et al. (2007) applied SMT methods for query expansion in answer retrieval, exploring a new area of research. They demonstrate how a full-sentence SMT model trained on question-answer pairs can improve retrieval performance, showcasing the power of SMT beyond traditional translation. Och & Ney (2002) also made an important improvement by introducing discriminative training with maximum entropy models. This addition changed the scope of SMT systems to include more feature functions such as applying word penalties and class-based language models, enhancing the translation quality.

Research developed alongside the progress of SMT. As systems grow in complexity, researchers become more concerned with their limitations and prospective transitions. El Maazouzi et al. (2017) offer a systematic account of the progression from SMT to neural machine translation (NMT). They describe the major shift to encoder-decoder frameworks with the development of attention technologies, which fundamentally changed the focus of research in machine translation. Also, Costa-Jussà et al. (2012) points out the advantages and likely shortcomings of phrase-based SMT, referring to the increase in translation quality but still presenting concerns with reordering and domain adaptation.

Hearne & Way (2011) form interdisciplinary relationships between linguistics and SMT by engaging with issues of how linguistic information can be used to build statistical models. Though lacking in empirical data, the paper adds to the discourse of SMT from a linguist's angle. All these works combined offer a journey depicting the evolution of SMT: from the theories and models to practical systems, only to later be supplanted by newer, more advanced, neural approaches. It showcases the evolution, strengths, and greatest weaknesses of SMT being the primary translation method used before the neural era.

2.2.3 Neural Machine Translation

The adoption of neural machine translation (NMT) systems with transformer models has been demonstrated to surpass the performance of traditional statistical machine translation (SMT) techniques for numerous languages and across evaluation metrics. Nguyen et al. (2021), the authors indicated that the performance of transformer models could be improved by preceding alignments that were either hand-labeled or automatically produced. They proposed a hybrid transformer model that incorporated word-to-lemma alignments and achieved an improvement in BLEU scores of 2.53 points from the baseline and 0.79 BLEU points over models trained with manual alignments. The follow-up work conducted by Nguyen et al. (2021) validated the claims with heavyweight prior alignments and attention averaging techniques, which resulted in a 4.37 BLEU point influx on English-Vietnamese translation.

The results of comparative assessments reinforce the multi-lingual dominance of Transformers. Kumar et al. (2024) reported that transformer-based multilingual NMT models outperformed their counterparts that utilized convolution and attention-based seq2seq architectures in translating German, French, and Czech to English, with the highest recorded BLEU score reaching 48.47. Additionally, Saengthongpattana et al. (2019) contested the RNN and SMT models in Thai-English and English-Thai translations and verified the supremacy of the transformer model. The analysis noted that while all models struggled with under or over translation, the transformer models made significantly fewer translation and word order errors.

Several surveys, alongside meta-studies, have provided an overview of the trends. Hu et al. (2024) gave a thorough overview of the NMT evolution, remarking on the fact that the Transformer captured long-range dependencies as well as their dynamic optimization techniques, such as learning rate and dropout scheduling, make them better and more efficient than SMT systems. A review by Ganesh et al. (2023) had already documented the shift from rule-based and statistical paradigms with deep learning in low-resource and cross-linguistically challenging settings. El Maazouzi et al. (2017) has also shared this insight and called attention to the increasing shift towards the use of attention-based NMT architectures due to their superior alignment and contextualization features.

Jassem & Dwojak (2019) analyzed a specialized English-Polish corpus and concluded that NMT yielded more fluent translations, while SMT performed better according to BLEU scores, thus demonstrating the tension between the criteria of fluency and adequacy. Focusing on NMT-specific strengths, some analyses have been reported. Bentivogli et al. (2018) reported that NMT systems, in comparison to phrase-based SMT systems, decreased the effort needed for post-editing by 26% and performed significantly better in error morphology, lexicon, and word order. They did report some progress in verb reordering, but the translation of proper nouns was still problematic. Regarding cross-architecture comparisons, Smirnov et al. (2022) demonstrated that of the multiple Transformer-based NMT models analyzed, M2M100 had the highest overall translation quality; while also noting it had the highest need for fine-tuning.

The combination of SMT and NMT techniques has also been researched. Akter et al. (2018) surveyed studies that attempted to merge statistical and neural models for enhanced translation quality. Such systems have promising potential, but in most cases pure Transformer-based NMT models still demonstrate superior performance.

To sum up, the research analyzed clearly agrees that transformer-based neural machine translation outperforms the older statistical methods in every aspect regarding quality of translation, fluency, and complexity of described linguistic features. While some studies accept that SMT could be at par in certain domains or with specific language pairs, the evolution of model design, training processes, and the embedding of prior information have made Transformers the leaders in today's machine translation. The ongoing development of these models, particularly the alignment and multilingual feature additions, points to more advancements in low-resourced languages or rich morphology languages.

2.3Evaluation Metrics for Machine Translation

An improvement in machine translation (MT) systems is always accompanied by great changes to their evaluation metrics, which determine the system's adequacy. One of the most prevalent metrics to evaluate the quality of translations is BLEU, and others are METEOR and TER. Bilingual Evaluation Understudy (BLEU) is perhaps the most frequently used metric among practitioners because it is straightforward and quick to use. Its scoring is based on n-grams, which measure how much a given translation aligns with a reference/baseline translation while also applying a brevity penalty for exceeding short translations. Unfortunately, it does not consider semantic inclusiveness, giving more weight to precision, which is severely strained in morphologically convoluted languages (Gupta et al., 2010).

As stated by Banerjee & Lavie (2005), Evaluation of Translation with Explicit Ordering (METEOR) attempts to resolve some of the BLEU shortcomings by adding more sophisticated linguistic elements like stemming, synonymous words, and loose word matching. Unlike BLEU, METEOR adjusts the balance between precision and recall, which correlates more with human evaluations, especially for translations with complex syntax or high diversity of languages. Lavie & Denkowski (2009) claimed in 2009 that METEOR behaved as syntactically sophisticated translations. Additionally, METEOR imposes penalties on the order of words, which makes it a stricter metric. This flexibility, however, increases the computational demand compared to BLEU, making it less efficient (Denkowski & Lavie, 2011).

BLEU is best suited for benchmarking and comparing systems when translation references are abundant or in situations where the grammar of the target language pair is simpler (Papineni et al., 2002). On the contrary, METEOR is the more favorable choice for educational exercises, particularly when reasoning about translations of more complex morphology or syntax languages due to the reliance on human judgment in these contexts (Banerjee & Lavie, 2005; Gupta et al., 2010). As demonstrated in m-BLEU and m-TER, both metrics have been augmented to incorporate additional features to measure linguistic diversity and align more closely with human evaluations (Agarwal & Lavie, 2008).

2.4Types of Prior Knowledge

The integration of existing knowledge into various Machine Translation (MT) systems has been an area of research focus over the years. In a way, this has been done with posterior regularization, which adds multiple sources of knowledge by turning them into features harnessed within a loglinear model on neural machine translation (NMT) (T. Zhang et al., 2020). Another approach is to embed prior translation knowledge within the source side of the training pipeline which is said to improve NMT performance (M. Chen, 2024). Both approaches demonstrate the impact of structure prior knowledge on the translation output quality.

Prior knowledge focused too much on disintegrating syntactic and semantic knowledge sources, which permits the application of language-specific grammar alongside a domain-dependent semantic knowledge base. For instance, Tomita & Carbonell (1986) designed a framework where syntax and semantics are resources distinctly integrated for a linguistically motivated, context-sensitive approach to understanding. In the same way, Mahesh & Nirenburg (1996) researched the design of meaning representation that links language-dependent lexicons to a language-neutral world model, enabling extensive multilingual translation. Meaning representation is defined here as the process of augmenting a foundational semantic structure with lexical, ontological, and textual attributes to produce meaning representation.

Such representations enable multilingual MT systems to more effectively perform language and domain-specific tasks as knowledge bridges the gap between languages and world knowledge sources. Researchers continue to leverage unlocked approaches to MT and challenge its accuracy and adaptability in extreme linguistic and contextual settings. The integration of syntactic prior knowledge into machine translation (MT) has been done before, and this prior work illustrates the knowledge possessed by the machine will notably improve translation quality.

Posterior regularization provides a broad approach to add more elements, such as syntax, to other knowledge types in NMT. It does so by placing poses using syntax structures as features in a loglinear model, thus guaranteeing that the translation considers the syntax of the source languages and does not ignore its intricacies. This offered handling for complex tracing of linguistic variations of (J. Zhang & Zong, 2020). Embedding prior translation knowledge, especially on the source language syntax, is another significant approach done to NMT on expecting it to improve its performance. The method has proven to enhance the efficiency of NMT by making it far more aligned with the grammar of the source (K. Chen et al., 2021). Focusing on syntax benefited the system by exposing it to a deeper understanding of the source language structure, enabling suitable, accurate, and contextually appropriate translation.

Earlier attempts in the development of knowledge-based MT systems focused on the disintegration of syntactic from semantic knowledge with an intention to enhance parsing and generation. Tomita & Carbonell (1986) suggested a model within which domain-independent but language-oriented syntactic knowledge was embedded in translation processes. This partition enables MT systems to consider the grammatical intricacies of specific languages while still having the flexibility to implement domain-specific semantics.

The contribution of syntax has also been studied in knowledge-aware NMT systems. K. Xu et al. (2018) proposed a model that applies syntactic features simultaneously with lexical ones while maintaining the original form of the source language. This method aids the correct positioning of linguistic elements between languages with more advanced grammar, where the placement of syntax, along with semantics and lexicon, becomes crucial. The incorporation of structural linguistic information, alongside meaning representations, enhances the role of syntax in MT. These representations integrate syntactic theories with language-dependent world models, thus empowering multilingual translation. Such designs facilitate linguistic information exchange among languages while respecting the grammatical structure of the donor language (Mahesh & Nirenburg, 1996).

The integration of prior syntactic knowledge into MT systems, as noted above, facilitates tackling the problem of preserving intricate grammatical relations of the source language. The methods described above, regarding posterior regularization, source-centric training pipelines, and knowledge infused models, underscore the importance of syntax in improving the quality of translations. When utilized correctly, syntactic structures can increase the precision MT systems achieve as well as their compliance with human linguistic standards, thus resulting in improved translation mechanisms.

2.5Methods for incorporating prior knowledge into MT models

The incorporation of prior syntactic knowledge into machine translation (MT) models has proven useful when it comes to translation quality and alignment improvement. One such method is posterior regularization; representations of syntax knowledge are built as features in a log-linear model. These features serve as constraints during the NMT model training and help incorporate source language structures (T. Zhang et al., 2020).

A different attempt to model syntactic knowledge eliminates the guiding of statistical word alignment models to soft probabilistic constraints, which leads to entropy-derived principles or bilingual latent semantic analysis. This approach helps achieve better results in word alignments as they enable the use of structured syntactic frameworks sourced from grammar translation. This method is especially useful for enhancing statistical machine translation (SMT) models and helps prohibit the source language syntax violations during translation (Deng et al., 2007). One approach incorporates linguistic structures directly into parsing models for use with neural MT systems, resulting in a hybrid. As an example, NMT and RNNG model implement an RNNG parser with attention-based NMT model, where the former adds syntactic parsing information during training to the latter, enabling deep structural understanding of the source language. Such a setup improves the accuracy and fluency of translations without dependency on the RNNG during the translation step itself (Eriguchi et al., 2017).

Along with the integration of syntactic knowledge, meaning representation frameworks have also been crucial. A prominent case is that of language meaning representations that are not dependent on a specific language but still have information from the language's syntax. This combines linguistic knowledge with explicit textual data to form a representation that supports multilingual translation systems. Such systems map lexicons of individual languages with the general syntactic rules of those languages, resulting in improved accuracy and consistency across different languages (Mahesh & Nirenburg, 1996).

Another innovative approach utilizes implicit word knowledge from statistical machine translation to improve neural machine translation models. SMT's syntax-influenced predictions are used to inform the NMT's word generation probabilities under this framework. Syntactic knowledge is incorporated into neural architecture through mechanisms like gating or competition. This allows the NMT system to structurally align the source language's syntax into the translation, improving overall performance (Wang et al., 2019).

T. Zhang et al. (2020) analysis has focused on exploring new approaches for applying prior knowledge into a neural machine translation (NMT) system with a focus on enhancing performance and adaptability. They implemented a posterior regularization framework that encodes multiple sources of knowledge as features into a log-linear model for structured guidance through model training. This example is one of many that show the incorporation of different types of knowledge into NMT systems is possible.

M. Chen (2024) came forward a universal framework for the incorporation of prior knowledge of translation into NMT models that significantly advanced performance in benchmark tasks. Their work underscores the role of previously available knowledge of translation in improving accuracy as well as pertinence in context. Syntactic knowledge has also been integrated successfully using graph convolutional networks (GCN) on dependency trees. Bastings et al. (2017) showed that using GCNs improves the translation quality for English-German and English-Czech language pairs by exploiting the structural relationships that lie within syntactic representations.

Currey & Heafield (2019) advanced this work by adding constituency parse information to the transformer NMT models. Their multitask learning with a mixed encoder model achieved success on outcomes for twenty languages in the target languages, especially in under-resourced situations. This demonstrates the effectiveness of incorporating syntactic knowledge for data-scarcity challenges. Multiple approaches utilize Graph Convolutional Networks for the integration of syntax and linguistics to augment neural machine translation.

One method enhances word representations using GCNs to deepen their sensitivity to the surrounding syntactic environment, capturing better syntactic dependencies in the source language (Bastings et al., 2017). This approach stands out in improving translation accuracy because of its use of the syntactic structure of the source language. Another approach uses GCNs to add semantic-role representations, blending semantics and syntax in NMT to boost translation performance (Marcheggiani & Titov, 2017). These approaches illustrate the promise of GCNs to better exploit source syntax and relations within semantics, resulting in more precise, context-aware translations within NMT systems. These innovations illustrate the ability to integrate linguistic knowledge alongside task-specific knowledge into NMT systems. The combination of interest in syntactic structures, translation strategies, advanced architectures, and other innovative approaches is expected to lead to more effective, precise, and flexible NMT systems, particularly in multilingual and low-resource situations.

As a final note, synergizing previous knowledge with MT models is arguably one of the most critical strategies to enhance translation accuracy, grammatical correctness, and contextual coherence. The adaptability of the integration of syntax into different MT paradigms is shown through posterior regularization, hybrid parsing models, and meaning representation frameworks. These methods utilize syntax to tackle problems posed by the variety of languages as well as develop more powerful and accurate machine translation systems.

2.6Graph Attention Networks

The integration of graph-based representations into neural machine translation emerged as a response to the limitations of purely sequential models in capturing syntactic and semantic structure. Early efforts focused on enriching translation models with structured linguistic information to improve alignment, fluency, and contextual coherence across languages.

Initial breakthroughs appeared in 2017, marking a pivotal shift toward graph-based modeling in NMT. Hashimoto & Tsuruoka (2017) introduced a model that jointly learned latent graph structures and translation, outperforming both sequential and syntax-based systems on English-Japanese translation tasks. In the same year, Bastings et al. (2017) incorporated syntactic dependency information into NMT using Graph Convolutional Networks (GCNs), achieving substantial BLEU score improvements for English–German and English–Czech translations. These studies demonstrated that explicitly modeling linguistic structure yields measurable gains over sequence-only approaches.

A major advancement followed with the introduction of Graph Attention Networks (GATs) by Velickovic et al. (2017). GATs extended GCNs by replacing fixed neighborhood aggregation with attention mechanisms, enabling models to dynamically weight neighboring nodes based on their relevance. This innovation significantly increased modeling flexibility and provided a more expressive framework for capturing heterogeneous and long-range linguistic dependencies in graph-structured data.

Subsequent research expanded graph-based NMT beyond syntax to incorporate semantic and discourse-level information. Marcheggiani et al. (2018) employed GCNs to encode semantic role representations, demonstrating that the joint modeling of syntactic and semantic graphs further improves translation quality. In parallel, K. Xu et al. (2018) proposed a graph-to-sequence architecture with LSTM-based attention that outperformed multiple baselines on tasks involving graph-structured inputs, highlighting the general applicability of graph encoders in sequence generation.

By 2019, graph-based techniques began influencing decoding strategies and translation efficiency. Xia et al. (2019) integrated graph-based translation memory into NMT decoding, achieving improvements in both performance and computational efficiency. These developments underscored the potential of graphs to support not only representation learning but also inference-time optimization.

The year 2020 and beyond saw a surge of innovations that combined graph learning with advanced attention mechanisms and pretrained models. Duan et al. (2020) proposed the Graph-Transformer, which explicitly captures multi-order graph dependencies using self-attention. K. Chen et al. (2021) introduced the Contextualized Graph Transformer (CG-Transformer) to model document-level context, while Nguyen et al. (2021) integrated Abstract Meaning Representation (AMR) graphs through GATs to improve English–Vietnamese translation. These approaches demonstrated that graph-based attention effectively captures both local syntactic relations and global contextual dependencies.

Further advancements emphasized document-level coherence and multi-scale linguistic modeling. K. Xu et al. (2018) and K. Chen et al. (2021) developed document graph representations and the Effective Graph Context Representation (EGCR), respectively, showing that GATs outperform recurrent and convolutional architectures in modeling discourse-level dependencies. Nguyen et al. (2021) also proposed a multi-level community-aware GNN to simultaneously model local and global language structures, reinforcing the adaptability of graph-based architectures.

More recent studies have explored syntactic augmentation and hybrid graph–language model integrations. Gong et al. (2022) demonstrated that injecting synthetic constituency structures through graph encoders improves English-Vietnamese translation performance. Dai et al. (2023) introduced the Syntax-Guided BERT (SGB) model, which integrates GATs with pretrained BERT representations to incorporate syntactic information without degrading BLEU scores. These works highlight the compatibility of GATs with large pretrained language models and modern NMT pipelines.

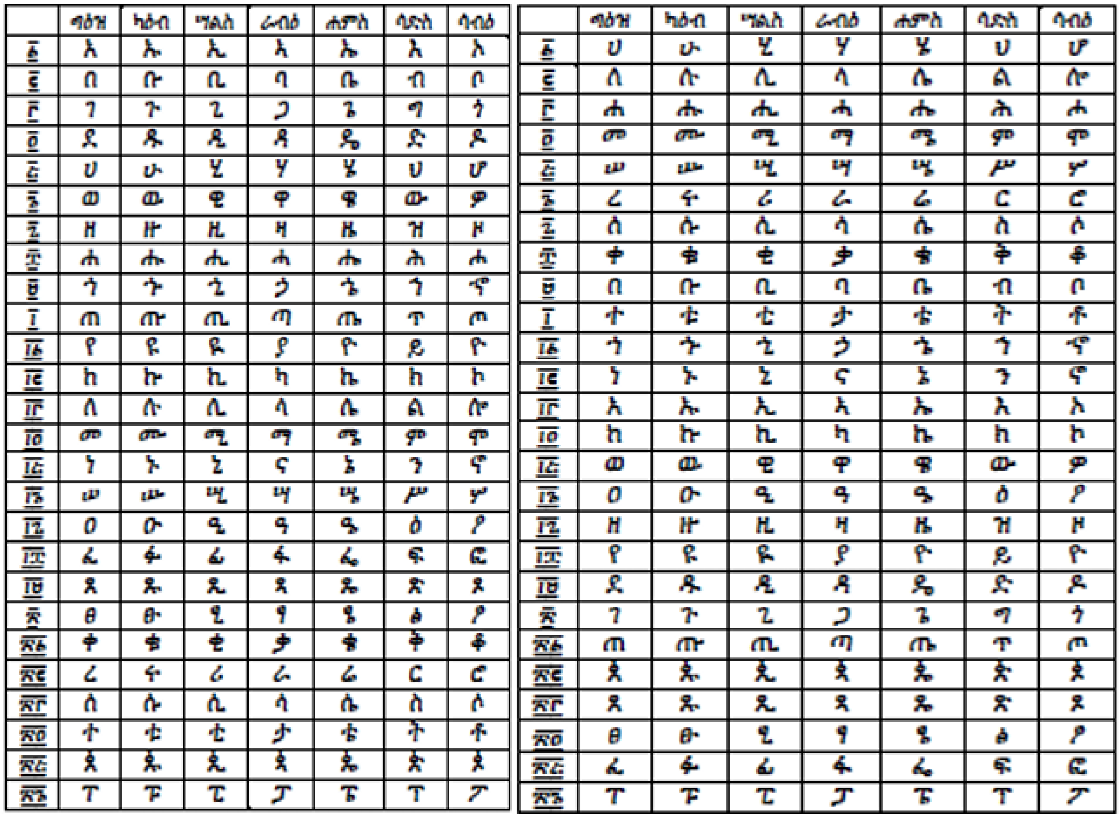
Collectively, research establishes graph-based attention mechanisms, particularly GATs as a powerful paradigm for advancing NMT. By enabling efficient, multi-level integration of syntactic, semantic, and contextual information, GAT-based models address core limitations of sequential architectures. This progression reflects a broader shift toward hybrid NMT systems that combine deep contextual embeddings with explicit linguistic structure, resulting in translations that are more accurate, coherent, and linguistically informed.

2.7 Amharic Language

2.7.1 Amharic Writing Systems

The purpose of the writing system is to represent a specific language visually or in a more understandable way. A symbol represents a sound, a punctuation mark, or a number. Six types of writing systems have been identified by studies (Coulmas, 2003). There are five primary alphabets (English, Russian, and Greek), six abjads or alpha syllabaries (Devanagari, Thai, Ge’ez, Amharic), four natural alphabets (Hangul), three syllabaries (Japanese, Cherokee), and one logographic alphabet (like Chinese). The Ge’ez language uses Abjad or Abugida derived from these. Until 330 A.D., the Abjad, which has 26 consonantal letters, was used, and vowels were not indicated (Tesfaye, 2020). Christian scripture influenced the development of abugida by adding a vocalic diacritic to consonantal letters. A variety of vowels, e, a, i, o, u, as well as diacritics, were combined with the consonants (Tesfaye, 2020). A left-to-right writing system was used before Aba Fremnatos, the first patriarch of the Ethiopian Orthodox Tewahedo Church. Geez, like Arabic, it was written from right to left (Tesfaye, 2020). The Amharic language also uses this form for its writing system. Ge’ez has two, the previous and the current alphabet arrangements. The previous alphabet arrangement uses the አቦጊዳ format, and the current alphabet arrangement uses the ሀሁ format. They use almost the same alphabetical arrangement, but there is a slight difference in some letters or fidel.

Table 2.2: The historical and modern ordering of the Ge’ez alphabet.



	ግዕዝ	ካዕብ	ሣልስ	ራብዕ	ኃምስ	ሳድስ	ሳብዕ
፩	አ	ኡ	ኢ	ኣ	ኤ	እ	ኦ
፪	በ	ቡ	ቢ	ባ	ቤ	ብ	ቦ
፫	ገ	ጉ	ጊ	ጋ	ጌ	ግ	ጎ
፬	ደ	ዱ	ዲ	ዳ	ዴ	ድ	ዶ
፭	ሀ	ሁ	ሂ	ሃ	ሄ	ህ	ሆ
፮	ወ	ዉ	ዊ	ዋ	ዌ	ው	ዎ
፯	ዘ	ዙ	ዚ	ዛ	ዜ	ዝ	ዞ
፰	ሐ	ሑ	ሒ	ሓ	ሔ	ሕ	ሖ
፱	ኀ	ኁ	ኂ	ኃ	ኄ	ኅ	ኆ
፲	ጠ	ጡ	ጢ	ጣ	ጤ	ጥ	ጦ
፲፩	የ	ዩ	ዪ	ያ	ዬ	ይ	ዮ
፲፪	ከ	ኩ	ኪ	ካ	ኬ	ክ	ኮ
፲፫	ለ	ሉ	ሊ	ላ	ሌ	ል	ሎ
፲፬	መ	ሙ	ሚ	ማ	ሜ	ም	ሞ
፲፭	ነ	ኑ	ኒ	ና	ኔ	ን	ኖ
፲፮	ሠ	ሡ	ሢ	ሣ	ሤ	ሥ	ሦ
፲፯	ዐ	ዑ	ዒ	ዓ	ዔ	ዕ	ዖ
፲፰	ፈ	ፉ	ፊ	ፋ	ፌ	ፍ	ፎ
፲፱	ጸ	ጹ	ጺ	ጻ	ጼ	ጽ	ጾ
፳	ፀ	ፁ	ፂ	ፃ	ፄ	ፅ	ፆ
፳፩	ቀ	ቁ	ቂ	ቃ	ቄ	ቅ	ቆ
፳፪	ረ	ሩ	ሪ	ራ	ሬ	ር	ሮ
፳፫	ሰ	ሱ	ሲ	ሳ	ሴ	ስ	ሶ
፳፬	ተ	ቱ	ቲ	ታ	ቴ	ት	ቶ
፳፭	ጰ	ጱ	ጲ	ጳ	ጴ	ጵ	ጶ
፳፮	ፐ	ፑ	ፒ	ፓ	ፔ	ፕ	ፖ

	ግዕዝ	ካዕብ	ሣልስ	ራብዕ	ኃምስ	ሳድስ	ሳብዕ
፩	ሀ	ሁ	ሂ	ሃ	ሄ	ህ	ሆ
፪	ለ	ሉ	ሊ	ላ	ሌ	ል	ሎ
፫	ሐ	ሑ	ሒ	ሓ	ሔ	ሕ	ሖ
፬	መ	ሙ	ሚ	ማ	ሜ	ም	ሞ
፭	ሠ	ሡ	ሢ	ሣ	ሤ	ሥ	ሦ
፮	ረ	ሩ	ሪ	ራ	ሬ	ር	ሮ
፯	ሰ	ሱ	ሲ	ሳ	ሴ	ስ	ሶ
፰	ቀ	ቁ	ቂ	ቃ	ቄ	ቅ	ቆ
፱	በ	ቡ	ቢ	ባ	ቤ	ብ	ቦ
፲	ተ	ቱ	ቲ	ታ	ቴ	ት	ቶ
፲፩	ኀ	ኁ	ኂ	ኃ	ኄ	ኅ	ኆ
፲፪	ነ	ኑ	ኒ	ና	ኔ	ን	ኖ
፲፫	አ	ኡ	ኢ	ኣ	ኤ	እ	ኦ
፲፬	ከ	ኩ	ኪ	ካ	ኬ	ክ	ኮ
፲፭	ወ	ዉ	ዊ	ዋ	ዌ	ው	ዎ
፲፮	ዐ	ዑ	ዒ	ዓ	ዔ	ዕ	ዖ
፲፯	ዘ	ዙ	ዚ	ዛ	ዜ	ዝ	ዞ
፲፰	የ	ዩ	ዪ	ያ	ዬ	ይ	ዮ
፲፱	ደ	ዱ	ዲ	ዳ	ዴ	ድ	ዶ
፳	ገ	ጉ	ጊ	ጋ	ጌ	ግ	ጎ
፳፩	ጠ	ጡ	ጢ	ጣ	ጤ	ጥ	ጦ
፳፪	ጰ	ጱ	ጲ	ጳ	ጴ	ጵ	ጶ
፳፫	ጸ	ጹ	ጺ	ጻ	ጼ	ጽ	ጾ
፳፬	ፀ	ፁ	ፂ	ፃ	ፄ	ፅ	ፆ
፳፭	ፈ	ፉ	ፊ	ፋ	ፌ	ፍ	ፎ
፳፮	ፐ	ፑ	ፒ	ፓ	ፔ	ፕ	ፖ

Table 2.3: Derived Ge’ez letters.

ሏ	ሗ	ሟ	ሧ	ሯ	ሷ	ቋ	ቧ	ቷ	ኗ	ኧ	ኳ	ቍ	ዟ	ዷ	ጓ	ጧ	ጷ	ጿ	ፇ	ፏ	ፗ

The Ge’ez language has 182 primary letters, as shown in the 2.2 table, with letters ordered as per the historical and the modern arrangement (comprising 7 groups of 26 letters each). Another table 2.3 shows the letters of Ge’ez derived from the original ones. Apart from several letters, the Amharic alphabet has the same letters as Ge’ez. As detailed in the table, the letters did not include total eight. This means there are 238 Amharic alphabets altogether, when these are added to the original set, as follows: (7 × 26) + (7 × 8) = 238 characters (which is 182 + 56). Amharic also has sets of derived alphabets, as shown in 2.3 table.

Table 2.4: Additional letters (fidels) in the Amharic language.

	ግዕዝ	ካዕብ	ሣልስ	ራብዕ	ኃምስ	ሳድስ	ሳብዕ
፩	ሸ	ሹ	ሺ	ሻ	ሼ	ሽ	ሾ
፪	ቸ	ቹ	ቺ	ቻ	ቼ	ች	ቾ
፫	ኘ	ኙ	ኚ	ኛ	ኜ	ኝ	ኞ
፬	ኸ	ኹ	ኺ	ኻ	ኼ	ኽ	ኾ
፭	ዠ	ዡ	ዢ	ዣ	ዤ	ዥ	ዦ
፮	ጀ	ጁ	ጂ	ጃ	ጄ	ጅ	ጆ
፯	ጨ	ጩ	ጪ	ጫ	ጬ	ጭ	ጮ
፰	ቨ	ቩ	ቪ	ቫ	ቬ	ቭ	ቮ

Table 2.5: Derived Amharic letters.

ሿ	ቿ	ኟ	ኧ	ዧ	ጇ	ጯ	ቯ

To conclude, the Amharic language has almost the same order of the alphabet as the Ge’ez language. The difference, however, is that there are more letters and derived characters. The extra alphabets and derivatives differ from Amharic to Ge’ez which means that letters are unique to one language. This shows the two languages are distinctly different from one another despite having the same basic script.

2.7.2 Syntax

Like many Semitic and Cushitic languages, Amharic primarily uses a Subject-Object-Verb (SOV) order. Nevertheless, Amharic shows some flexibility in its syntax: it may also occasionally use the

Object-Subject-Verb (OSV) order. This flexibility may be due to Amharic's particular linguistic situation, where it channels both Semitic and Cushitic elements (Coulmas, 2003).

Conversely, characteristic of the languages in the Ethiopian family, Geography-wise the order of the Subject-Verb-Object (SVO), Verb-Subject-Object (VSO), and Object-Verb-Subject (OVS) constructs can be used as such. Take for instance the following illustrative example in Amharic:

> አስተማሪው ተማሪዎችን ለማስተማር ወደ ክፍል ገባ።
> ("The teacher enters to class to teach students.")

Ge'ez allows for the use of any of the orders described above as illustrated in the following sentences:

Table 2.6: Examples of Amharic Sentences with Different Word Orders

Word Order	**Amharic Sentence**
SVO	አስተማሪው ወደ ክፍል ገባ ተማሪዎችን ለማስተማር
VSO	ወደ ክፍል ገባ አስተማሪው ለማስተማር ተማሪዎችን
OVS	ተማሪዎች ወደ ክፍል ገባ ለማስተማር አስተማሪው

Concretely, the sentence as illustrated in Amharic has the same subject as Ge'ez "አስተማሪው" which works the same in both languages. The object in Amharic, "ሰማይንና ምድርን" aligns with the Ge'ez object "ተማሪዎች" while the verb "ገባ እና ለማስተማር" translates in both languages. Amharic "ገባ" means 'enter' can be used to describe a completed or stative action depending on the context. The comparison of the use of syntax shows while Amharic seems more rigid in comparison to the Ge'ez flexible structures. The different historical evolutions and influences on the languages explain the variations in structures.

2.7.3 Similar Letters

As depicted in table 2.7 letters represent highly correlated phonetic sounds, but are spelled differently, or differently orthographically. Although they sound similar, meaning each letter is different, and affects the meaning of the word, phrase, or sentence. These different shapes and meanings are crucial for distinguishing the semantics of language. There is a total of nine letters which have this phonetic similarity and orthographic difference (Tesfaye, 2020).

Table 2.7: Similar letters of Amharic alphabet.

ውስሰት	**ክልሰት**	**ክልደት**	**ክልዐት**
ሐ	ሠ	ዐ	ጸ
ሀ	ሰ	አ	ፀ
ኀ			

2.7.4 Word Classes

Any language consists of thousands of elements; however, communication and dialogue usually only need a small number of them. Even though the vocabulary of a language may consist of tens of thousands of words, the average speaker only uses and recognizes a pinpoint of them in a regular discourse. Effective communication also depends upon the ability to combine single lexemes and perform this in a multitude of diverse hierarchically organized structural forms of language (Bache & Davidsen-Nielsen, 2010). In Amharic the word ‘ሰዋሰው’ encompasses all the relevant structural and combinatorial rules of a language and includes the arts of designing and in constructing phonology and syntax. The Ge’ez language also uses all the seven elements of the Greek language. The Ge’ez and Greek languages include all the needed parts of communication: nouns, adjectives, verbs, adverbs and conjunctions (Tesfaye, 2020).

Noun in Amharic Language

In Amharic as in any language, nouns are used to denote things, feelings, places, animals, and people, as well as abstract ideas. There are two general types of nouns in Amharic language: naturally occurring nouns and nouns that are derived from verbs. Naturally constructed nouns are those that exist independently as names of things or beings. These include አልጋ (bed), አህያ (donkey), አንገት (neck), ግመል (camel), and ጸጉር (hair). Such nouns are usually concrete, not abstract, and as such directly reference tangible things or living beings.

Nouns in Amharic can also be derived from verbs and used as formal nouns in clauses, phrases, and sentences. These verbal nouns relate to the core actions of the associated verbs. They include ማሰብ (to think) and ሀሳብ (thought), ባረከ (to bless) and መባረክ (blessing), and ጤና (health) and ቡራኬ (blessing). Such verbal nouns enrich the language by allowing actions to be described using nouns that reflect the abstract or related concept.

In Amharic, pluralization of nouns means attaching fidels to the singular form of the noun. Some pluralization markers include ኣ, ት, ን, ያን, ያት, ው, and ል. Of these, ኣ is a prefix while the others are suffixes. Amharic depends on these pluralization systems to form and differentiate meaning in sentences.

Pronoun

Pronouns are essential substitutes for nouns or noun phrases. They save speakers and writers the hassle and repetition of saying the same thing repeatedly. Communication would be cumbersome and longer without the use of pronouns. Each reference made would have the explicitly state the noun every time. Amharic language uses nine pronouns to achieve the same goal.

Table 2.8: Adding fidels to make the noun plural form.

Added fidels	Original known	Inflected to
አ	ደብር (ተራራ)	አድባር (ተራሮች)
አ……ን	ገብር (አገልጋይ)	አግብርት (አገልጋዮች)
ት	ንጉሥ	ነገሥታት
ን	ባዕድ	ባዕዳን
የን	ዘማሪ	ዘማሪያን
የት	ውዳሴ	ውዳሴያት
ው	አብ (አባት)	አበው (አባቶች)
ል	ኪሩብ	ኪሩቤል

However, in addition to replacing nouns, the pronouns in these languages also substitute some adverbs and adjectives and even some pronouns. This increases the multifunctionality and flexibility of these pronouns in the grammatical system. This role multifunction promotes clarity and the conciseness of expression of language. In Amharic, there are seven pronouns in the language along with their forms and functions are summarized in Table 2.9.

Table 2.9: Amharic Pronouns List

አማርኛ	**English**	**አማርኛ**	**English**
እኔ	I	አንተ	You
እኛ	We	እሱ	He
አንተ	You	እሷ	She

እንቺ	You (pl.)	እነርሱ	They
እናንተ	You (pl.)	-	-

As ‘verb to be’, pronouns also expressed as past tense.

Table 2.10: Amharic Pronouns as Verb “To Be”.

English	አማርኛ
Be, is, was	ደረሰኝ ነው፤ነበር
Be, is, was/will be	ነበረች፤ነርከች፤ነርሰች
Are, were	ነበሩ፤ነርሱ፤ነርሰሉ
Are, were/will be	ነበሩ፤ነርሰሉ፤ነርሰሉ
Are, were	ነኝ፤ነበር፤ነርሰል
Are, were/will be	ነርሰል፤ነርሰል፤ነርሰል
Are, were/will be	ነበሩ፤ነርሰሉ፤ነርሰሉ
Are, were/will be	ነበሩ፤ነርሰሉ
Am, was/will be	ነኝ፤ነበር፤ነርሰሉ
Are, were/will be	ነበሩ፤ነርሰሉ፤ነርሰሉ

Pronouns as demonstrative pronouns፡

Table 2.11: Ge’ez Pronouns as Demonstrative Pronouns.

አማርኛ	**English**	**አማርኛ**	**English**
ይህ፣ ይህው	This	ያ፣ ያው፣ ያውና	That
ይች፣ ይችው	This (Feminine)	ያች፣ ያችው፣ ያችውና	That (Feminine)
እኒህ፣ እኒሁ	These	እነዚያ፣ እነዚያው፣ እነዚያውና	Those
እኒህ፣ እኒሁ	These (Feminine)	እነዚያ፣ እኒያው	Those (Feminine)

Adjective

An adjective is described as a word that qualitatively expands or defines or contains identification for a noun or pronoun. A noun is showcased as a legal title of a thing, while an adjective is showcased as a secondary title on a thing that displays its attributes. Examples of these attributes are type, color, property, shape, and size, among others. These attributes also constitute distinguishing qualities or distinguishing features of objects (Tesfaye, 2020).

There are various ways of forming adjectives in Amharic. A simple Amharic morphological rule entails changing a verb into an adjective by changing the last letter into the 3rd letter of a specific order. For instance, the verb ፈጠረ (to create) turns into the adjective ፈጣሪ, describing a person that creates. A second rule transforms the verb into a 6th letter ending and considers ተግሀ (to be strong) changing to ትጉህ, which means strong or powerful. A third rule also transforms the verb by the prefix ‘መ’ (the 15th fidel) as in ዘመረ (to sing) and changes to መዘምር (related to singing or musical). You can also form adjectives from nouns by attaching the suffixes ‘ዊ’ or ’ዊት’. These signify characteristic quality or relatedness. For example, ገሊላ (Galilee), ገሊላዊ, or ገሊላዊት translates to ”of or pertaining to Galilee,” and ኢትዮጲያ (Ethiopia) transforms into ኢትዮጲያዊት, which means “Ethiopian” (in adjectival form).

Verb

Amharic considers verbs as action words containing the core of a predicating phrase in the sentence. Amharic verbs change as Amharic verbs gain additional prefixes and suffixes in the construction of a sentence to reflect the changes of the grammatical categories of person, number, voice, tense, and gender. The position of the verb in a sentence as the last component is characteristic of the Amharic language.

The verbs of Amharic maintain grammatical agreement with their subjects and objects, which stabilizes coherence in sentence formation. Other verbs in the language are built on these primary root verbs or are derived from them, and they follow the same morphological structure.

Adverb

An adverb is a word that qualifies, modifies, or adds description to other words, predominantly verbs, but also adjectives and other adverbs. In Amharic, adverbs can be categorized into seven groups based on their functions. They are time, frequency, place, manner, reason, and interrogative adverbs. Time adverbs clarify when an action happens. They anchor a sentence in time and can reference the present, past, or future. Sentences can also include temporal connectors, such as ”while” and ”during”. Examples are ነገ meaning “tomorrow” and ትናንትና, meaning ”yesterday”. An adverb of frequency states how often an action occurs. It indicates the regularity of the action and whether it is repeating. For instance, ሁሌ means “always”, and ሁልጊዜ translates to ”every time” or “on each occasion”. An adverb of place indicates where an action takes place. It provides spatial context to the surrounding elements in the sentence. Examples are እዚህ, meaning ”here” and ቅርብ, meaning ”near”.

An adverb of manner states how an action is performed. It conveys the quality or style of the action, such as እሙን, which means ”certainly” or ”indeed,” or በግልጽ, meaning “clearly” or ”distinctly”. An adverb of reason states the purpose of an action and explains why it is performed. For instance, the word ስለ means “about” or “approximately” and ስለሆነም translates to “because of” or “due to”. Adverbs of question are those which are used to ask questions, specifically questions related to various components of an action or an event. እንዴት means “how”, ምን translates to “what” and የቱ means “which”.

The different categories of adverbs are essential to the construction of Amharic sentences. They provide additional layers of meaning to the expressions, especially with regards to importance and include specification to time, duration, frequency, place, manner, causation, and other related questions concerning the actions and conditions.

Prepositions

Prepositions are grammatical elements which sit before a noun or pronoun. They define the relationship a noun or pronoun has with another word in the sentence. In English and Amharic, prepositions are relationship definers of people, things, time, and places. They ascertain and define the positions, directions, and timing of a given situation. For example, the word ከስር means ‘under’ defines a position, በጊዜው means ‘during’ defines timing, and አጃቢ means ‘accompany’ defines a relationship.

Including a pronoun or noun with a conjunction gives relevant information. As such, a sentence can attain a full meaning. It gives context, defines the situation, and gives clarity to people, places, or things within a sentence. A sentence can be accurate and valuable only if prepositions give information and context appropriately. Table 2.12 presents Amahric and English languages preposition.

Table 2.12: Some prepositions of Amharic and English languages.

አማርኛ	English
ላይ፣ በ - ላይ፣ ከ - ላይ	On, above
ታች፣ በ - ታች፣ ከ - ታች	Under
ውስጥ፣ በ - ውስጥ (ከ ውስጥ)፣ በመካከል	In/inside / in the middle
	Before/ after

ፊት - (በፊት)፣ ኋላ/ኋላለ ወደ ስለ ከ፣ ከ - ይልቅ	To About From

Conjunctions

Conjunctions are tools that connect different parts of sentences (clauses or phrases) or group words together in a single clause. They connect and logically relate various elements of language. They enable the speaker or writer to express relationships of addition and contrast, cause and effect, alternation, and condition. In other words, they enable the construction of sophisticated sentences.

In the Amharic language, conjunctions perform precisely the same functions. They integrate clauses or sentences to construct elaborate, meaningful expressions, and to achieve coherence and cohesion. These also enable the construction of thought compounds and complex expressions. The table 2.13 illustrates some of the common Amharic conjunctions with their uses and English translations. This illustration serves to demonstrate the importance of conjunctions to language.

Table 2.13: Some conjunctions of Amharic language.

ግለሰብ አማርኛ	**English**
እንደ/እንድ/እንዳ/እንዲ	As …As
ስለ	About
በ - ጊዜ	In …time
እና/ያህል/ስለ	And/Due to

2.7.5 Comparison Between English and Amharic Languages

We can summarize the discussion on the similarities and differences between English and Amharic in the following table.

Table 2.14: Differences and Similarities Between English and Amharic Languages.

Feature	**Similarity/Difference**
Writing System	Different
Syntax (Word Order)	Different
Numeral System	Different

Letters/alphabets	Different
Word Classes	Different

English and Amharic share no similarities in their writing system, letter forms, Syntax (word order), numerical system, letter similarities and word classes. They employ the different alphabetic order derived from different scripts, though Amharic has more additional letters to represent sounds than English. In addition, there are notable differences in syntax and numeral systems. Amharic predominantly follows a Subject–Object–Verb (SOV) word order, while English language follows order like Subject–Verb–Object (SVO) structures. In terms of numbers, Amharic possesses a richer set of numeral representations than English.

2.8Evolution and Future Prospects of Amharic-English Machine Translation

We conducted this section using the systematic literature review (SLR) methodology for data retrieval. The SLR process is defined as the methodical and systematic approach of locating primary studies to develop and evaluate a certain research question. Average literature surveys usually do not require as much attention to detail when presenting data within a systematized review. A systematic review aimed at combining and summarizing all information concerning a research issue offers greater validity in the conclusions than individual studies. The results are reported based on the latest PRISMA (preferred reporting items for systematic reviews and meta-analyses) framework (Antoniou et al., 2021).

2.8.1 Eligibility Criteria

To capture the full evolution of Amharic-English machine translation, this study imposed no restrictions on the year of publication. Titles, keywords, and abstracts of identified research articles were screened for relevance based on clearly defined inclusion and exclusion criteria.

Inclusion Criteria

The following criteria were used to include studies in the review:

- Research papers published in peer-reviewed journals or conference proceedings.
- Studies specifically focusing on machine translation between Amharic and English.
- Articles written in English.
- Only full-text articles were considered.

- Studies from any publication year were eligible.

Exclusion Criteria

Studies were excluded from the review based on the following conditions:

- Research on machine translation involving Amharic and English that also involved other languages without isolating Amharic-English translation.
- Studies that did not describe or evaluate Amharic machine translation.
- Research focusing on areas outside machine translation (e.g., speech recognition, information retrieval).
- Review articles, abstracts, commentaries, posters (short papers), or editorials.
- Articles that were not fully accessible or retrievable.
- Duplicate publications.

2.8.2 Source of Information

To identify relevant publications, we conducted comprehensive searches across five widely used databases and libraries: ScienceDirect (https://www.sciencedirect.com), Google Scholar (https://scholar.google.com), ACM Digital Library (https://www.acm.org), IEEEXplore (https://ieeexplore.ieee.org), and SpringerLink (https://www.springerlink.com). This approach ensured a focus on reputable and high-quality sources. Our search included examining titles, keywords, and abstracts.

2.8.3 Search strategies

We conducted a thorough search for published literature using the specified resources and search terms. We included all relevant literature without any limitations on publication dates, except for journals from the social science fields. After numerous iterations of trial and error, the final search query is as follows:

[("Machine translation" OR MT OR "Computer translation" OR "Automatic translation" OR "Automatic text conversion") AND ("Amharic text" OR "Amharic translation" OR "Amharic text translation" OR "Amharic") AND ("English language" OR English)].

Table 2.15: Keywords used for each database,

Database	Keywords	Search results	Included in the review
Science direct (ScienceDirect)	(“Machine translation” OR MT OR “Computer translation” OR “Automatic translation”) AND (“Amharic text” OR “Amharic”) AND (“English language” OR English)	40	0
Association for Computing Machinery (ACM) Digital Library	+("Machine translation""automated translation""automatic translation")+("Amharic")+("english")	37	1
Institute of Electrical and Electronics Engineers (IEEE) Xplore	(“Machine translation” OR MT OR “Computer translation” OR “Automatic translation”) AND (“Amharic text” OR “Amharic”) AND (“English language” OR English)	9	5
Google Scholar	“"Machine translation" "Machine translation" OR MT OR "Computer translation" OR "Automatic translation" OR "Amharic text" OR Amharic OR "English language" OR English "Amharic AND English"”	96	24
Springer Link	(“Machine translation” OR MT OR “Computer translation” OR “Automatic translation”) AND (“Amharic text” OR “Amharic”) AND (“English language” OR English)	19	1

Table 2.15 summarizes the results from various academic databases regarding machine translation, specifically its application to Amharic and English. In addition, it provides a clear overview of the literature available across different databases on the topic of machine translation between Amharic and English.

2.8.4 Selection process

The publication selection process involves applying inclusion and exclusion criteria to identify primary sources relevant to our research questions. Specifically, we select research papers, conference proceedings, and book chapters that focus on machine translation between Amharic and English. Any other publications not related to Amharic machine translation were excluded. After finalizing the selection, we compile all the chosen papers from the five search engines into a single CSV file.

2.8.5 Data collection process

We initiated the data collection process by identifying relevant databases and selecting appropriate keywords for our search. The chosen databases included well-known academic resources such as Google Scholar, IEEE Xplore, ACM Digital Library, SpringerLink, and Scopus. Using a set of predefined keywords related to Amharic and English machine translation, we conducted comprehensive searches within these databases.

Upon obtaining the search results, we meticulously examined the titles of the retrieved works to determine their relevance to our research objectives. This initial screening helped us filter out irrelevant papers and focus on those directly addressing machine translation between English and English. Through this systematic approach, we collected a robust set of research works for further analysis.

Next, we applied our inclusion and exclusion criteria to further refine the selection process, ensuring that only pertinent studies were considered. The final step involved consolidating all the selected papers from the different databases into a single CSV file, facilitating organized and efficient data handling for subsequent analysis stages.

2.8.6 Study selection

We employed broad searching strategies and selection criteria to retrieve 201 studies from different databases. The article distribution is as follows: ScienceDirect had 40, ACM Digital Library 37. There were 9 in IEEE Xplore, 96 in Google Scholar, and 19 in Springer Link which were all based on our set search words. While following these steps, we collected a lot of records that we later purged of duplicates, bringing the number down to 22. Every single record was then screened against the set inclusion and exclusion criteria to determine their suitability with our research goals. The complete PRISMA 2020 framework is given in Figure 2.2, which shows the flow of information through the different phases of selection focusing on the eligibility criteria.

These steps guarantee the set of articles left after this process are suitable to form a basis for research on machine translation between English and English.

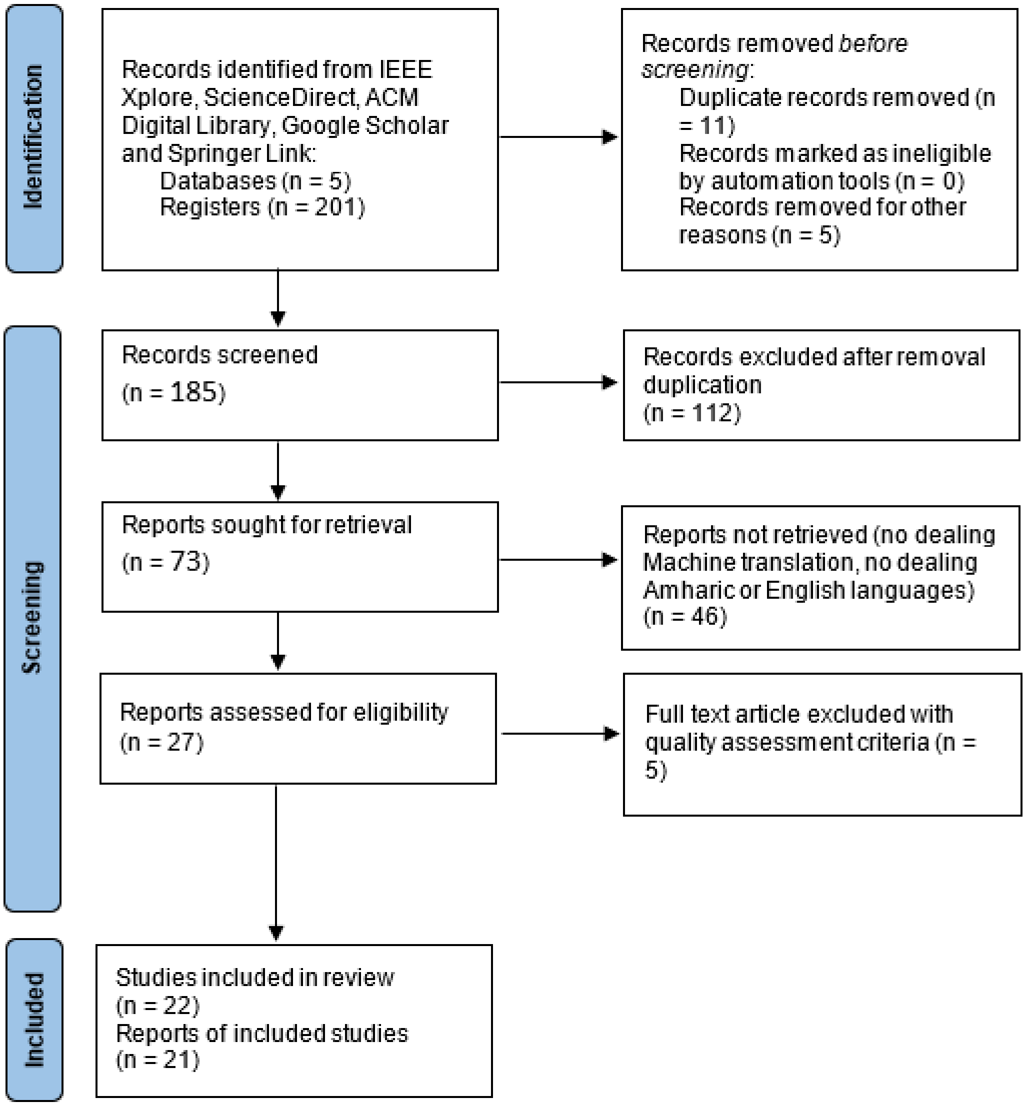


Figure 2.2: The PRISMA framework for the research screening process.

2.8.7 Document Historical Development and Milestones

The publication timeline for Amharic to English machine translation spans from 2012-2023, as depicted in Figure 2.3. The inaugural publication emerged in 2012, marking the beginning of scholarly interest in this field. Over the years, there has been a noticeable fluctuation in the number of publications. Certain years stand out with a higher volume of research output, indicating periods of intensified academic focus and advancements. The most significant surge in publications occurred in 2022, which recorded the highest count within the given timeframe. This trend reflects the growing interest and development in the domain of machine translation between Amharic and English. The overall publication pattern exhibits variability, with some years experiencing a greater influx of studies than others do, highlighting the evolving nature of research activity in this area.

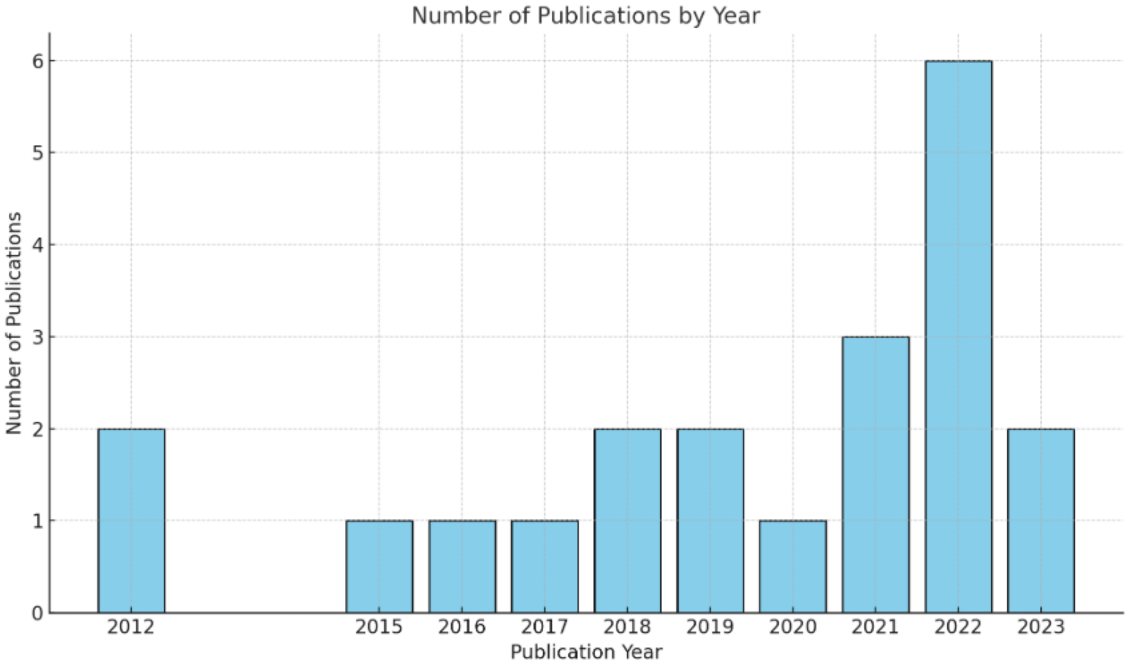


Figure 2.3: Number of studies per year of publication.

Figure 2.3 illustrates the distribution of machine translation methodologies in percentage. These methodologies include Rule-Based Translation, Statistical Machine Translation (SMT), Neural Machine Translation (NMT), and the Transformer model.

- *Rule-Based Translation (10%):* Only two studies employed rule-based approaches, highlighting their limited use compared to modern techniques. De Pauw et al. (2012) and Kore et al. (2017) focused on implementing rule-based systems for translating English to Amharic and proper noun transliteration, respectively.
- *Statistical Machine Translation (SMT, 45%):* SMT dominated earlier research, with nine studies exploring its application for translating between Amharic and English languages. SMT experiments often served as baselines for comparison with NMT methods.

- *Neural Machine Translation (NMT, 35%):* Seven studies utilized NMT, reflecting a shift towards neural approaches. These works included developing attention-based architectures (Gashaw & Shashirekha, 2019), leveraging subwords for handling inflectional morphology (A. M. Gezmu et al., 2021), and hybrid methods combining contextual information (Ashengo et al., 2021).
- *Transformer Models (30%):* The adoption of Transformers is growing, with six studies emphasizing their effectiveness in low-resource language pairs. Researchers like Hadgu et al. (2022) and Destaw Belay et al. (2022) demonstrated superior performance compared to previous methodologies, using techniques like homophone normalization and corpus augmentation.

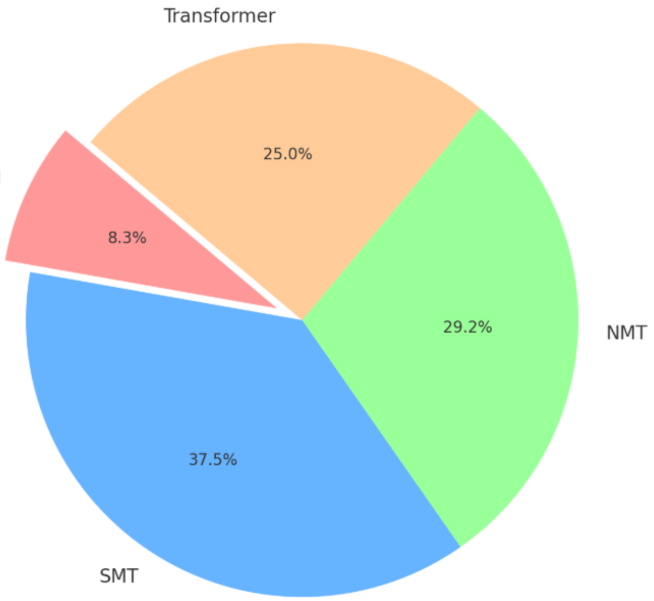


Figure 2.4: Distribution of methodologies in machine translation research.

The selected publication contains detailed information on various research papers focused on machine translation involving Ethiopian languages such as Amharic, Tigrinya, and Ge'ez (which have the same scripture). Table 2.16 is organized into columns of authors, methodologies, and contributions. Each entry lists the authors of the paper; the specific machine translation methodology used (including rule-based, statistical, neural, and transformer-based approaches); and a summary of the research contributions. For example, studies by M. Teshome et al. (2015) explored statistical machine translation (SMT) between English and Amharic, achieving incremental improvements in BLEU scores through different experimental setups. Other studies by (Gashaw & Shashirekha, 2019) and colleagues examined neural machine translation (NMT) between Amharic and Arabic, comparing LSTM and GRU models. Some works, like A. M. Gezmu et al. (2022), focus on creation and enhancement of parallel corpora specifically aimed at improving SMT and NMT models using a large Amharic-English corpus, while other, Hadgu et al. (2020), developed the Lesan translation system which aids low-resource language translation by using transformer models and back-translation methods. Those enabled and insightful dataset reflects the most important milestones, issues, and prospects of machine translation research and development for Ethiopian languages, demonstrating different methods and notable progress in translation quality.

Table 2.16: Summary of Key Studies in English-Amharic Machine Translation Research

Author	Methodology	Contribution
De Pauw et al. (2012)	Rule-based machine translation	Described key aspects of an ongoing project to implement a rule-based English-to-Amharic and Amharic-to-English machine translation system.
M. G. Teshome & Besacier (2012)	Statistical machine translation	Discussed the experiment conducted to translate from English to Amharic using the Statistical Machine Translation (EASMT) approach.
M. Teshome et al. (2015)	Statistical machine translation	Focused on improving translation quality by applying phonemic transcription on the target side, resulting in a BLEU score improvement.
Tedla et al. (2016)	Statistical machine translation	Presented initial research on English-to-Tigrinya SMT, addressing morphological segmentation of Tigrinya words to reduce data sparseness and improve translation quality.
Kore et al. (2017)	Rule-based machine translation	Proposed a rule-based machine transliteration technique for English to Amharic proper nouns, achieving 90.08% precision in correct transliterations.
S. Abate et al. (2018)	Statistical machine translation	Described the development of parallel corpora for English and Ethiopian Languages for bidirectional SMT experiments, highlighting the impact of morphological richness on SMT performance.
Gashaw & Shashirekha (2019)	Neural Machine Translation	Developed Amharic-Arabic NMT models using Attention-based Encoder-Decoder architecture, comparing LSTM and GRU models, and found that LSTM outperforms GRU and Google Translation system.

S. T. Abate et al. (2019)	SMT	Described the development of parallel corpora for English and Ethiopian Languages for bidirectional SMT experiments, highlighting the impact of morphological richness on SMT performance.
Gashaw & Shashirekha (2020a)	Neural Machine Translation	Constructed a small parallel Quranic text corpus for Amharic-Arabic NMT experiments, comparing LSTM and GRU based models with Google Translation.
A. M. Gezmu et al. (2021)	Neural Machine Translation	Described neural machine translation between Amharic and English using a new transliteration technique for Amharic and subwords to handle highly inflectional morphology.
Biadgligne & Smaïli (2021)	SMT and NMT	Developed an English-Amharic parallel corpus and conducted SMT and NMT experiments, with the corpus freely shared for research. SMT achieved 26.47 BLEU and NMT achieved 32.44 BLEU.
Ashengo et al. (2021a)	Neural Machine Translation	Investigated a new approach combining context-based machine translation (CBMT) with RNNMT for English-Amharic translation, showing performance improvement over simple NMT.
Biadgligne & Smaïli (2022)	SMT and NMT	Investigated the effect of corpus augmentation on English-Amharic MT quality, showing improved BLEU scores for both SMT and NMT models.
Hadgu et al. (2022)	Transformer	Presented Lesan, an MT system for low-resource languages using a custom OCR system and Transformer model, outperforming Google Translate and Microsoft Translator for Tigrinya, Amharic, and English translations.
Destaw Belay et al. (2022)	Transformer	Developed bidirectional Amharic-English NMT models using the Facebook M2M100 pretrained model, achieving BLEU scores of 37.79 for Amharic-English and 32.74 for English-Amharic translation, and explored the effects of Amharic homophone normalization.
Biadgligne & Smaïli (2022)	Transformer	Applied corpus transliteration and augmentation techniques to improve English-Amharic MT performance, achieving the highest BLEU score for the language pairs using Transformer models.
A. Gezmu et al. (2023)	morpheme-based NMT	Investigated morpheme-based NMT models for low-resource fusion languages, showing that morpheme-based models outperform conventional subword models on benchmark datasets.
Getachew & Yayeh (2023)	Transformer	Proposed a bidirectional NMT model for Ge'ez-English translation using the Transformer model, achieving BLEU scores of 27.19 for English-Ge'ez and 29.39 for Ge'ez-English translation, despite dataset scarcity.

A. Gezmu et al. (2022)	SMT and NMT	Described the acquisition, preprocessing, segmentation, and alignment of an Amharic-English parallel corpus, demonstrating that NMT models outperform SMT models by approximately six to seven BLEU points.
Negia et al. (2023)	Transformer	Attempted to design Amharic-Kistanigna bidirectional MT using various deep learning models, concluding that the Transformer model achieved the highest BLEU scores of 7.73 for Amharic-Kistanigna and 4.43 for Kistanigna-Amharic translations, but highlighted the need for more parallel corpora.
Andargachew MG et al. (2023)	morpheme-based NMT	Investigated morpheme-based NMT models for low-resource fusion languages, demonstrating that morpheme-based models outperform conventional subword models on benchmark datasets, and created a new dataset for a low-resource language.

This systematic review revealed that the field of Amharic to English machine translation (MT) has undergone significant developments and milestones over the years, reflecting broader advancements in machine translation technologies. The following is a summary of the key historical developments and milestones:

Early developments: Starting in 2012: Rule-based approaches (De Pauw et al., 2012; Kore et al., 2017)

Initial Efforts: The earliest attempts at machine translation between Amharic and English relied primarily on rule-based approaches. These systems were built via linguistic rules and require extensive knowledge of both languages' grammar and syntax.

Challenges: These early systems faced challenges due to the complex morphology of Amharic and the lack of extensive digital resources.

Statistical Machine Translation (SMT): 2012 was the beginning of SMT for Amharic (M. G. Teshome & Besacier, 2012).

Parallel Corpora Development: The development of parallel corpora, such as the Amharic-English Bible corpus, provided essential data for training SMT models.

GIZA++ and Moses Toolkit: Tools such as GIZA++ for word alignment and the Moses toolkit for phrase-based SMT became instrumental in developing Amharic-English SMT systems.

Notable Works: Research projects and academic efforts during this period focused on leveraging SMT techniques, resulting in moderate improvements in translation quality using Ge'ez text (Tedla et al., 2016; E. Teshome, 2013; M. G. Teshome & Besacier, 2012). These systems benefit from bilingual dictionaries and aligned texts, but their performance is still limited by the scarcity of large, high-quality parallel corpora.

Neural Machine Translation (NMT): 2019 was the beginning of the NMT for Amharic text (Gashaw & Shashirekha, 2019).

LSTM and GRU Models: The introduction of neural machine translation models using long short-term memory (LSTM) and gated recurrent units (GRUs) marked a significant shift. These models were better at handling the complexities of Amharic grammar and provided improved translation accuracy compared with SMT (Biadgligne & Smaïli, 2022; A. M. Gezmu et al., 2022).

Parallel Corpus Expansion: Efforts have been made to expand parallel corpora, incorporating news articles, government documents, and other bilingual texts to train more robust NMT systems (A. M. Gezmu et al., 2022).

2022 - Present: Transformer models for the Amharic language (Destaw Belay et al., 2022)

Transformer Architecture: The adoption of transformer models, as exemplified by OpenNMT and similar frameworks, revolutionized the field. The transformers offered superior handling of long-range dependencies and contextual information, leading to substantial improvements in translation quality (Hadgu et al., 2022).

Back-Translation and Data Augmentation: Techniques such as back-translation, where monolingual Amharic texts are translated into English and then used to train the model, help mitigate the issue of limited parallel corpora (Biadgligne & Smaïli, 2022).

Overall, this development highlights the importance of technological innovation, custom-made methodologies and data enlargement in addressing the unique challenges of translating between Amharic and English. Future improvements in this domain will prospectively focus on further refining these methods and expanding data resources to attain even greater translation accuracy and accessibility.

2.8.8 Identifying Key Challenges and Limitations

Table 2.17 summarizes the key challenges and limitations faced by various research efforts in the field of Amharic-to-English machine translation, as documented in multiple studies from 2012-2023. The table is organized into two columns: Authors, and challenges and limitations.

Table 2.17: Challenges and Limitations in English-Amharic Machine Translation Research

Author	**Challenges and Limitations**
M. G. Teshome & Besacier (2012a)	Limited computational linguistic resources and integrated linguistic knowledge. The unique Ge'ez-based writing system of Amharic complicates the adaptation of existing tools designed for languages with different scripts.
M. G. Teshome et al. (2015b)	The linguistic diversity between Amharic and English presents significant challenges for machine translation (MT). Capturing nuances, idiomatic expressions, and cultural references is difficult due to the distinct characteristics of the two languages.
S. Abate et al., (2018); S. T. Abate et al., (2019)	Due to the linguistic barrier, there is a shortage of data that hinders training of translation models. The existence of high structural differences within Ethiopian languages creates considerable issues for Statistical machine translation (SMT). Moreover, the scarcity of linguistics and NLP tools for African languages makes the problem even worse.
Gashaw & Shashirekha (2019, 2020)	Difficulties in domain adaptation, rare terms, lengthy sentences and phrases, as well as word alignment discrepancies are among the shortcomings of NMT. The absence of parallel corpora for the NMT Amharic-Arabic language pair, along with the rich morphology of Amharic, the lack of capitalization, and small sized machine-readable lexicons add to the complexity of the problem.
Mekonnen Gezmu et al. (2021)	Amharic-English machine translation poses a distinctive challenge due to morphological divergence of languages. NMT suffers heavily from the lack of data for effective training. Carefully designed implementation is required for subword-based models to outperform word-based models in translation. Linguistic translation tasks are further rendered difficult due to the divergence of linguistic syntactic structures.

Ashengo et al. (2021a)	Large parallel corpora are essential for fluent machine translation, yet context unawareness in approaches like phrase-based machine translation (PBMT) hampers performance. While combinational approaches improve over simple NMT, rare words and uncommon vocabulary remain problematic.
Hadgu et al. (2022), Destaw Belay et al. (2022), (Negia et al., 2023)	The scarcity of datasets for low-resource languages limits the development of effective translation systems. Obtaining large-scale parallel corpora is challenging, and translation quality remains an issue for languages like Amharic. Additionally, limited Amharic linguistic resources and a small number of parallel sentences constrain deep learning experiments. Dependency on handcrafted features in rule-based MT further hampers progress.
Getachew & Yayeh (2023)	Dataset scarcity restricts extensive experimentation for improved results. Translating Ge'ez to English is time-consuming due to the script's longer word counts and its agglutinative nature (i.e Amharic/Ge'ez languages combine multiple morphemes (word units) into single words, leading to longer and more complex expressions. This can make it difficult for translation models to accurately parse and understand the intended meaning.), which adds to the computational complexity and increases training time for translation models. Additionally, Ge'ez is morphologically rich, meaning that words can take on several forms based on grammatical context. This richness can complicate encoding and require more sophisticated handling in machine translation task.
A. Gezmu et al. (2023)	Variations in morphological typology present challenges in determining optimal vocabulary sizes for subword NMT models. Nondeterministic training processes and the lack of specified stopping criteria for NMT model training further complicate development.
Andargachew MG et al. (2023)	Neural machine translation requires substantial training data and parallel corpora. Ensuring faithful and fluent translations is challenging due to language variations. Recurrent neural networks (RNNs) struggle with long-distance dependencies, which are critical for accurate translations.

Throughout the years, Amharic to English machine translation (MT) has had its share of challenges, hurdles, and constraints. One of the major problems is deeply technological and linguistic in nature. Some of the challenges involve complex inflectional Amharic morphology which enables a word to encapsulate a great deal of grammatical information such as tense, aspect, person, gender, and number. These statistical and rule-based systems have a considerable amount of difficulty dealing with such complicated forms because their accuracy in parsing and generating complex forms is abysmal.

Key challenges

a) Data scarcity

Parallel Corpora: There is a significant lack of parallel corpora for Amharic and other Ethiopian languages, which is crucial for training both statistical and neural machine translation models. Although a number of efforts have been made to develop parallel datasets for these languages (S. Abate et al., 2018; Destaw Belay et al., 2022; Gashaw & Shashirekha, 2020a; Hadgu et al., 2022), the scarcity of such resources continues to hinder the development and improvement of translation systems. We therefore encourage further research and collaboration to address this gap and better support these low-resource languages.

NLP Resources: The shortage of basic linguistic resources, such as morphological analyzers, machine-readable lexicons, and annotated datasets, impacts the effectiveness of translation models (S. Abate et al., 2018; S. T. Abate et al., 2019; Gashaw & Shashirekha, 2020a; A. M. Gezmu et al., 2022).

The lack of parallel corpora and essential linguistic resources for Amharic and other Ethiopian languages hinders the advancement of machine translation systems drastically. The development of parallel texts which are both high quality and sufficient is needed for the training of statistical and neural machine translation models. Furthermore, none of these resources is available, such as annotated datasets and morphological analyzers, which makes the construction of reliable translation models even more difficult. There is no doubt that without overcoming these shortages, the use of machine translation in Amharic and other Ethiopian languages will continue to be ineffective and inadequate.

b) Morphological Complexity

Inflectional and Derivational Morphology: Amharic's rich morphological structure, where single words carry extensive grammatical information, poses a substantial challenge. The ability of language to create new words through various prefixes and suffixes increases the complexity of translation systems (S. T. Abate et al., 2019; Gashaw & Shashirekha, 2019; A. Gezmu et al., 2023; Mekonnen Gezmu et al., 2021).

Agglutination: The frequent combination of multiple morphemes into a single word adds another layer of difficulty for parsing and generating accurate translations (A. Gezmu et al., 2023).

c) Syntactic Structure

Word Order: The syntactic difference between Amharic's subject-object-verb (SOV) order and English's subject–verb-object (SVO) order necessitates complex reordering algorithms to maintain grammatical coherence and meaning during translation (S. Abate et al., 2018; Mekonnen Gezmu et al., 2021).

Syntactic Divergence: The divergence in sentence structure requires advanced handling to preserve the intended meaning and fluency of the translation (A. M. Gezmu et al., 2021; Mekonnen Gezmu et al., 2021).

d) Domain Mismatch and Lexical Issues

Domain Mismatch: The challenge of translating domain-specific content due to differences in vocabulary and context between the source and target languages (Gashaw & Shashirekha, 2019)

Rare Words and Long Sentences: The presence of rare words and long sentences further complicates the translation process, particularly in NMT systems (Ashengo et al., 2021a; Gashaw & Shashirekha, 2019).

Limitations

a) Unique Writing System

Capitalization and Diacritics: The absence of capitalization and the critical role of diacritics in Amharic add another layer of complexity to accurate translation (Gashaw & Shashirekha, 2019).

b) Handling Nuances and Cultural References

Idiomatic Expressions and Proverbs: Amharic has numerous idiomatic expressions and proverbs that do not have direct equivalents in English, requiring a deep understanding of the cultural context (M. G. Teshome et al., 2015b).

Honorifics and politeness: Variations in the use of honorifics and levels of politeness between Amharic and English necessitate careful handling to maintain appropriate tone and respect in translation (M. G. Teshome et al., 2015b).

c) Technical Constraints

Training Data Requirements: Both statistical and neural machine translation models require large amounts of training data, which are challenging to obtain for low-resource languages such as Amharic (S. Abate et al., 2018; Ashengo et al., 2021a; Biadgligne & Smaïli, 2022).

To sum up, doing Amharic-to-English machine translation is laudable, but accomplishing it is seemingly enveloped by insurmountable obstacles of inadequate technology, data, and the language's intricacy. The lack of well-formed parallel corpuses and requisite NLP tools is greatly compounded by the rich fusional nature of Amharic as well as its comparatively more complex structural composition in relation to English which makes it almost impossible to create effective machine translation devices. In addition, American and Ethiopian cultural subtleties together with some specialized domains make it hard to improve the quality of the translation. Overcoming these problems calls for enhancing the current systems and resources by combining them with novel approaches, such as investing in deep text-to-text transform networks, particularly increasing the funding. Filling these voids will improve the process of translating Amharic to English making it accurate and more culturally relevant.

2.8.9 Assess Current Trends and Future Directions

Table 2.18 presents research trends and future directions in machine translation for Amharic and other low-resource languages because of works published in 2022 and 2023. To assess current trends and future directions, we consider papers published in 2022 and 2023.

Table 2.18: Current Trends and Future Directions in English-Amharic Machine Translation.

Author	Current Trends	Future Directions
Andargachew Mekonnen Gezmu	Corpus augmentation enhances MT models for under resourced	Investigate corpus augmentation impact on other underresourced language translations.

& Andreas Nürnberger Ernesto William De Luca Michael Gasser (2023; Biadgligne & Smaïli, 2022; A. Gezmu et al., 2023)	languages. Token-level augmentation manipulates text to retain original semantics. Morpheme-based and subword-based NMT models outperform conventional models. Automated metrics such as BLEU and ROUGE offer a standardized and reproducible method for evaluating translation models, minimizing subjective biases that can arise in human assessments.	Explore advanced tokenization techniques for further translation quality enhancement. Incorporate linguistic knowledge into NMT models for future research. Investigate efficacy of morphological segmentation tools in low-resource NMT. Explore morphological segmentation tools for low-resource NMT of fusion languages. Increase the size of the Amharic-English parallel corpus for NMT.
Hadgu et al. (2022)	Data preprocessing and model architecture of Lesan MT system contributed to its promising results for low-resource languages. Lesan outperforms Google Translate and Microsoft Translator in human evaluation. Lesan's MT models are implemented using OpenNMT toolkit.	Leverage Lesan for broader language support on online platforms. Enhance Lesan's translation model for more low-resource languages.
Destaw Belay et al. (2022; Getachew & Yayeh, 2023)	Transformer model dominates NMT paradigm for machine translation tasks. Normalization of Amharic homophones enhances Amharic-English machine translation performance. Limited studies on Amharic-English translation due to scarce linguistic resources. Ge'ez-English NMT using Transformer models shows promising results. Highly used evaluation metrics of machine translation is BLEU.	Expand datasets for more languages and use data augmentation techniques. Explore alternative pretrained language models for Amharic-to-English translation to serve as trainers, helping to address the challenges posed by low-resource settings. Use more corpora for higher quality results in future studies.
, A. M. Gezmu et al. (2022; Negia et al., 2023)	Extended parallel corpus for Amharic-English Machine Translation using SMT and NMT.	Increase the size of the Amharic-English parallel corpus for NMT. Enhance

	Amharic-Kistanigna Bidirectional Machine Translation using Deep Learning.	bidirectional translation capabilities using advanced deep learning techniques.

Current research trends in Amharic to English machine translation focus on leveraging advanced NMT techniques, enhancing linguistic resources, and improving evaluation metrics. Future directions aim to expand these efforts by incorporating more sophisticated models, increasing data resources, and further refining translation techniques.

Current trends

1. Corpus Augmentation and Token-Level Manipulation

Researchers are focusing on corpus augmentation to enhance machine translation (MT) models for under-resourced languages. Token-level augmentation, which manipulates text while retaining the original semantics, is particularly effective (Biadgligne & Smaïli, 2022; A. Gezmu et al., 2023).

Morpheme-based and subword-based neural machine translation (NMT) models are gaining traction, outperforming conventional subword models by leveraging automated evaluation metrics (Andargachew Mekonnen Gezmu & Andreas Nürnberger Ernesto William De Luca Michael Gasser, 2023; A. Gezmu et al., 2023).

2. Transformer Models and Normalization Techniques

Transformer models dominate the NMT paradigm for machine translation tasks, especially for low-resource languages such as Amharic (Destaw Belay et al., 2022; Getachew & Yayeh, 2023). In addition, normalization techniques, such as addressing Amharic homophones, have been shown to enhance Amharic-English translation performance (Destaw Belay et al., 2022).

3. Extended parallel corpora and bidirectional translation

Research has focused on extending parallel corpora for Amharic-English machine translation, which is crucial for improving translation quality (A. M. Gezmu et al., 2022). Bidirectional translation models using deep learning techniques are being developed, showing promising results in improving translation accuracy (Negia et al., 2023).

Future Directions

a. Advanced Corpus Augmentation and Tokenization

Future research should investigate the impact of corpus augmentation on other underresourced language translations and explore advanced tokenization techniques to further enhance translation quality (Biadgligne & Smaïli, 2022; A. M. Gezmu et al., 2022). The incorporation of linguistic knowledge into NMT models and improvements in morphological segmentation tools are suggested for advancing translation models (A. M. Gezmu et al., 2022).

b. Enhancing Transformer Models and Data Resources

Expanding datasets for more languages, using data augmentation techniques, and exploring other pretrained language models are crucial for improving Amharic-English translations (Destaw Belay et al., 2022; Getachew & Yayeh, 2023). The use of more corpora and enhancing translation quality through transformer models are essential next steps (Getachew & Yayeh, 2023).

For the most part, the latest analysis in the field of Amharic-English translation focuses on the application's augmentation with modern neural machine translation (NMT) approaches such as adding transformer models, corpus enlargement, and manipulation at token levels. Moreover, parallel corpora extensions and the inclusion of cross-directional translation models have shown remarkable progress towards solving problems related to this language pair of limited resources. Further development of the translation system requires substantial improvement in tokenization techniques, data resource enhancement, and corpus expansion. Besides, the integration of pre-trained language models and the enrichment of the datasets with new topical areas need to be dealt with to advance the Amharic-English machine translation system capabilities.

2.9Related Work

In this part, we study the intricacies of machine translation of the Amharic language. We highlight various methods and the gaps in this field, alongside the lack of attention offered to the Amharic language, which is considered a low-resource language because of sparse parallel corpora (Destaw Belay et al., 2022).

Various approaches have been attempted in translating Amharic and English language pair, though none have been entirely successful. M. G. Teshome & Besacier (2012a) reported a BLEU score of 35.32% using the statistical machine translation (SMT) method, which increased to 37.53% when phonemic transcription was used (M. Teshome et al., 2015). In 2020, Hadgu et al. (2020a) did an evaluation of the existing Amharic machine translation (MT) systems to test their quality. While the study confirmed the expectations of potential in these systems, they also reported their low BLEU scores. From a different angle, all these studies highlight the opportunity to improve the performance of English-Amharic translation using SMT and phonemic transcription techniques.

Ashengo et al. (2021b) analyzed the performance of Content-Based Machine Translation (CBMT) and Recurrent Neural Network Machine Translation (RNNMT) working in conjunction for English-to-Amharic translation to understand how effective this hybrid model would be. They explained how this hybrid method outperformed plain neural machine translation (NMT) while exploring how dictionaries impact translation quality. The results indicated that the combination of CBMT and RNNMT yielded better performances on English-to-Amharic translation, especially with larger datasets like the New Testament Bible. Furthermore, the accuracy of the dictionary used by CBMT certainly has a major impact on the system's performance.

Destaw Belay et al. (2022) claimed that normalizing Amharic homophone characters can have a substantial positive effect on machine translation performance between Amharic and English in both directions. The particular focus of their study was an extensive parallel corpus of Amharic-English sentences and sought to explore the effects of the normalization of Amharic homophones on machine translation, with special emphasis on improving Amharic-to-English translation. Their results showed that normalizing homophonic characters in Amharic improves the performance of machine translation. The results also indicate that the performance of the M2M-100 model exceeds that of other transformers, and building in homophone normalization increases the efficacy of the NMT system.

Regarding the field of neural machine translation (NMT), the addition of some background information is known to improve translation quality and is regarded as an important addition. It helps even more with problems on disambiguation involving both lexical and syntactic ambiguity (Moussallem et al., 2018). Coupled with best choice identification for some sentences and the use of translation memory, the similarity-aware NMT integration has also alleviated human translator workload (J. Zhang & Zong, 2020). Moreover, recent studies regarding the role of context in NMT have pointed out that the use of a broader unstructured context scope is crucial for improving translation quality, as noted in Popescu-Belis (2019).

There is increasing curiosity about the use of graph neural networks (GNNs) in combination with natural language processing (NLP) over the last few years as a result of their ability to model relationships and structures associated with language data. This new area of research has been advanced by scholars like Wu et al. (2023) and Liu & Wu (2022) whose surveys provide an overview of the impact of GNNs on the different subdomains of NLP and detail the other various aspects that these disciplines include. Their surveys provide deep analyses of several areas of NLP research, starting from sentiment analysis to machine translation, and analyzing the incorporation of GNNs into those fields. They also give a summary of benchmarks and overall metrics used for evaluating GNN models. This makes it easier to understand the performance capabilities.

Particularly, Wu et al. (2023) contributed to the discussion by suggesting a type of taxonomy of GNNs for NLP, which divided the area of research into the construction of the graph, learning the representation of the graph, and graph-based encoder-decoder models. This categorization helps to make sense of the different approaches and practices that exist in GNN, NLP research. In combination, this body of research illustrates the increasing attention given to GNNs in the field of NLP, as well as highlighting the remaining issues that need to be researched, including how to improve scalability, interpretability, and generalization for different language tasks and languages.

Graph convolutional networks (GCNs) are one of the more promising methods that utilize the predicted syntactic dependency parse trees of the source sentences to form representations or hidden states of the encoder that are syntactically informed neighborhoods. GCNs are easily added to encoders such as bidirectional RNNs and convolutional neural networks as layers. An evaluation by English-to-German and English-to-Czech translators showed that the addition of syntax significantly outperformed all setups over the syntax-agnostic versions. This strategy emphasizes the opportunity introduced by using syntax-aware features for the enhancement of neural machine translation models (Bastings et al., 2017).

Table 2.19: Summary of Related Literature

No	**Authors & Year**	**Title & Journal Name/Conference Name**	**Major Findings / Contributions & Conclusion**	**Strength And limitation of the Articles**

1	M. Teshome et al. (2015)	Phoneme-based English-Amharic Statistical Machine Translation and IEEE	BLEU score results for the phoneme-based English-Amharic SMT system is scores 37.53%.	They use the syllable of Amharic language to increase. Not awarding the language structure very well and ignoring the conjunction terms.
2	A. M. Gezmu et al. (2021)	Neural Machine Translation for Amharic-English Translation and ICAART	By adjusting the hyperparameters for low-data scenarios, they applied the transformer-based neural machine translation architecture.	They have worked on orthography and morphology of Amharic words. But this language needs to focus on its structure rather than orthography and morphology.
3	Gashaw & Shashirekha (2019)	Amharic-Arabic Neural Machine Translation and arXiv	Short-Term Memory (LSTM) and Gated Recurrent Units (GRU) based Neural Machine Translation (NMT) models are developed using Attention-based Encoder-Decoder architecture and LSTM based OpenNMT in BLEU registered the 12%.	Due to the small size of corpus and not considering the structure of Amharic language the performance of the model is registered low.
4	Destaw Belay et al. (2022)	The Effect of Normalization for Bidirectional Amharic-English Neural Machine Translation and IEEE	They build bidirectional Amharic-English translation models by finetuning the existing Facebook M2M100 pre-trained model and achieved a BLEU score of 37.79 in Amharic-English translation.	Due to normalizing the homophonic letter of Amharic, they have got a better performing model. But they have not seen the structure of the language and its effect.

Chapter Three: Research Methodology

3.1 Introduction

This chapter outlines the research methods used to build the English-Amharic machine translation model. It describes the step-by-step processes from the beginning phases of acquiring data through to the final phases of the model's execution and assessment. Each given step in the pipeline has been meticulously tailored to reinforce the translation system in terms of reliability, precision, and operational competency.

We first describe the bilingual corpus construction, the most primary resource for the model training and assessment, and the data collection process. After that, the chapter discusses the various methods used to preprocess the data and the diverse techniques of cleaning, normalizing, and structuring data to improve the model's performance. Next, we focus on the English language syntactic analysis, where we build syntactic frameworks that capture the grammatical connections, within and across, the sentences. Such syntactic data are subsequently converted to graphs that capture the model's input within a definitional space and are used in embedding processes. These graph embeddings deepen the model's comprehension of the structure of the source sentence.

This chapter continues to focus on the fundamentals of model architecture and explains the graph encoder and how it serves to retrieve features and contextual relations of the embeddings. Lastly, the decoder of the transformer is used to produce the goal Amharic sentences by applying attention and the created representations for the purpose of efficient and proficient translations. This chapter highlights the extensive detailing of these aspects and how it offers a clear vision of the methodological framework serves as the steppingstone for the development of the English-Amharic machine translation system, illustrating the theory as well as the practices used to conduct the research.

3.2Research Design

This study examines the different conditions under which the English–Amharic machine translation model operates. This is why experimental research design is used. In experimental research, dependent variables are manipulated to determine the extent to which one or more independent variables are influenced. A cause-and-effect relationship can be determined to a high degree through research. As opposed to other designs, this design was defined to evaluate the effects of configuration and model parameters on translation quality because of the configured experimental translation parameters.

A focused study enhances model performance parameters. In a managed environment, external factors are minimized or controlled. In this way, you can be sure that any change in the result is caused by the parameters you have changed. In this study, the independent variables are hyperparameters and machine translation model architecture, while the dependent variables are translation performance metrics like METEOR and BLEU, as well as semantic and grammatical accuracy.

Iterative experimentation involved conducting multiple runs, each with a different parameter variation, to observe the impact of each parameter. This process enabled model tuning and configuration optimization. The design also maintained the required scientific rigor, as the model was built and evaluated with validation sets, baseline comparison, and reproducible metrics to ensure the results are trustworthy and valid.

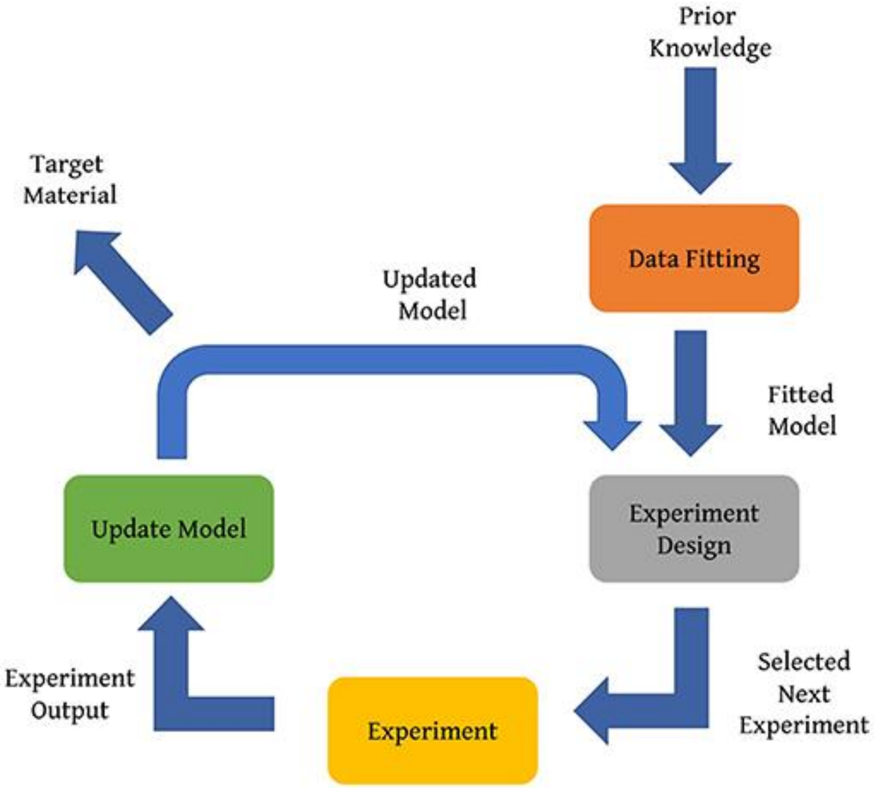


Figure 3.1: Iterative experiment design process (Talapatra et al. (2019)).

The justification for utilizing this design approach is the possibility of identifying causal relationships. Through parameter adjustments, model layers, and the addition of syntactic structures, this work attempts to comprehend the nuances that improve the accuracy of translation. This approach is also flexible and customizable, as multiple hypotheses with minimal constraints can be investigated around the behavioral and training structures of the model, which is important in the challenging area of neural machine translation. It also captures theoretical constructions about model performance empirically. To conclude, the outlined research design allows for a systematic and in-depth study of the English-Amharic machine translation model exploring its potential. It also allows the study to identify and examine performance determinants and offer suggestions for the enhancement of machine translation practice. It thus addresses both the practice and the theory of machine translation.

3.3System Architecture

The proposed Graph2Seq system integrates a graph-based syntactic encoder with a Transformer-style decoder to perform English-to-Amharic machine translation. By explicitly modeling syntactic structure in the source language, the architecture enhances both translation accuracy and fluency. Figure 3.2 illustrates the overall workflow of the system, which comprises interconnected modules spanning data preprocessing, graph construction, encoding, and sequence generation.

The translation pipeline begins with a parallel corpus consisting of aligned English source sentences and Amharic target sentences, which serves as the foundation for training and evaluation. Prior to model ingestion, the raw text undergoes a series of preprocessing steps to improve data quality and consistency. These steps include text cleaning, such as removing noise (e.g., HTML tags, special characters, and misaligned sentence pairs), and normalization. English text is standardized through lowercasing, while Amharic text undergoes normalization to resolve orthographic variations among phonologically equivalent characters and to ensure consistent script representation.

Following normalization, the system tokenizes each sentence into word-level and subword-level units and subsequently performs vectorization by mapping tokens to numerical indices within a fixed vocabulary. These representations enable efficient processing within the neural architecture while supporting open-vocabulary translation.

Next, the system applies syntactic dependency parsing to each English source sentence to generate a dependency tree that encodes grammatical relationships such as subject, object, and modifier. This tree captures the hierarchical structure of the sentence beyond its linear word order. The dependency tree is then transformed into a graph representation, where words correspond to nodes and syntactic relations correspond to edges. The graph structure enables the explicit modeling of relational dependencies that are difficult to capture using sequence-based encoders alone.

Each node and edge in the graph is embedded into a continuous vector space. Node embeddings encode lexical and semantic information, while edge embeddings represent syntactic relations. This joint embedding strategy allows the model to simultaneously encode semantic content and structural information. The resulting embedded graph is processed by a multi-head graph attention mechanism, which enables the encoder to attend to multiple relational substructures in parallel and to capture long-range dependencies across the sentence graph.

The attention outputs pass through position-wise feed-forward neural networks that apply non-linear transformations to refine node representations. The encoder ultimately produces enriched contextual embeddings that integrate syntactic structure with semantic meaning. These embeddings form the input to the Transformer decoder.

The decoder follows the standard Transformer architecture and generates the Amharic translation autoregressively. At each decoding step, the decoder performs multi-head attention over the encoded graph representations to retrieve relevant source-side information. The attended representations are then processed through feed-forward layers to construct context-aware target embeddings. A linear projection maps these embeddings to logits over the target vocabulary, and a softmax function produces a probability distribution over possible next token.

During inference, the decoder selects tokens sequentially from the predicted distributions until it generates an end-of-sentence token, thereby producing a complete Amharic translation. By incorporating syntactic dependency graphs into the encoding process, the Graph2Seq architecture enables the decoder to better model word order variations, grammatical agreement, and long-distance dependencies between English and Amharic.

Overall, the modular design of the system from preprocessing and syntactic graph construction to graph-aware encoding and Transformer-based decoding ensures a seamless integration of prior linguistic knowledge with neural sequence modeling. This pipeline substantially improves translation coherence, grammatical correctness, and fluency for English-to-Amharic machine translation.

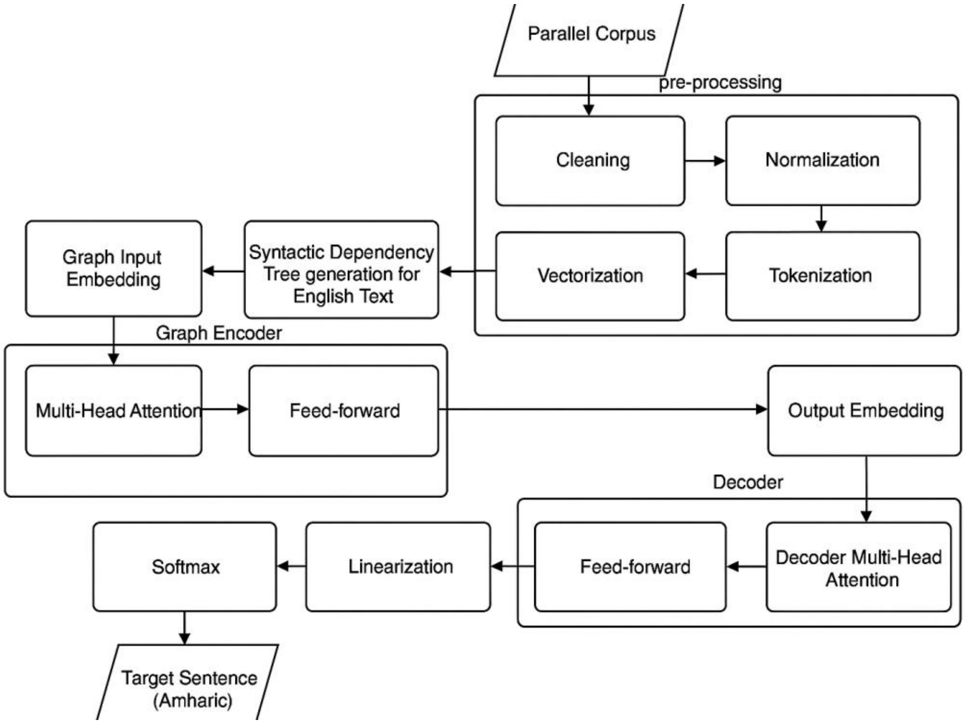


Figure 3.2: Proposed system architecture of English-Amharic machine translation.

3.4Data Collection and Corpus Preparation

The dataset used for training and evaluating English-to-Amharic machine translation is a parallel corpus compiled by Destaw Belay et al. (2022). This corpus is larger than previous ones and is freely available for research purposes. It was used to train neural machine translation models. Importantly, the neural machine translation models, especially those using subword units, achieved the highest BLEU scores, demonstrating their exceptional performance in this task ((A. M. Gezmu et al., 2022).

Table 3.1 illustrates that Biadgligne et al. (2023) compiled the most extensive parallel corpus for English-to-Amharic language pairs. In addition to the datasets outlined in Table 3.1, Destaw Belay et al. (2022) 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 Corporation, 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.

Table 3.1: Available Amharic and English parallel data.

Data source	**# Sentence pairs**	**Accessible**
Am-En ELRA-W0074	13,347	yes
Biadgligne & Smaïli (2021)	225,304	yes
Horn MT	2,030	yes
Am-En MT corpus	53,312	yes
A. M. Gezmu et al. (2021)	145,364	yes
S. T. Abate et al. (2019)	40,726	yes
Lison & Tiedemann (2016)	562,141	yes
Tracey & Strassel (2020)	60,884	no
Admasethiopia	153	yes
MT Evaluation Dataset	2,914	yes
Destaw Belay et al. (2022)	33,955	yes
Total	**1,140,130**	**yes**
Unique sentence pairs	**888,837**	**yes**

3.4.1 Corpus Cleaning and Filtering

The bilingual corpus used in this study contained substantial noise, inconsistencies, and sentence misalignments that adversely affected model performance. To address these issues, we applied a rigorous cleaning and filtering pipeline to ensure data quality, consistency, and semantic reliability prior to training.

First, we removed exact duplicate sentence pairs to eliminate redundancy and prevent the model from overfitting to repeated examples. This step ensured that the corpus contained diverse and informative translation instances rather than multiple occurrences of identical data. Next, we verified sentence-level alignment to confirm that each English source sentence accurately corresponded to its Amharic translation. We identified and removed misaligned pairs that failed to convey equivalent meaning by employing a hybrid approach that combined automated alignment verification with targeted manual inspection. This strategy balanced scalability with accuracy and reduced the propagation of semantic noise into the training process.

We then applied language identification techniques to ensure that each sentence appeared in its intended language. This step filtered out instances of code-switching, mixed-language content, and off-topic material that could confuse the translation model and degrade performance. Finally, we performed noise filtering to remove low-quality sentences, including those containing excessive special characters, HTML tags, or incomplete and ill-formed phrases. We also discarded sentences that fell outside predefined length thresholds, as extremely short or excessively long sentences often provide limited linguistic value or introduce alignment ambiguity.

Through this systematic cleaning and filtering process, we refined the bilingual corpus into a collection of high-quality, semantically aligned, and linguistically coherent sentence pairs. This curated dataset provides a robust foundation for training reliable and high-performing English–Amharic machine translation models.

3.4.2 Normalization

Following corpus cleaning and alignment, we applied additional normalization procedures to enhance data consistency, linguistic coherence, and model robustness. These normalization steps targeted both general textual inconsistencies and language-specific characteristics of English and Amharic.

At the corpus level, we standardized character encoding using Unicode Normalization Form to ensure consistent representation of characters across the dataset. We removed non-printable and control characters that could introduce noise during tokenization and model training. We also eliminated residual duplicate sentence pairs to maintain dataset neutrality and prevent biased learning caused by overrepresented examples.

At the token level, we applied normalization to reduce orthographic variability. For English text, this process included lowercasing, standardizing punctuation (e.g., converting curly quotation marks to straight quotes), and removing redundant whitespace. Lowercasing reduces vocabulary sparsity by collapsing capitalization variants (e.g., *Head* vs. *head*) into a single form. This step decreases vocabulary size, improves memory efficiency, accelerates model convergence, and promotes better generalization by preventing the model from learning redundant representations for lexically identical words. Lowercasing also mitigates out-of-vocabulary issues by increasing the likelihood that rare or unseen words share representations with observed forms.

In contrast, we did not apply lowercasing to Amharic text, as the Geʿez script does not encode capitalization. Instead, we performed language-specific normalization tailored to Amharic's phonetic and orthographic properties. Specifically, we unified characters that share identical phonetic realizations but differ orthographically by mapping them to a single canonical form. For example, characters such as ሀ and ሐ, which represent the same phoneme, were normalized to a unified representation. This process reduces orthographic ambiguity, limits vocabulary inflation, and improves alignment consistency between parallel sentences.

Additionally, we applied Amharic-specific morphological preprocessing, including stemming or lemmatization where appropriate, to mitigate data sparsity caused by rich inflectional morphology. We further employed automatic language identification tools to detect and remove incorrectly labeled or linguistically inconsistent sentence pairs that survived earlier filtering stages.

Collectively, these normalization procedures improved the linguistic integrity, calibration precision, and adaptability of the dataset. By reducing orthographic variation, vocabulary fragmentation, and language inconsistencies, the normalization pipeline enhanced the effectiveness of downstream tokenization and model training, thereby contributing to more stable and accurate English–Amharic machine translation.

3.4.3 Corpus Statistics and Analysis

The dataset utilized in this study consists of parallel corpora in two languages: English and Amharic. The dataset is divided into three subsets commonly used in machine learning and natural language processing tasks: Training (Train), Testing (Test) and Validation (Val) sets. Table 3.2 summarizes the vocabulary sizes for each language across the subsets.

Table 3.2: Dataset unique vocabulary size.

Subset	**English Vocabulary Size**	**Amharic Vocabulary Size**
Train	276,618	610,623
Test	139,800	351,020
Val	85,953	219,866

Vocabulary size refers to the number of unique tokens (words or sub words) present in each subset of the dataset. The vocabulary sizes indicate the diversity and richness of the textual data in each language. Amharic exhibits substantially larger vocabulary sizes across all subsets compared to English in this dataset. Training set of English language contains approximately 276,618 unique tokens. Validation and test sets contain 85,953 and 139,800 unique tokens, respectively. The relatively smaller vocabulary size compared to Amharic reflects the morphological characteristics of English, which tends to have less inflectional variation and fewer unique word forms. Whereas the training set Amharic language shows a large vocabulary size of 610,623 unique tokens, more than twice the size of the English training vocabulary. Validation and test sets also demonstrate large vocabulary with 219,866 and 351,020 unique tokens respectively. This larger vocabulary size can be attributed to the highly inflectional and morphologically rich nature of Amharic, which results in many word forms derived from roots and affixes. The larger vocabulary size indicates increased linguistic complexity and diversity in the Amharic text data.

The training set contains the largest vocabulary in both languages, as expected, since it encompasses the most data and is used to learn model parameters. Whereas the validation set has the smallest vocabulary, which helps in tuning model hyperparameters and preventing overfitting. And the test set vocabulary size is intermediate, used for evaluating final model performance on unseen data.

The disparity in vocabulary sizes presents challenges for building effective multilingual or translation models. Handling the large and morphologically complex Amharic vocabulary requires specialized tokenization techniques such as subworld segmentation like Byte Pair Encoding to reduce sparsity. The model must be capable of managing large vocabulary efficiently to avoid overfitting and ensure generalization. The difference in vocabulary sizes also affects model architecture choices, embedding layer sizes, and computational resource requirements.

The vocabulary statistics highlight the linguistic differences between English and Amharic and underscore the need for tailored preprocessing and modeling strategies to effectively leverage the dataset. The rich and extensive Amharic vocabulary reflects the complexity of the language and the challenges involved in natural language processing tasks involving Amharic text.

Word Frequency in English

Figure 3.3 presents the top 50 most frequent English words identified from the English side of the parallel corpus used in this study for English–Amharic machine translation. Like the Amharic side, stop words were removed before analysis to ensure that only semantically significant words were retained. The x-axis displays the English words in descending order of frequency, while the y-axis represents the number of times each word appears in the corpus. The distribution shows a strong right-skewed pattern typical of natural language text, where a few words occur very frequently, and the majority have lower frequencies and observation consistent with Zipf's law.

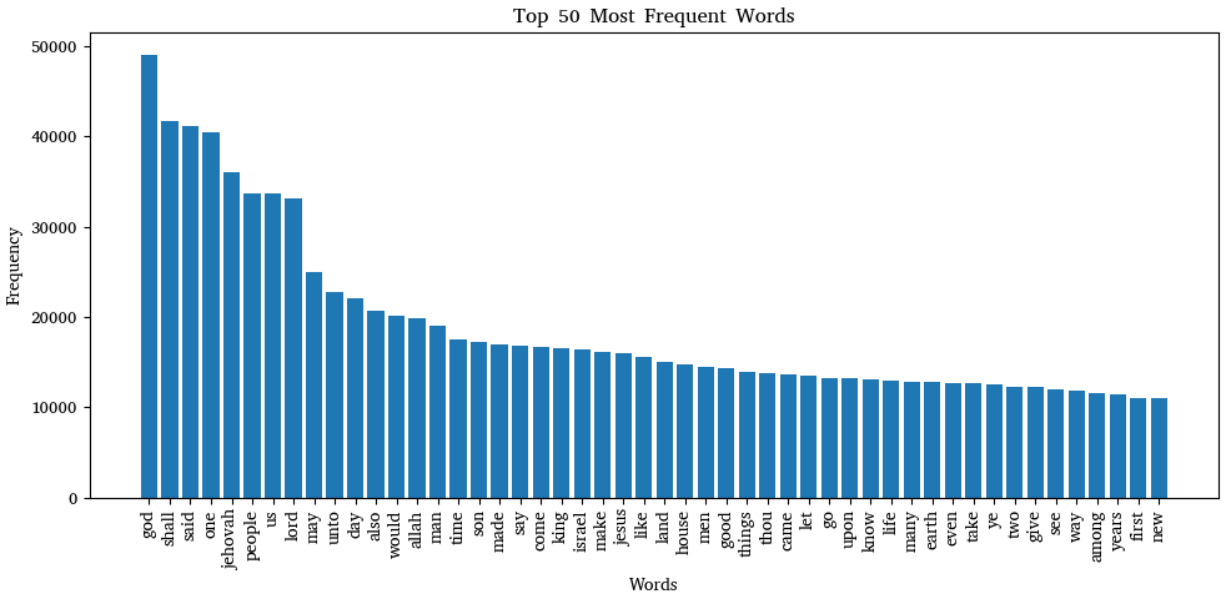


Figure 3.3: Top 50 most frequent English words.

The most frequent words include high-content lexical terms such as “god”, “said”, “lord”, “people” and “king” indicating the corpus's domain orientation toward religious or historical texts. The predominance of such words suggests that the corpus is rich in narrative and descriptive content, which could influence the stylistic and contextual patterns learned by the translation model.

Overall, this visualization highlights the lexical distribution of the English dataset after stop word filtering, providing insights into the corpus composition and the vocabulary base essential for effective English–Amharic neural machine translation.

Word Frequency in Amharic

The top 50 Amharic words most frequently used, taken from the Amharic side of an English-Amharic parallel corpus used in the machine translation this study, are presented in the figure 3.4. For analysis, stop words were removed so that the remaining words would be linguistically and semantically important, especially in translation. The x-axis shows the unique Amharic words, organized by frequency, and the y-axis shows the frequency of each word in the corpus. The words are arranged in order from most frequent to least frequent. The most frequent words are on the left side of the figure 3.4.

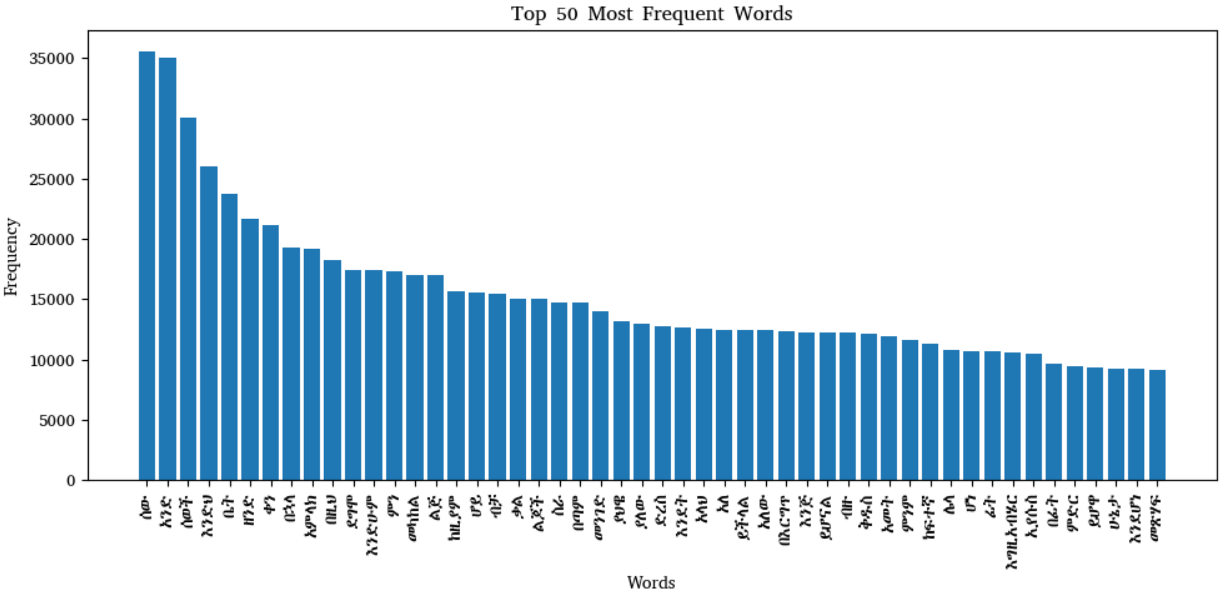


Figure 3.4: Top 50 most frequent Amharic words.

The skew in the frequency distribution implies that only a few words tend to be used, and most words are used infrequently, as described in Zipf's law. Within the corpus, the early, frequently used words reflect the basic structure, morphology, and syntax of Amharic text. To focus the analysis, stop word removal also eliminates high-frequency, semantically empty, words that would otherwise dominate the analysis, like "እኔ" (I), "እሱ" (he), and "ይህ" (this). As a result, the visualization captures high-frequency core lexical items that are critical in learning translation models. Overall, this visualization offers important information regarding the Amharic dataset's lexis and assists in the assessment and understanding of vocabulary richness and the dominance of certain words, in addition to the system's the English-Amharic neural machine translation system's possible imbalances in terms of words coverage.

Sentence Length in English

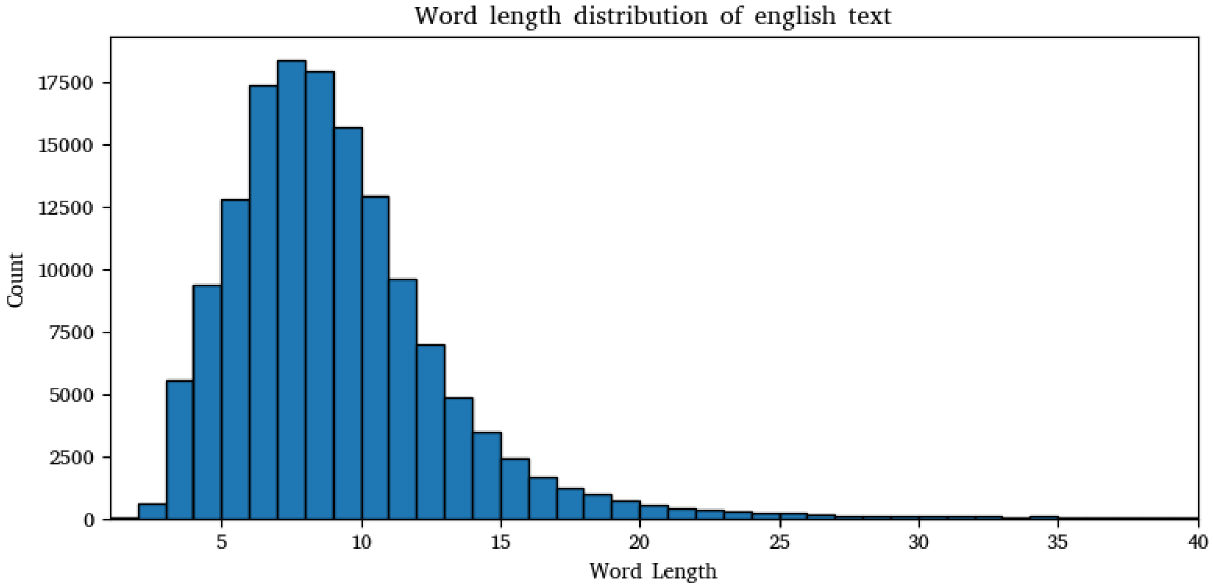


Figure 3.5: Sentence length in English text.

Figure 3.5 describes the distribution of sentence lengths in the corpus. The X-axis displays the length of the sentence in number of words, and the Y-axis shows the frequency of sentences of that length. Figure 3.5 shows the number of sentences that fall into certain length ranges to describe an idea of the type of corpus we are working with.

Sentence lengths range from very short ones of one to two words, to longer ones of about forty words. Sentences longer than thirty words, however, are infrequent. The length of the sentence that occurs most frequently is seven words, with nearly 18,000 occurrences, confirming that the corpus is made up of short sentences. The distribution is right-skewed, that is, there are a lot of short sentences and fewer longer ones. The sentence counts increase dramatically from one to about seven words, and then slowly decrease. Most sentences are of four to twelve words, which is the average sentence length in this corpus in the bilingual corpus.

The predominance of short to medium-length sentences most likely points to the fact that the dataset is made up of simple, more direct constructions, intended to make machine translation easier and effective. Using short sentences works to our advantage because they are more likely crossover between different languages, thus improving the quality of bilingual parallel corpora. On the other hand, the less common longer sentences may be more complex, and therefore, there are more gaps and errors that could come into the translation and alignment.

Considering the practical aspects, this length distribution carries certain consequences regarding machine translation. When training the model, working with sentences that are between four and twelve words will capture most of the available data. For preprocessing, we could enhance translation quality by segmenting or simplifying sentences longer than twenty words. For evaluation, knowing the overall distribution of sentence length aids in ensuring that we evaluate translation models with sentence lengths that are comparable to the training data. In neural MT systems, having and employing certain sentence length norms are significant in establishing maximum sequence lengths and padding techniques during the tokenization process.

Sentence length in Amharic

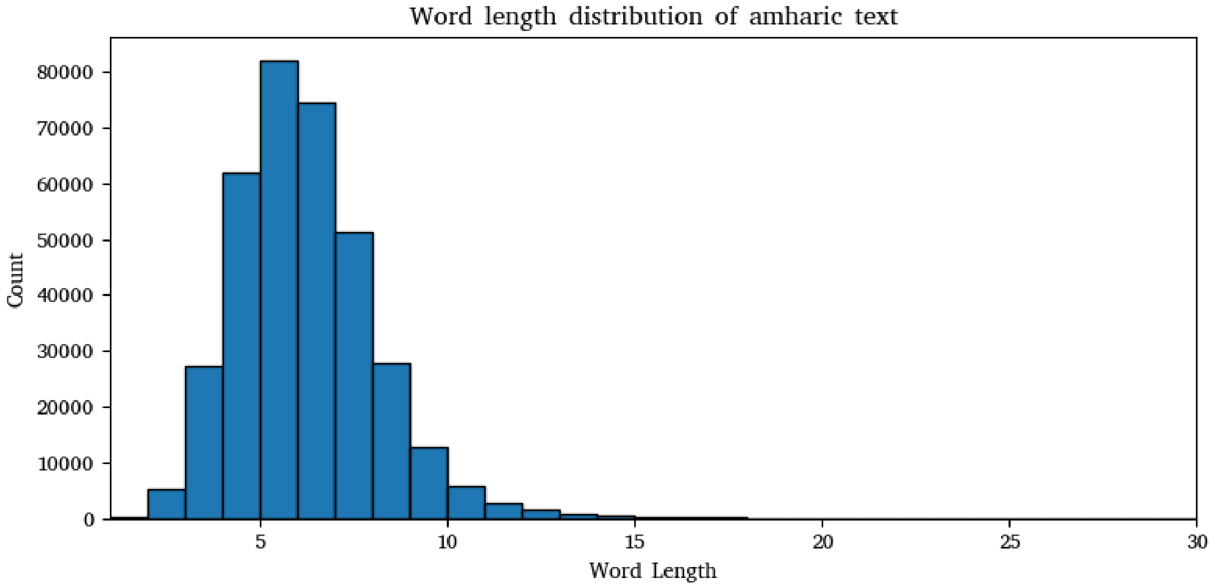


Figure 3.6: Sentence length in Amharic text.

Figure 3.6 clarifies how sentences vary in length in the Amharic section of the machine translation corpus. Here, the X-axis is the sentence length in the Amharic text in number of words, whereas

Y-axis is the frequency of sentences for each length. Figure 3.6 depicts the range of sentence lengths for the Amharic corpus, providing insights into the compositional structure of the corpus. Sentences in the corpus are of varying lengths, and estimates suggest that they range from one to thirty words. However, longer sentences (more than fifteen words) are rare. The predominant sentence length is six words, and there are approximately 82,000 sentences of this length. This suggests that shorter sentences are the most predominant in the Amharic corpus, pointing to a tendency toward greater conciseness in the sentence structure.

The overall distribution shows a right skew, indicating that there are a greater number of short sentences and that longer sentences are progressively fewer. Additionally, there is a rapid increase in sentence count with lengths between one to six words, and then a gradual decrease. The range of four to eight words constitutes most sentences, and very short (one to three words) and longer (more than ten words) sentences are considerably less frequent.

The average sentence length of six words suggests and is in line with corpus linguistics and linguistics predictions that Amharic text has a greater tendency to dense and concise constructions. Decisions made during design of the corpus might have aimed to facilitate alignment of sentences for machine translation. Filters that processed length sentences, or the syntax and morphology of Amharic, could explain why longer sentences are relatively uncommon.

There are practical consequences that this distribution of sentence length poses for machine translation. Since most of the data contained in the corpus is within the four-to-eight-word sentence length, model training should focus on this length. Long sentences could be subjected to more aggressive splitting or segmentation during the preprocessing stage to improve translation quality. Evaluation should consider the model's performance on the frequent shorter sentences, and on the infrequent longer sentences. Furthermore, the sentence length distribution observed here could be used to refine padding and batching strategies in neural machine translation for greater efficiency and more accurate translation.

Table 3.3: Comparison of English and Amharic Sentence Length Distributions.

Feature	**English Text**	**Amharic Text**
Most frequent sentence length	7 words	6 words
Peak sentence count	~18,000 sentences	~82,000 sentences
Sentence length range	1 to ~40 words	1 to ~30 words
Distribution shape	Right skewed	Right skewed

Concentration range	4 to 12 words	4 to 8 words

3.4.4 Dependency Parsing

Parsing is an essential step in NLP that involves analyzing the syntactic structure of sentences to determine the relationships between words. This structural information aids in understanding the semantics and context of the sentence. Dependency parsing models the syntactic structure of a sentence as a directed graph where individual words are nodes, and directed edges denote syntactic relations (e.g., subject, object, modifier). In the sentence” I booked a ticket to Adama,” the verb “booked” functions as the root node connected to its subject “I,” direct object “ticket,” and the prepositional phrase “to Adama”.

This study utilized the StanfordNLP toolkit for dependency parsing, a widely recognized and reliable tool that produces labeled syntactic dependency graphs. Each edge in the graph is assigned a label corresponding to the type of syntactic relation, providing rich contextual information.

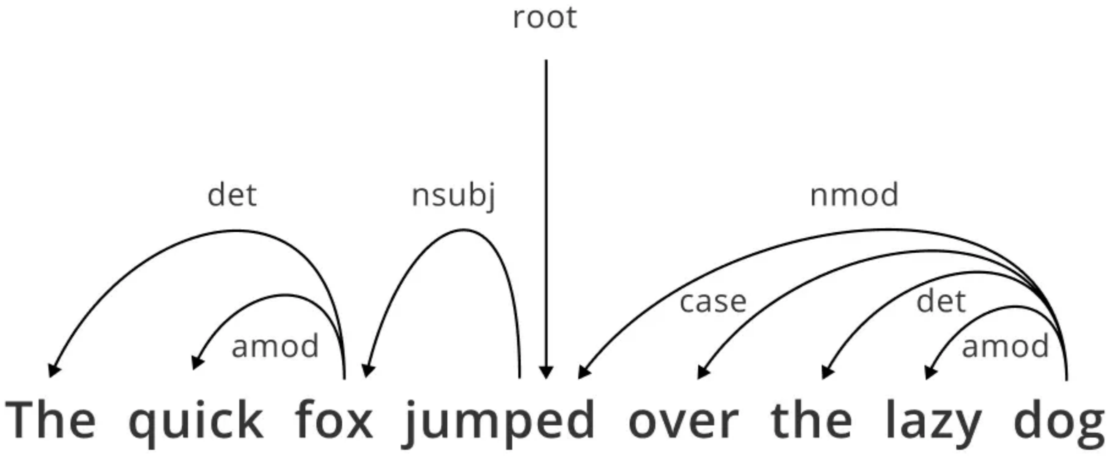


Figure 3.7: Syntactic dependency tree for the sentence “The quick fox jumped over the lazy dog”.

3.4.5 Graph Representation of Dependency Trees

A graph G= (V, E) consists of a set of vertices V (or nodes) and edges E that connect pairs of vertices. In the context of dependency parsing, nodes represent words, while edges represent grammatical relations. Each sentence is converted into a graph where words correspond to nodes and syntactic dependencies to directed edges. This transformation allows the model to capture the multi-relational and non-linear structure of language beyond sequence order. Each node is associated with a feature vector, typically a word embedding, that encapsulates semantic and syntactic information. Edges may carry labels identifying the dependency relation type, adding another layer of information for encoding.

Representing sentences as graphs preserves structural dependencies and hierarchical relationships, enabling models to better capture context and syntactic nuances than linear sequence representations.

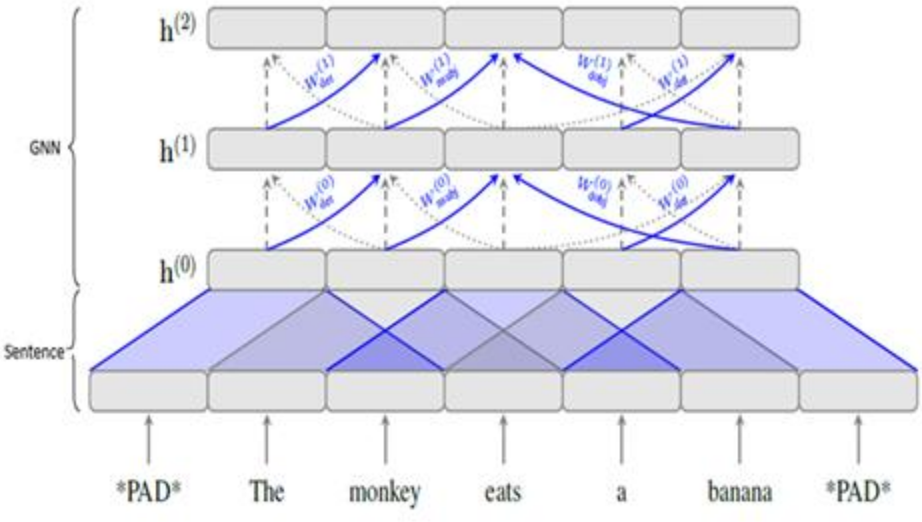


Figure 3.8: Two layers GNN representation of dependency tree.

3.5Graph Neural Networks

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 (Bignoli, 2021). 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 can encode 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 (D. Lin et al., 2022).

The node embeddings produced by GNNs contain informative parts that can be processed and encoded in the form of walls, word relations, or any feature that 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 make it possible to accurately contextualize relevant sequences during decoding (G. Tan, 2025).

It is possible to learn from graphs in ways knowledge graphs, by using graph attention networks (GATs). GATs use an attention mechanism to assign different weights to nodes and edges of a graph based on their relevance and importance (L. H. B. Nguyen et al., 2020). Our proposed machine translation model uses GAT, which improves translation quality and captures syntactical information from the source language. Considering the syntax dependency tree, we have a set of node features in the input layer.

Layer input:

$$h \ = \{h_1, h_2, \dots, h_n\}, h_i \ \in \ R^d$$

where:

- h is the input to the layer,
- *n* is the number of nodes,
- *d* is the number of features for each node.

Layer output:

$$h' = \{h'_1, h'_2, \dots, h'_N\}, h'_I \ \in \ Rd'$$

where:

- h′ is the output of the layer,

- N is the number of nodes,
- d' is the dimensionality of the new feature representation for each node.

This transformation is computed using multi-head attention in all intermediate layers for each node.

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

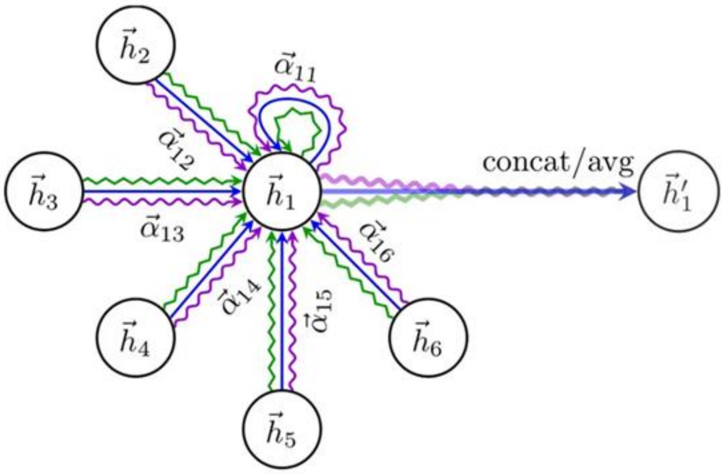


Figure 3.9: An illustration of multihead attention (with K = 3 heads) by node 1 on its neighborhood.

3.6Baseline GNN Encoders

3.6.1 Graph Convolutional Network Encoders

Recent advances in neural machine translation (NMT) have highlighted the effectiveness of Graph Convolutional Networks (GCNs) in modeling syntactic and semantic structures that are not adequately captured by purely sequential architecture. By operating directly on graph-structured representations, GCNs enable the integration of linguistic dependencies into neural encoders, thereby enriching contextual representations of source sentences.

One of the earliest and most influential studies in this direction was conducted by Bastings et al. (2017), who introduced dependency-based GCNs for word-level encoding in NMT. Their approach leveraged predicted syntactic dependency trees to construct graph representations of source sentences, which were subsequently processed using GCN layers on top of BiRNN- and CNN-based encoders. Experiments on English-German and English-Czech translation tasks demonstrated significant improvements in translation quality, providing early empirical evidence that syntactic structure enhances neural translation models.

Building on this foundation, Marcheggiani et al. (2018) extended GCN-enhanced NMT by incorporating semantic-role information derived from PropBank-style annotations. By integrating semantic relations into BiRNN- and CNN-based GCN encoders, their model captured both syntactic dependencies and predicate–argument structures. Evaluations on the WMT16 English-German dataset yielded consistent BLEU score improvements, underscoring the complementary role of semantic information in improving translation performance.

From a different linguistic perspective, Hao et al. (2022) proposed a graph-to-sequence framework for Mongolian-Chinese translation using Densely Connected Graph Convolutional Networks (D-GCNs), inspired by DenseNet architectures. To address the scarcity of high-quality syntactic resources, they constructed an auxiliary Mongolian dependency tree corpus, enabling richer syntactic representations. Their model outperformed conventional NMT baselines, demonstrating the effectiveness of densely connected graph architectures for low-resource and morphologically complex languages.

Extending beyond sentence-level modeling, M. Xu et al. (2021) introduced a document-level graph-based NMT approach that integrates GCNs with Transformer architectures. Their method represents entire documents as graphs that encode syntactic dependencies, clause adjacency, and cross-sentential relationships within paragraphs. By capturing both local and global contextual dependencies, the model achieved substantial improvements across multiple evaluation metrics, highlighting the importance of discourse-level structure in translation.

To sum up, these studies demonstrate that GCN-based encoders consistently improve NMT performance by incorporating syntactic structure and broader contextual information. Motivated by these findings, we adopt GCN-based encoders as a strong baseline for evaluating the effectiveness of the proposed syntax-aware translation framework.

3.6.2 Gated Graph Neural Networks Encoder

Gated Graph Neural Networks (GGNNs) are a type of neural network designed to handle graph structured data. They extend the capabilities of traditional Graph Neural Networks (GNNs) by incorporating gating mechanisms, which help control the flow of information and improve the learning process.

As a subset of Gated Graph Neural Networks (GGNNs), GGNNs are quite useful in handling structured data for neural machine translation (NMT). It is notable that some scholars of machine translation employ GGNNs. Still, many utilize graph-based frameworks that embody GGNN ideas specifically, iterative message passing along with gating mechanisms to preserve long-range dependencies as well as structural information. For example, the Multi-level Community awareness Graph Neural Network (MC-GNN) developed by Vo et al. (2024) replaces self-attention in Transformers with GNN-based layers, thus improving the modeling of both local and global interdependencies between words while overcoming semantic sparsity.

These trends illustrate the use of encoders in a sequence that is being improved upon by graph-structured learning, in which methodologies like GGNNs contextually improve GGNN and translation leverage structural linguistic cues. GGNNs have demonstrated their ability to mature and model complex sentence dependencies, providing valuable sentence tools for fostering NMT or low-resource languages.

3.7Syntactic Encoder

This study integrates Graph Neural Networks (GNNs) into the translation architecture to explicitly incorporate source-language syntactic structure into the encoding process. The syntactic encoder models each source sentence as a dependency graph, enabling the translation system to exploit grammatical relationships among words rather than relying solely on linear word sequences.

Specifically, the system applies syntactic dependency parsing to English source sentences to derive dependency trees that represent hierarchical grammatical relations, such as head-dependent structures and modifier attachments. These dependency trees define syntactic neighborhoods in which each word’s representation becomes sensitive to its surrounding grammatical context (G.

Tan, 2025). By representing sentences as dependency graphs, the model explicitly captures structural relationships among words as defined by the syntactic tree.

The GNN operates directly on word-level representations as both input and output. It refines these representations by propagating information across syntactic edges, thereby allowing each word embedding to encode not only its semantic content but also its grammatical role within the sentence. In practice, the GNN layers are stacked on top of conventional encoders such as bidirectional recurrent neural networks or convolutional neural networks following the framework proposed by Bastings et al. (2017). This layered design enables the model to retain the strengths of sequential encoders while enriching them with structure-aware representations.

By incorporating syntactic information through GNNs, the encoder gains access to linguistically meaningful dependencies without requiring rigid constraints on model architecture or extensive manual feature engineering. This flexibility allows the syntactic encoder to adapt to language-specific grammatical structures and to selectively emphasize syntax-relevant information while down-weighting less informative patterns (Gong et al., 2022).

In summary, the proposed syntactic encoder leverages GNNs to integrate dependency-based structural knowledge into the neural encoding process. This approach enhances the model's ability to capture grammatical relationships and long-distance dependencies in the source language, leading to improved contextual understanding and more accurate, fluent translations.

3.8 The Proposed Graph2Seq Machine Translation Model

The structure of the sequence-to-sequence model (encoder-decoder), which is widely used, is depicted in Figure 3.10. The goal of this study is to add prior knowledge to the encoder side. Seq2Seq models have many flexible and expressive renderings, but they suffer from a serious limitation, which is that 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 are a crucial aspect of many difficult problems, can be done using graphs (Bignoli, 2021). 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.

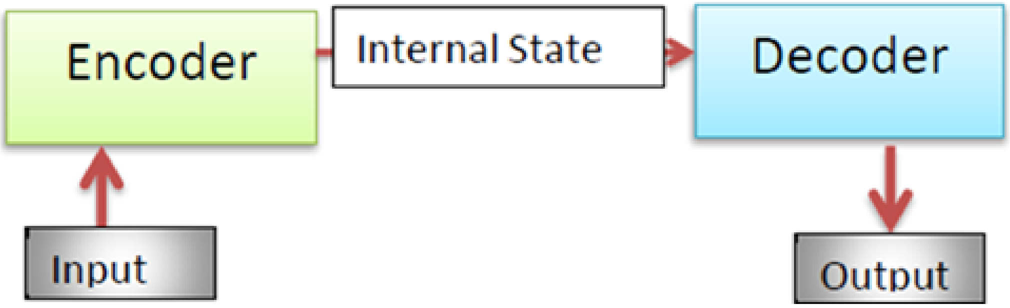


Figure 3.10: Encoder-decoder model.

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 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.

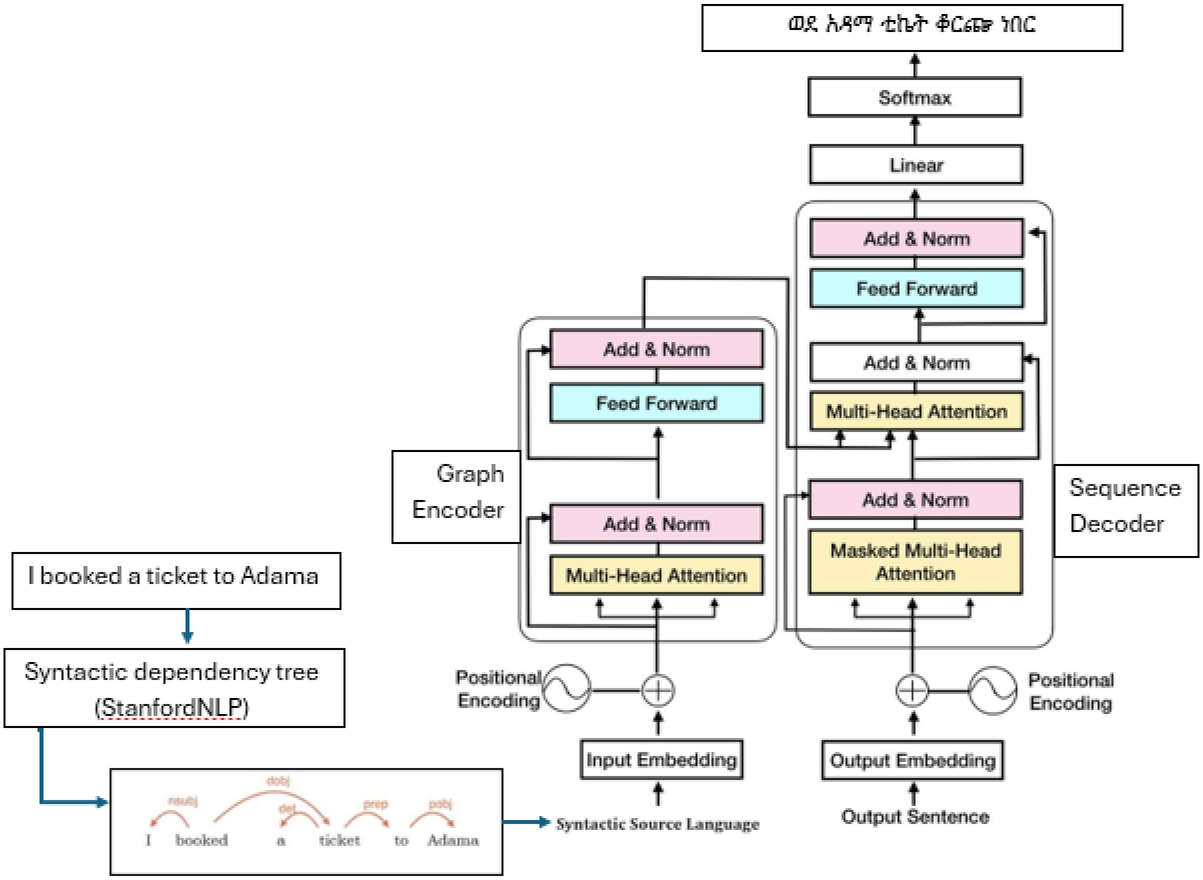


Figure 3.11: Transformer Encoder-Decoder architecture (source: Vaswani et al. (2017)); the proposed

Figure 3.11 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 undergoes 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:

$$\mathrm{C_x^{(l)}} = \mathrm{MultiHead}\left(\mathrm{H_x^{l-1}}, \mathrm{H_x^{l-1}}, \mathrm{H_x^{l-1}}\right), \quad (3.1)$$

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 position wise feedforward forward networks $FFN(\ M_x^l)$ to generate textual node states$H_x^{(l)}$

$$H_x^{(l)} = FFN(\ M_x^l), \quad (3.2)$$

where $M_x^{(l)} = [M_{xi}^{(x)}]$ 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 mask self-attention and encoder-decoder attention:

$$E^l = MultiHead(\ S^{l-1}, S^{l-1}, S^{l-1}), \quad (3.3)$$

$$T^l = MultiHead\left(\ E^l, H_x^{(Le)}, H_x^{(Le)}\right), \quad (3.4)$$

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:

$$S^{(l)} = FFN(S^{(l)}) \quad (3.5)$$

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:

$$P(Y\ |X) = \prod_t Softmax(WS_t^{Ld} + b) \quad (3.6)$$

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.

Chapter Four: Experiment Setup

4.1 Introduction

Using the representative datasets described in the preceding chapter, this study systematically evaluates the effectiveness of the proposed translation architectures. A series of carefully designed experiments assesses English-Amharic machine translation performance under different modeling assumptions, enabling a structured comparison of translation quality across architectures with varying degrees of linguistic awareness.

The experimental framework comprises three independent setups, each employing a distinct model architecture and training strategy. This design facilitates a comprehensive analysis of how architectural choices and linguistic integration influence translation accuracy, fluency, and overall quality.

The first experiment employs a standard Transformer model without incorporating any explicit syntactic information from the source language. This syntax-agnostic Transformer serves as a baseline, providing a controlled reference point against which the impact of subsequent syntax-aware approaches can be measured.

The second experiment fine-tunes the pretrained M2M100 multilingual model, which contains approximately 418 million parameters. Initially trained on large-scale multilingual data, M2M100 is subsequently adapted to the English-Amharic parallel corpus. Owing to its extensive cross-lingual pretraining, this model functions as a strong benchmark, benefiting from transfer learning across related language pairs and capturing broader multilingual representations.

The final experiment evaluates the proposed Graph2Seq framework, which explicitly integrates source-language syntactic dependency information into the translation process. By modeling sentences as syntactic graphs, this approach captures grammatical relations among words and encodes deeper structural dependencies. As a result, the Graph2Seq model aims to enhance translation accuracy, syntactic coherence, and fluency beyond what sequence-based models can achieve.

In summary, these experiments enable a comparative analysis of syntax-agnostic and syntax-aware translation paradigms, as well as an assessment of the impact of large-scale pretrained multilingual models. The findings elucidate how syntactic structure, dependency modeling, and knowledge transfer contribute to improved English–Amharic machine translation. Ultimately, the study identifies specific architectural components and linguistic enhancements that are critical for advancing translation quality in morphologically rich and low-resource language settings.

4.2 Dataset Preparation and Preprocessing

The construction and preprocessing of a high-quality parallel corpus constitute a critical foundation for developing an effective English–Amharic machine translation system. For this study, over 1.14 million aligned English–Amharic sentence pairs were collected from diverse, reliable, and publicly available sources, including web-based documents, established bilingual datasets, and NLP research repositories. The scale and diversity of this corpus provide substantial linguistic coverage, enabling the translation models to learn robust representations across varied domains, contexts, and linguistic expressions.

Despite its size, the raw corpus required extensive preprocessing to enhance data quality, coherence, and suitability for model training. The preprocessing pipeline comprises four principal stages: cleaning, normalization, tokenization, and vectorization. Each stage plays a crucial role in transforming noisy, unstructured text into a linguistically consistent and model-ready dataset.

During the cleaning stage, the corpus undergoes systematic filtering to address common issues in raw parallel data. This process removes exact duplicate sentence pairs to eliminate redundancy and potential bias, filters out misaligned or semantically inconsistent sentence pairs, and excludes incomplete entries. Additionally, non-linguistic artifacts such as HTML tags, extraneous symbols, and malformed characters are removed, ensuring that the remaining data consists of well-formed and meaningful linguistic content.

Normalization follows cleaning and aims to enforce consistency across the corpus. For English, all text is converted to lowercase to reduce vocabulary size and minimize sparsity. For Amharic, Unicode normalization is applied to ensure consistent representation of Ethiopic script characters, including the unification of orthographic variants that share identical phonetic realizations. Further normalization steps include standardizing punctuation, removing redundant whitespace, and addressing language-specific orthographic inconsistencies, thereby improving textual uniformity.

After normalization, the data proceeds to the tokenization stage, where sentences are segmented into smaller linguistic units. English text is tokenized using standard word-level tokenizers capable of handling contractions, punctuation, and compound forms. For Amharic, a custom tokenizer tailored to its script and rich morphological structure is employed. To further address vocabulary sparsity and out-of-vocabulary issues in both languages, subword segmentation techniques such as Byte Pair Encoding (BPE) or SentencePiece are applied, enhancing model generalization and robustness.

The final preprocessing stage involves vectorization, which converts tokenized text into numerical representations suitable for neural network processing. Separate vocabularies are constructed for English and Amharic, incorporating special tokens such as padding, unknown, start-of-sequence, and end-of-sequence markers. Each sentence is mapped to a sequence of token indices, and padding or truncation is applied to enforce a fixed maximum sequence length, enabling efficient batch-based training.

In general, these preprocessing steps yield a clean, consistent, and linguistically coherent parallel corpus. By systematically refining the input data, the study enables the translation models to learn more accurate and generalizable English–Amharic interlingual mappings, leading to improved translation performance and more reliable evaluation outcomes. Figure 4.1 illustrates the complete preprocessing pipeline employed in this study.

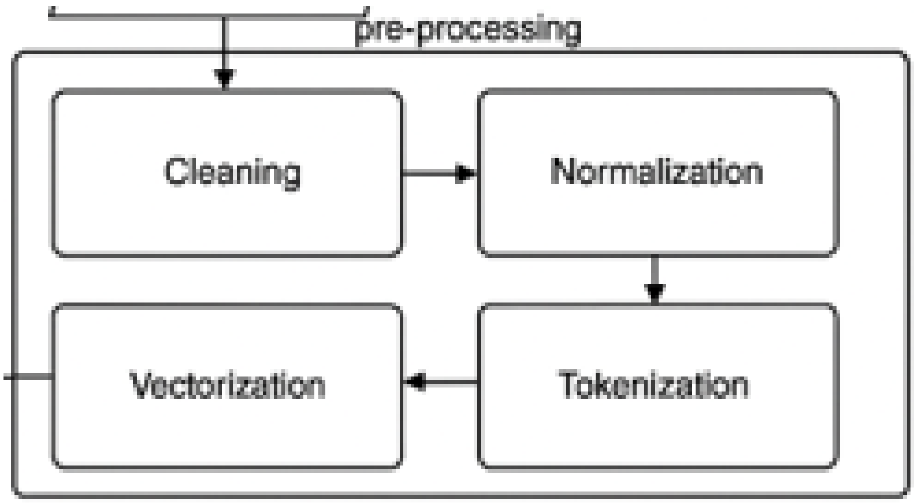


Figure 4.1: Steps for data preprocessing.

4.3 Experiment I: Transformer Model

4.3.1 Introduction

The Transformer model is an innovation in MT because of its self-attention mechanism. It outperformed recurrent and convolutional architectures because it captures long-range dependencies better. The model surpassed RNNs and LSTMs because it does not process sequentially and allows parallel processing of tokens. This greatly increases the efficiency of training and the quality of translations (Vaswani et al., 2017).

The self-attention mechanism and encoder-decoder structure create the most significant advancements in machine translation research. The improvements in deep transformer architectures like DLCL and novel techniques of normalization have increased translation quality and computational efficiency, as well as the ability to adapt to various languages simultaneously.

4.3.2 Transformer Model Architecture

Most state-of-the-art neural sequence transduction models utilize an encoder-decoder framework (Bahdanau et al., 2014). In this setup, the encoder transforms an input sequence of symbols $(x1, . . . , x_n)$ into a sequence of continuous vector representations $z = (z_1, . . . , z_n)$. The decoder then produces an output sequence $(y_1, . . . , y_m)$ symbol by symbol based on these encoded representations. During generations, the model operates in an auto-regressive manner (Vaswani et al., 2017), meaning it incorporates previously generated symbols as part of the input when predicting the next symbol. The Transformer architecture implements this general structure by stacking self-attention mechanisms and point-wise fully connected layers within both the encoder and decoder components, as illustrated in the left and right sections of Figure 4.2, respectively.

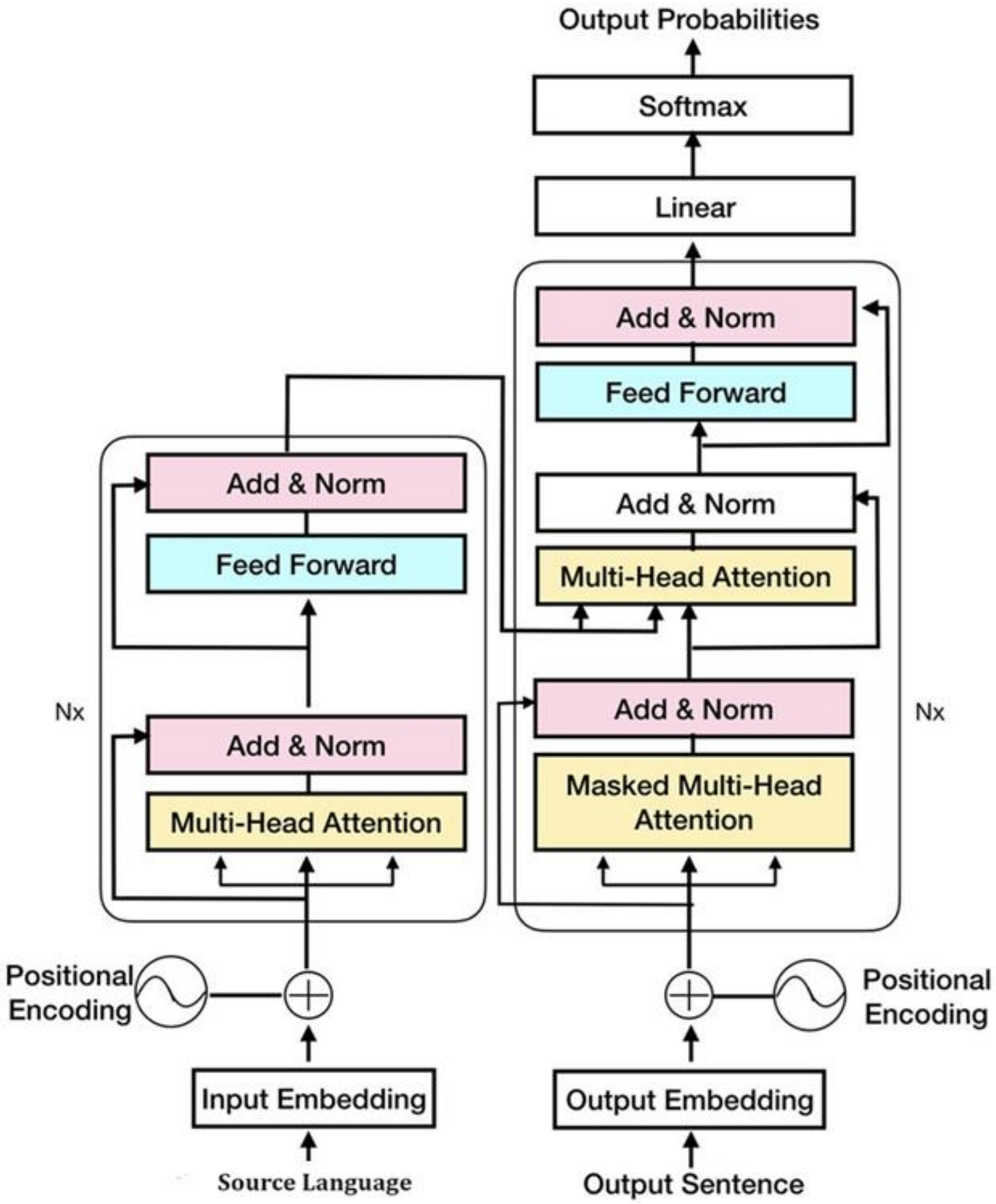
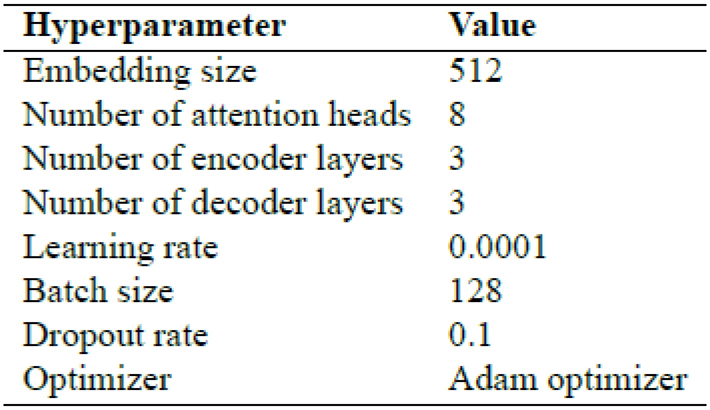


Figure 4.2: Transformer model architecture (Vaswani et al. (2017)).

4.3.3 Hyperparameters

Hyperparameters define the learning process of a machine translation model. Hyperparameters are set before training and differ from the model parameters (weights and biases) which are learned during training. Hyperparameters help shape how well the model performs.

Table 4.1: Hyperparameters used in the Transformer Model.



Hyperparameter	**Value**
Embedding size	512
Number of attention heads	8
Number of encoder layers	3
Number of decoder layers	3
Learning rate	0.0001
Batch size	128
Dropout rate	0.1
Optimizer	Adam optimizer

4.4 Experiment II: Pre-trained Model

4.4.1 Introduction

The integration of pre-trained models via fine-tuning has significantly enhanced the quality of translations in NMT. Leveraging pre-trained linguistic knowledge also improves the generalization and robustness of the model, derives faster convergence, and constitutes a step forward in realizing effective and high-quality robust neural machine translation systems.

4.4.2 Pre-trained Fine-Tune Model Architecture

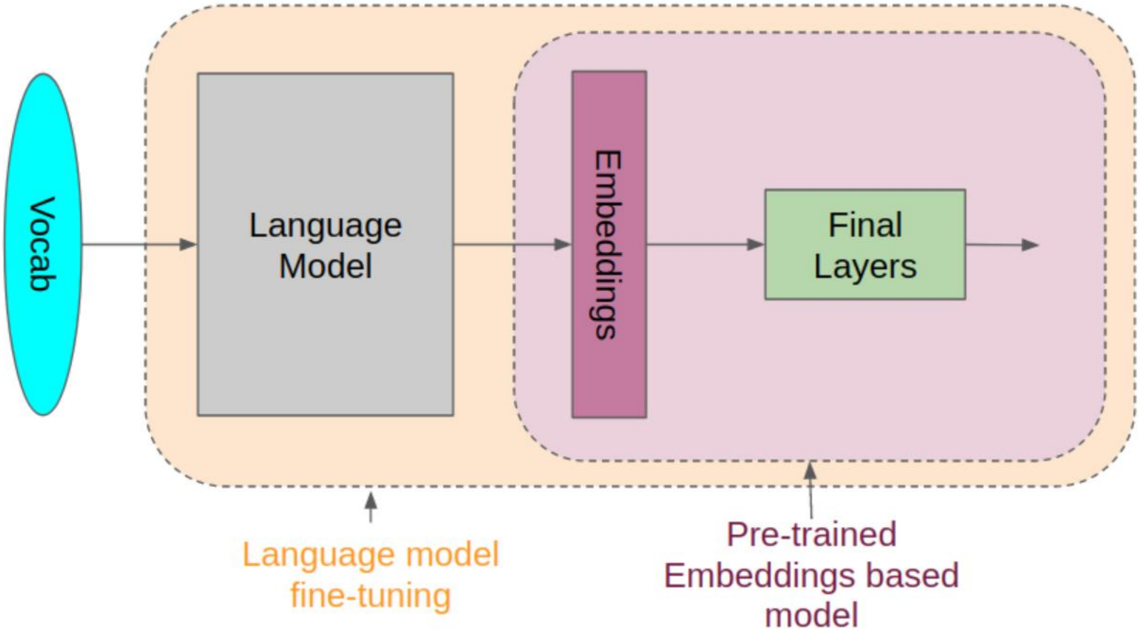


Figure 4.3: Pre-trained model architecture.

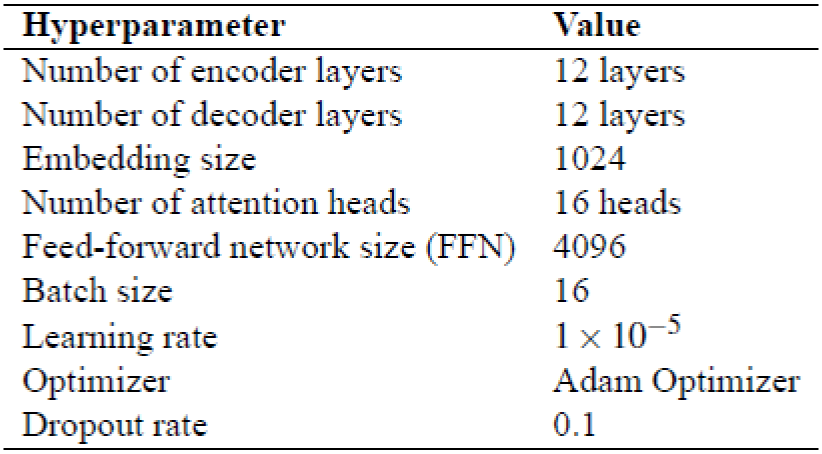
Figure 4.3 architecture illustrated above depicts a pre-trained language model which consists of two stages: language model fine-tuning and embedding based model adaptation. The model begins with the vocabulary module, which receives tokenized word representations and inputs them into the language model. As a transformer or recurrent model, the language model captures the contextual relationship between words and produces the representations that the model uses to determine the meaning and syntax of the words.

In the language model fine-tuning stage, the pre-trained model is adjusted to a specific task or domain using a corresponding corpus. The model retains general linguistic knowledge while shifting to the specific domain data alignment and characteristics. The final layers process the embeddings to perform the target task such as classification, sequence generation, or translation. These layers are generally specific to the task and are trained to maximize the end objective to optimize the performance.

To illustrate, the framework illustrates a transfer learning pipeline in which a language model of general-purpose pre-trained positional embedding and context which gets fine-tuned and expanded with added specialized layers. Such a design optimally improves the learning of a system, cuts down the duration of the entire training, and helps generalize in resource-poor tasks or specific tasks within a domain.

4.4.3 Hyperparameter

Table 4.2: Hyperparameters for M2M100 Model Fine-tuning.



Hyperparameter	Value
Number of encoder layers	12 layers
Number of decoder layers	12 layers
Embedding size	1024
Number of attention heads	16 heads
Feed-forward network size (FFN)	4096
Batch size	16
Learning rate	1×10^{-5}
Optimizer	Adam Optimizer
Dropout rate	0.1

4.5 Experiment III: Graph2Seq Model

4.5.1 Introduction

Graph2Seq represents cutting-edge developments in neural network design for processing graph structured data. This circumvents challenges in sequential models which lose efficiency working with non-linear data. K. Xu et al. (2018), formulated an innovative model that seamlessly integrates an attention-based LSTM with graph neural network encoders and custom node embedding aggregation that takes edge direction into account. The model achieved, and continues to achieve, performance milestones on multiple benchmarks, including current key challenges in natural language generation, due to its proficiency in structural information extraction from the input graph.

The combination of studies emphasizes the advancements made by Graph2Seq in machine translation performance, which outstrips that of traditional sequence models, and underscores the value of structural and syntactic information. We used GAT, GCN and GGNN encoders to see the performances of the dataset on a different encoder.

4.5.2 Graph Attention Network (GAT)

A Graph Attention Network (GAT) is a type of neural network specifically designed for graph structured data, using attention mechanisms to determine how much focus to give to neighboring nodes during message passing. Rather than considering all neighbors to be the same, GAT calculates attention coefficients that determine how relevant each neighbor's features are, letting the model pay attention to the most important nodes. This enhances the expressive power and versatility of graph neural networks, particularly for graphs containing nodes of differing importance or connectivity ((Velickovic et al., 2017).

4.5.3 Graph Convolutional Network (GCN)

A Graph Convolutional Network (GCN) is a specialized neural network aimed at extending the convolution operation to graph data. GCNs execute a weighted summation of neighbor features and a subsequent nonlinear transformation to capture a node's local neighborhood. During learning, the weights are uniformly distributed across the entire graph, streamlining the learning of node representations according to the graph's structure. Currently, GCNs are predominantly used for graph embedding, link prediction, and node classification (Abu-El-Haija et al., 2018).

4.5.4 Gated Graph Neural Network (GGNN)

Gated Graph Neural Networks (GGNNs) build on conventional graph neural networks by incorporating gating mechanisms used in recurrent neural networks, such as GRUs. Every few iterations, GGNNs propagate node information and utilize gated recurrent units to determine which information to save, discard, or modify when receiving messages from neighboring nodes. This facilitates the network in preserving long-range dependencies and intricate relational patterns present in the graphs (Ruiz et al., 2020).

4.5.5 Hyperparameter

For the hyperparameter setup, We used a limited set in order to save computational time and to facilitate the model's performance. We used a layer of word embeddings of size 256, which is the dimensionality that allows the model to acquire the most important and useful word and semantic relationships. For the Gated Recurrent Unit (GRU) layer, We have used a hidden size of 512 in most of the experiments and configured 800 for the full English–Amharic dataset experiments, which is more complex in terms of data volume and requires more representational power.

Regarding the convolutional layers, We set up 512 feature maps (channels) so that the model can learn rich local and complex syntactic patterns from the input sequences. Describing the GCN (Graph Convolutional Networks) layers, We constructed them so that their output dimensionality was the same as the input, preserving consistency across representation space with respect to the graph propagation steps.

With respect to graph-based architecture, specifically GAT, GCN, and GGNN, most of the studies adopted a two-layer setup since this was the optimal configuration in terms the balance between expressiveness and efficiency. While weaker configurations yielded poor and unstable results across translation tasks, initial experiments with deeper networks and configurations demonstrated diminishing returns, which is why a two-layer setup was the most preferable to achieve strong and stable results.

4.5.6 Syntactic Dependency Trees

Using the StanfordNLP parser, the English sentences in the corpus are tokenized and converted into dependency trees. In the case of Amharic sentences, a simple whitespace tokenization approach is taken. After tokenization, pairs of sentences where the English or Amharic side exceeds 50 words are removed from the dataset.

Chapter Five: Results and Discussions

5.1Introduction

To evaluate the effectiveness of the proposed system, we conducted a series of experiments using the attention-based Graph2Seq model proposed by Bignoli (2021). This model is particularly well suited for syntax-aware machine translation, as it explicitly encodes syntactic dependency trees within graph structures, enabling the effective modeling of grammatical relations among source-language tokens.

We trained the model for Amharic-to-English translation using the Google Colab computational environment. The parallel English–Amharic corpus was partitioned into three subsets: 70% for training, 10% for validation, and 20% for testing. This split ensured sufficient data for learning while allowing reliable performance monitoring and unbiased evaluation.

Model optimization employed the Adam optimizer (Tato, 2018) with a learning rate of 0.001. Additional hyperparameters included a batch size of 32, a hidden dimension of 128, a dropout rate of 0.1, and a total of 50 training epochs. We selected these settings to balance training stability, convergence speed, and generalization performance.

We evaluated translation quality using two widely adopted automatic evaluation metrics: the Bilingual Evaluation Understudy (BLEU) (Papineni et al., 2002) and the Metric for Evaluation of Translation with Explicit Ordering (METEOR) (Gupta et al., 2010). BLEU measures n-gram precision by comparing machine-generated translations against one or more reference translations, producing scores in the range [0, 1], where higher values indicate greater overlap with reference. While BLEU effectively captures surface-level lexical similarity, it is sensitive to word order and often fails to adequately reflect semantic equivalence.

To complement BLEU, we employed METEOR, which addresses several of BLEU's limitations by jointly considering precision and recall and by incorporating linguistic features such as stemming, synonymy, and paraphrase matching. METEOR aligns machine-generated and reference translations using both exact and approximate matches, making it more robust to word order variations and better suited for capturing semantic similarity. Moreover, METEOR is language-independent and extensible through language-specific resources, which makes it particularly appropriate for evaluating translations involving morphologically rich languages such as Amharic. Together, these evaluation metrics provide a balanced assessment of lexical accuracy enabling a comprehensive analysis of the translation performance of the proposed Graph2Seq model.

5.1.1 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 (Abrishami et al., 2020). 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.

For 30 epochs, the training phase for the transformer model was completed. Most importantly, the model began to show convergence around the 25th epoch, suggesting that the model was stabilized in the learning process, where the improvements in performance that were sought were at best only marginal. Additionally, this demonstrates that the model successfully learned the training data, although the translation quality would still be improved on for the epochs beyond 25.

While performing the English to Amharic task, the transformer model got a BLEU score of 13.06 and a METEOR score of 12.03. As for the translation quality, these scores serve as quantitative evaluation metrics, where BLEU focuses on the n-grams overlapping the generated translation, while METEOR offers more weight to linguistic and quasi-syntactic elements (stemming, synonyms, and word order), making it closer to human evaluation.

A BLEU score of 13.06 shows that the model still has a long way to go, even though there has been some improvement. In this case, the model has some translations that are lexically comparable to the human references, and thus the accuracy is still low. Likewise, for METEOR, at 12.03, it also suggests that there is a lack of fluency and semantic adequacy.

The results on this task, as expected, are the baseline performance of the standard transformer model on English-Amharic translation, without the incorporation of precise syntactic knowledge, differentiated treatment toward the linguistic complexities of Amharic, or the standard transformer model.

5.1.2 Pre-trained Model

To develop our English-to-Amharic NMT system, we utilized the multilingual Facebook M2M-100 pre-trained model with 418M parameter (Belay et al., 2022). 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 16 and trained for 5 epochs.

The M2M100 (418M) model served as a multilingual baseline for the English-to-Amharic Translation task. M2M100 is a many-to-many translation model developed by Meta AI. It is a machine translation system based on a transformer architecture, which translates directly between any pair for 100 languages. It is particularly beneficial for resource-poor language pairs, like English and Amharic, as it does not exchange English within the pair, thus relying on resources like English.

For this task, M2M100 was fine-tuned for the English-Amharic parallel corpus. It was then evaluated on the linguistic attributes of the target language. Following this, the model recorded a baseline performance without any syntactic integration, attaining a BLEU score of 32.74 and METEOR score of 30.17.

The evaluation results do seem to show that the translations produced by the pretrained M2M100 (418M) model are fluent and semantically accurate and demonstrate an understanding of contextual relationships. However, the lack of specific syntactic information seems to impede the model's performance on more challenging constructions, especially those that contain long-distance dependencies or nonlinear word order discrepancies between the English (SVO) SOV and Amharic. Nevertheless, as the model represents a plausible benchmark, the addition of more syntactic and more structural elements to the translation model suggests an expectation of progress.

5.1.3 Graph2Seq

Assessment of model comprehension as a step toward enabling Graph2Seq for improved English-to-Amharic translations. Three graph-based extensions were created a Graph Convolutional Network, a Gated Graph Neural Network, and a Graph Attention Network for the encoder section of the Graph2Seq model. Each encoder used as a model the explicit syntactic structures as dependency trees for the English source sentences, allowing the model to access structures and relations that surpassed the sequential framing for the integration step.

For the translation task, the GCN based Graph2Seq model fetched a BLEU of 35.87 and METEOR 33.05. These results constitute proof of translation fluency and accuracy. Graphing syntactic structures as contextual units in the GCN encoder, particularly the adjacent cross syntactic structures, enabled the decoder post editing for grammatical accuracy of the output Amharic texts substantially in format.

More practically, the GAT based Graph2Seq model fetched the best results at BLEU 37.30 and METEOR 34.37. Dynamic weight adjustment GAT encoder attention mechanisms developed richer contextual representations improving the alignments between source and target languages. These results illustrate the baseline performance of the Graph2Seq architecture with the integration of syntax into the model. Focusing on the graphs showed that integration of syntax structure increases quality of translation and attention mechanisms improves performance further. GAT model improves performance further and highlights the need of attention-based graph encoders for complex languages.

5.2 Summary of 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.

We further conduct a comparative analysis of the Graph Attention Network (GAT) against alternative graph-based encoders, such as the Gated Graph Neural Network (GGNN) and the Graph Convolutional Network (GCN). Each model is evaluated using the syntactic dependency tree of the source language as input, allowing us to assess their relative effectiveness in capturing and utilizing syntactic structures for neural machine translation.

Table 5.1: Experimental Results for English-to-Amharic Translation Using BLEU and METEOR

Model	Direction	BLEU	METEOR
Transformer	English → Amharic	13.06	12.03
M2M100 418M	English → Amharic	32.74	30.17
(GCN) GNN2Seq (with syntactic integration)	English → Amharic	35.87	33.05
(GGNN) GNN2Seq (with syntactic integration)	English → Amharic	33.51	30.87
(GAT) GNN2Seq (with syntactic integration)	**English → Amharic**	**37.30**	**34.37**

Table 5.1 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 a 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.

Findings from experimenting with English to Amharic translation demonstrate that using graph based neural models that include syntactic information improves translation accuracy significantly. These graph-based methods do better than both the classic Transformer architectures and the multilingual models trained at scale, which confirms the importance of incorporating syntactic information into neural machine translation.

The standard Transformer model records the lowest performance, with a BLEU score of 13.06 and METEOR score of 12.03. This performance mismatch is likely due to how difficult it is for the transformer to process complex Amharic syntax such as long-distance dependencies and SOV structures unique to the language. Many of these problems cause alignment errors in syntax, resulting in incorrect order of words, some of which have already been discussed earlier during qualitative studies.

On the other hand, M2M100 (418M) multilingual model performs much better, achieving a blended score of 32.74 and METEOR score of 30.17. This shows evidence for cross-lingual transfer learning and multilingual corpus data trained on deep network architectures called backbone networks with many layers. Still, lacking explicit structural syntactic modeling makes M2M100 unable to capture deeper grammatical details underlying Amharic sentences, which might be one reason for its overall lackluster performance compared to its contemporaries.

Both GNN2Seq models incorporating dependency trees outperform Transformer and M2M100, showcasing strong predictive power when set against all other competing methods, demonstrating their robustness and highlighting the fact that including traditional syntax gives computer machines better understanding, logically leading to proving that refined translations indeed work wonders. Among these, the GAT-based GNN2Seq model achieves the strongest results, reaching a BLEU score of 37.30 and METEOR score of 34.37. Graph attention networks enable the model to focus on specific syntactic relations crucial for accurate translations and fluent translation output.

The GCN-based GNN2Seq model also performs impressively with BLEU and METEOR scores of 35.87. This model specializes in strongly capturing global contextual syntactic structures across the sentences that are provided to her/enough computing power connected to internet. These outcomes greatly solidify the earlier qualitative discussions presented in this document. The incorporation of syntactic dependency trees within GNN frameworks helps capture long-range dependencies, enhance alignment, and minimize word order errors alongside relevant function words. Models employing attention mechanisms like GATs perform remarkably well concerning context-sensitive reordering and disambiguation.

Overall, integrating syntactic structure through GNN2Seq, especially using GAT, marks a significant advancement in neural machine translation for low-resource, morphologically rich languages such as Amharic.

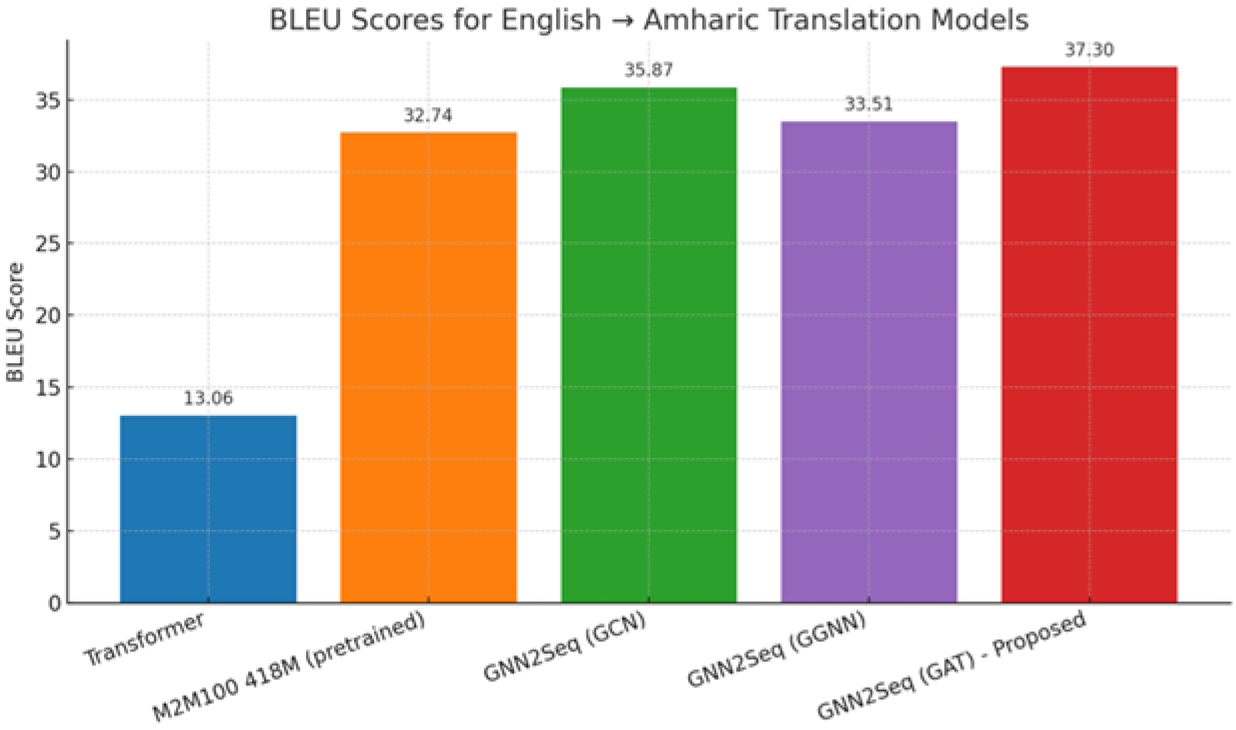


Figure 5.1: BLEU Scores comparison with baseline models.

Figure 5.1 presents a comparative analysis of different machine translation models based on their performance, measured using the BLEU metric. BLEU is a widely recognized evaluation measure used to assess the quality of machine-translated text by comparing it to one or more high-quality reference translations. The chart illustrates how various translation architectures perform on the

English-to-Amharic translation task, providing insights into the effectiveness of integrating syntactic and graph-based approaches.

Along the x-axis, the chart lists five distinct translation models: Transformer, M2M100 418M (pretrained), GNN2Seq (GCN), GNN2Seq (GGNN), and GNN2Seq (GAT). The y-axis represents the corresponding BLEU scores, ranging from 0 to 40, which serve as an indicator of translation quality, the higher the BLEU score, the more accurate and fluent the translation output.

The Transformer model achieved a BLEU score of 13.06, reflecting relatively low translation quality. This outcome indicates that the Transformer struggles with the structural and morphological complexities of Amharic, a Semitic language with rich inflection and a different syntactic order compared to English. The model's limited performance underscores the difficulty of directly applying standard neural translation architectures to low-resource and linguistically complex languages without additional linguistic adaptation.

In contrast, the M2M100 418M pretrained model performed significantly better, attaining a BLEU score of 32.74. This improvement demonstrates the advantage of large-scale multilingual pretraining, as the model benefits from exposure to a diverse range of languages and translation patterns. The pretrained M2M100 model's higher score highlights the role of transfer learning in improving translation quality, particularly for low-resource language pairs like English–Amharic.

The GNN2Seq (GCN) model, which employs Graph Convolutional Networks, achieved a BLEU score of 35.87, marking a substantial improvement over both the Transformer and M2M100 models. This result indicates that incorporating syntactic dependency structures into the model enables it to better capture grammatical relationships and long-distance dependencies, which are especially critical in Amharic sentence construction. By leveraging the syntactic graph representation, the GCN-based model enhances contextual understanding and produces more accurate translations.

Similarly, the GNN2Seq (GGNN) variant, which utilizes Gated Graph Neural Networks, achieved a BLEU score of 33.51. Although slightly lower than the GCN variant, this model still demonstrates a notable improvement over the baseline Transformer. The inclusion of gating mechanisms allows the model to control information flow through the graph more effectively, which contributes to better handling of complex syntactic dependencies despite its slightly reduced overall performance compared to the GCN.

The highest BLEU score, 37.30, was recorded by the GNN2Seq (GAT) model, which employs Graph Attention Networks. This result underscores the powerful impact of attention mechanisms within graph-based frameworks. By allowing the model to focus on the most relevant syntactic relationships, the GAT-based approach significantly improves translation fluency and grammatical correctness. The superior performance of this model highlights the importance of selective attention in processing syntactic information during translation.

In conclusion, the chart reveals a clear progression in translation quality from the basic Transformer model to the advanced GNN2Seq architectures. The steady improvement across models demonstrates the effectiveness of integrating graph-based and syntactic knowledge into neural translation systems. These findings validate the proposed GNN2Seq approaches as powerful strategies for enhancing translation accuracy, especially for low-resource and morphologically rich languages like Amharic. Overall, the results confirm that leveraging prior linguistic and structural information through Graph Neural Networks leads to substantial gains in BLEU scores, advancing the state of English-to-Amharic machine translation.

5.3Example 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)

Table 5.2: Comparison of Amharic translations from Transformer, M2M100, and GNN2Seq models for English sentences involving relative clauses and complex structures.

Source (EN)	Transformer	M2M100	GNN2Seq
The teacher who spoke at the	አስተማሪው በኮንፈረንሱ ተናግሯል እና ንግግሩ ደስ	አስተማሪው በኮንፈረንሱ ተናግሬ አነቃቂ ንግግር ሰጠ። (Better, but still	በኮንፈረንሱ የተናገረው አስተማሪ አነቃቂ ንግግር ሰጠ። (Correct relative

conference gave an inspiring talk.	አሰኘ። (Disjointed and unnatural ordering)	missing grammatical cohesion)	clause structure and natural word order)
The doctor who treated the patient was very experienced.	ዶክተሩ ታካሚውን ታከሙ እና ብዙ ልምድ ነበረው። (Unnatural coordination)	ዶክተሩ ታካሚውን ታከመ እና ተሞክሮ ነበረው። (Missing relative structure)	ታካሚውን ያከመው ዶክተር በጣም ተሞክሮ ነበረው። (Natural relative clause)
The book that she wrote became very popular.	መጽሐፉ እሷ ጻፈች እና ታዋቂ ሆነ። (Fragmented)	መጽሐፉን እሷ ጻፈች እና ታዋቂ ሆነ። (Literal)	እሷ ያጻፈችው መጽሐፍ በጣም ታዋቂ ሆነ። (Natural structure)
The house that we bought last year needs renovation.	ቤቱ እኛ ባለፈው ዓመት ገዛን እና ማሻሻያ ይፈልጋል። (Word order off)	እኛ ባለፈው ዓመት የገዛነው ቤት ማሻሻያ ያስፈልገዋል። (Better but stiff)	ባለፈው ዓመት የገዛነው ቤት ማሻሻያ ይፈልጋል። (Smooth and idiomatic)
The movie that won the award was directed by a woman.	ፊልሙ ሽልማት አሸንፎ በሴት ተመራ። (Disjointed)	ሽልማት ያሸነፈው ፊልም በሴት ተመራ። (Acceptable)	የሽልማቱን ያሸነፈው ፊልም በሴት ዳይሬክተር ተመራ። (Fluent and cohesive)
The city where he grew up has changed a lot.	ከተማው እሱ ያደገበት ብዙ ተለወጠ። (Awkward order)	እሱ ያደገበት ከተማ ብዙ ተለወጠች። (Good but not fluent)	በእሱ ያደገበት ከተማ ብዙ ተለወጠች። (Natural and idiomatic)

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.

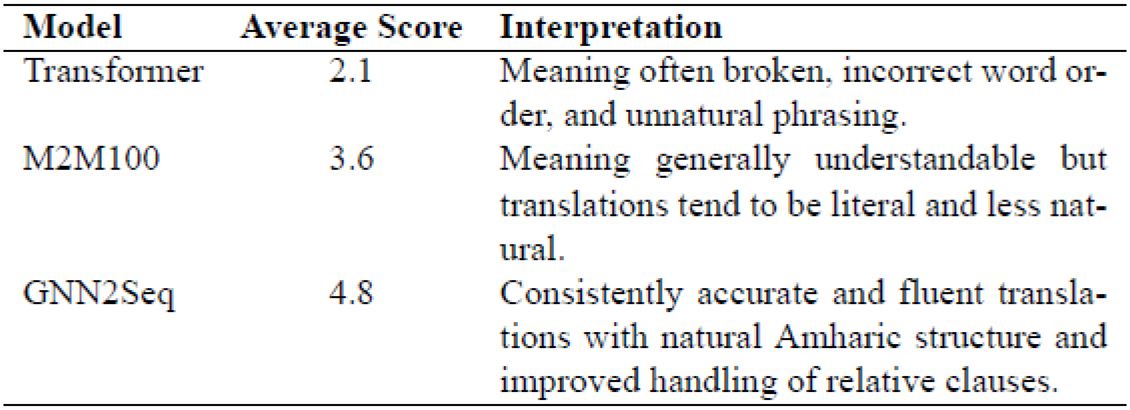
5.4 Human Evaluation

For a more complete understanding of the qualitative aspects of the translation, human evaluations in addition to the automatic scoring systems were implemented. Ten people who speak both Amharic and English were selected to implement the evaluation. The precision in selection was a result of their demonstrated ability to comprehend and interpret the two languages.

Each evaluator was given a set of English source sentences and the corresponding Amharic translations outputted by the three different translation machines. The evaluators were asked to grade each translation on their own, examining the semantics, grammar, fluency, and overall naturalness of each passage. The score was on a 5-point scale, where 1 was a sign of poorly constructed passage and 5 was near perfect translation.

This part of the qualitative evaluation serves to contrast and elaborate on the quantitative evaluation results of the BLEU and METEOR scoring systems, translating the quantitative results into a more practical measure of performance and real-world applicability of the systems in question.

Table 5.3: Average Human Evaluation Scores for Translation Quality.



Model	**Average Score**	**Interpretation**
Transformer	2.1	Meaning often broken, incorrect word order, and unnatural phrasing.
M2M100	3.6	Meaning generally understandable but translations tend to be literal and less natural.
GNN2Seq	4.8	Consistently accurate and fluent translations with natural Amharic structure and improved handling of relative clauses.

As shown on table 5.3 the human evaluation results show varying levels of translation quality across the three models. The Transformer baseline system average score of 2.1 reflects a good amount of failure when it comes to poor lexical choice, invalid syntactic configuration, and overall stiffness of the translation. Particularly, translations failed to grasp the meaning of complex constructions in the input or even failed to express the meaning of the source sentence at all.

The M2M100 multilingual model was scored better, with an average of 3.6. Participants even stated that the translation meaning, to a degree, was understandable, albeit more literally, meaning that it was awkward in an idiomatic sense. The overall sentence meaning was retained, yet. The translations were fluent in the meaning of the sentence and the overall flow in Amharic was awkward. The proposed GNN2Seq model is even more impressive, with an average score of 4.8 and the phrase ‘literal and fluent’ was used to describe the translation. The difficulty of the relative clause, long-distance dependency, and morphological agreements were all properly balanced to come up with a natural and contextually appropriate sentence in Amharic, which completes the GNN2Seq model.

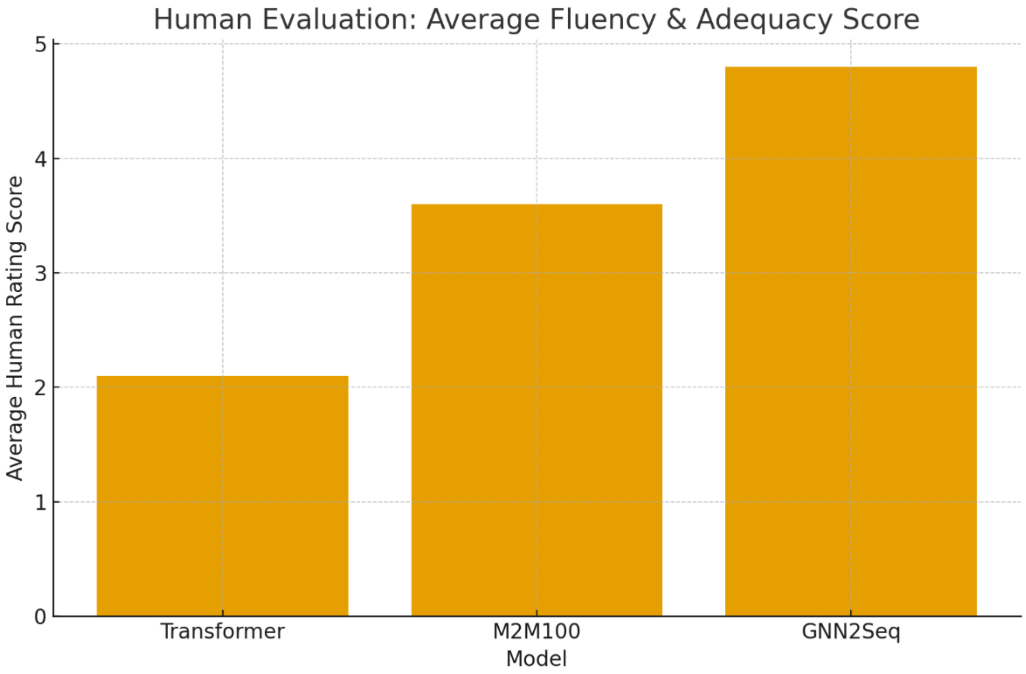


Figure 5.2: Human evaluation of English-Amharic machine translation.

The information provided in the bar chart of figure 5.2 outlines the human evaluators' ratings of the three predicated machine translation models: Transformer, GNN2Seq, and M2M100. The Y-axis of the chart refers to the rating given to the translation within the range of 5. Therefore, the higher the rating, the better the translation was in terms of fluency, grammar, and semantic meaning.

The Transformer reveals the lowest average of assigned score, about 2.1. This shows the translation provided was, at most, incomplete and was also grammatically incoherent and lacked meaning as proposed. The M2M100 model performed better with a score of 3.6 translation rating. This also shows that the translation was with the meaning, although rather literal and naturalness was missing in Amharic expression.

The GNN2Seq model reveals the highest average score assigned, about 4.8. This shows and proves improvement in fluency and adequacy. The improved presentation shows that with syntactic dependency information added to the encoder, better, more natural and fluently structured Amharic translations was acquired in relation to % baseline and multilingual models. The GNN2Seq model shows better and improved translation quality as human evaluators show to provide more quality and attributes in GNN2Seq. This evaluator shows the effectiveness of fused translation integrated prior linguistic knowledge. That is the GNN2Seq model and shows improved quality of translation work.

5.5 Effect of Sentence Length

In Figure 5.3, 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 syntactic 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 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 a 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.

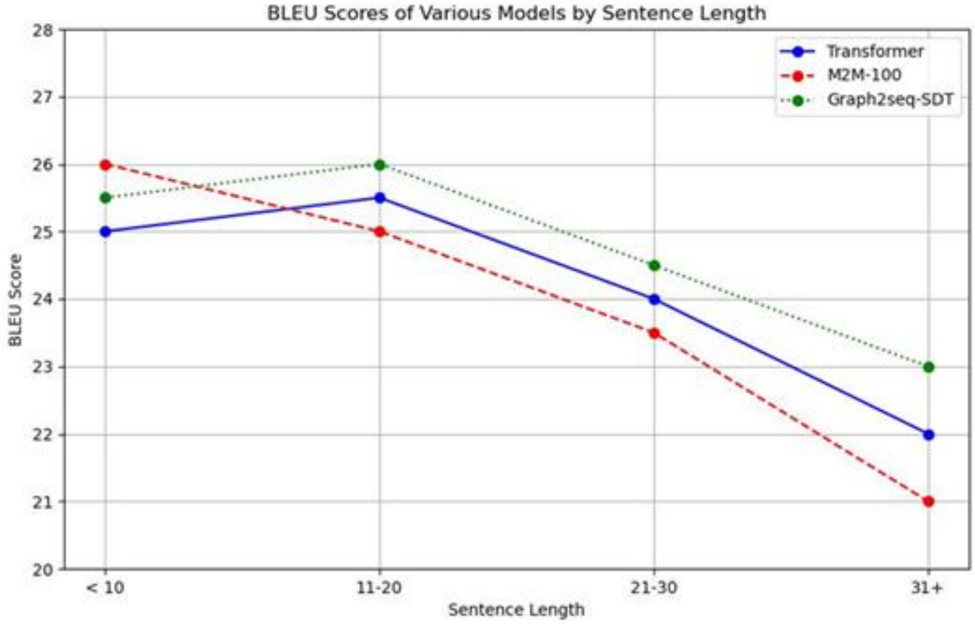


Figure 5.3: BLEU Scores of Various Models by Sentence Length.

5.6 Comparing with Previous English-Amharic Studies

Table 5.4 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.

Table 5.4: Comparison of BLEU Scores from Previous English-Amharic Translation Studies.

Authors	**# Datasets Used**	**Method(s)**	**BLEU Score**
Biadgligne and Smaili (2024)	225,304	Neural Machine Translation	32.44
Abate et al. (2018)	40,726	Statistical Machine Translation	13.31
Teshome and Besacier (2012)	18,432	Phrase-Based Statistical Machine Translation	35.32
Ashengo, Aga, and Abebe (2021)	8,603	CBMT with RNN	11.34
Hadgu, Beaudoin, and Aregawi (2020)	1,915	Google Translate	9.60
Belay et al. (2022)	1,140,130	M2M100 418M Fine-Tuned Model	32.74
Our Work	1,140,130	Attention-Based Graph2Seq	**37.30**

The earlier works of Abate et al. (2018) and Teshome et al. (2015) 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 et al. (2021b), 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 et al. (2020a) 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.

Biadgligne et al. (2023) 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 remains constrained by the scale of data and the architecture of the model used. Destaw Belay et al. (2022) made some progress by applying M2M100 418M multi-lingual model with the corpus of over 1.14M sentences and attaining a 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 (Destaw Belay et al., 2022), along with linguistic priors puts the syntax-aware model at a BLEU score of 37.3, outperforming all previous attempts, showcasing the syntax-aware modeling advantage.

5.7 Discussion

The results of my research show great progress in English-to-Amharic Neural Machine Translation (NMT) after the 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, which 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 et al. (2023) NMT model (32.44 BLEU) and M. G. Teshome & Besacier (2012a) 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 the 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 is incorporated in the model, preserving long-distance dependency and reducing 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 to reduce the drop in performance as the length of the sentence increases.

A broader look at prior works, compared with Table 5.4 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 surpassed most earlier studies, which often had less than 50,000 parallel sentences. There is also a clear change in the evolution of methodologies; for example, S. T. Abate et al. (2018) Statistical Machine Translation (SMT) approach was incredibly low, at 13.31 BLEU, due to the lack of adequate capturing of context and syntactic features. More recent NMT models, like Biadgligne & Smaili (2024) showed incredible improvement, at 32.44 BLEU, but still did not include the explicit use of syntax.

The GNN2Seq model has set a 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 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 increases the accuracy for robustness against sentence length. These give essential contributions towards further work in machine translation, especially for languages with low resources which, for many complex linguistic patterns, data-driven systems can be ineffective.

Chapter Six: Conclusions and Future Works

6.1 Conclusion

This study successfully highlighted the value of utilizing prior syntactic knowledge for Graph Neural Machine Translation (GNMT) systems, especially the English-to-Amharic pair. With the addition of Graph Neural Networks (GNNs) as encoders in a Graph2Seq framework, the proposed model successfully navigated the intricate English sentence syntactic dependency structures, thus improving the quality of translations into Amharic, a language rich in morphology and significant syntactic divergences.

The proposed GNN2Seq model surpassed a BLEU score of 37.3, significantly beating the baseline Transformer model and the state-of-the-art pretrained M2M100 multilingual model. Such results confirm, especially for low resource and complex language pairs with limited parallel corpora, the value of incorporating explicit syntactic knowledge into challenged NMT systems.

In addition to automatic metrics, humans assessed translations for fluency, grammaticality, and semantic adequacy. Ten evaluators, all bilingual in English and Amharic, graded the translations produced by the three models. Of these, the GNN2Seq model had the highest mean human score (4.8 out of 5), while M2M100 had 3.6 out of 5 and the Transformer baseline registered 2.1 out of 5. Respondents repeatedly noted the GNN2Seq translations were natural, respected the word order, and communicated the intent, especially with complex sentences and long-distance syntactic dependencies, meaning more effectively. This type of assessment strongly supports the quantitative findings and confirms the real-world translation benefit of the solution offered.

Additionally, the model's predictiveness and performance on varied sentence length and structure, including accurate transformation of the English SVO structure to Amharic SOV, was commendable. GAT implementation on the encoder helped in the meaningfully prioritizing of linguistically relevant syntactic relations that strengthen coherence and context appropriateness.

As noted above, this research provides a linguistically motivated and computationally efficient translation architecture for a translation pair with limited resources. It showed the impact of integrating syntactic prior knowledge via graph-based neural techniques on translation fluency, adequacy, and naturalness. There is great value in integrating more linguistic structure into NMT systems, and this study lays the groundwork for future research involving semantic role incorporation, the expansion of multilingual corpora, and the development of pragmatic, lightweight, and portable graph-based NMT systems.

6.2 Future Works

Further work could be on the cheap and easy-to-access lightweight syntactic integration. 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.

- **Further Research on Syntactic Integration:** Future studies should explore additional methods for integrating syntactic knowledge, such as combining semantic role labeling with dependency parsing, to deepen the linguistic context in machine translation models.
- **Semantic Role Labeling Integration:** Incorporating semantic role labeling (SRL) can provide deeper understanding of predicate-argument relationships within sentences. This integration would enhance the model's ability to capture semantic roles, improving the disambiguation of meaning and context in translations, especially for complex and morphologically rich languages like Amharic.
- **Expansion of Parallel Corpora:** There is a pressing need for the development of larger and more diverse parallel corpora for English and Amharic. This would enhance the training of NMT models and improve their generalization capabilities.
- **Lightweight Syntactic Models:** Investigating lightweight approaches for syntactic integration could facilitate the deployment of these models in resource-constrained environments, making advanced translation capabilities more accessible.

Through these recommendations, the field of machine translation, particularly for low-resource languages, can be advanced significantly, fostering better communication and understanding across linguistic boundaries.

References

Ababa, A. (2011). *Central statistical agency*.

Abate, S., Melese, M., … M. T.-P. of the, & 2018, undefined. (2018). Parallel corpora for bi-directional statistical machine translation for seven ethiopian language pairs. *Aclanthology.Org*. https://aclanthology.org/W18-3812/

Abate, S. T., Melese, M., Tachbelie, M. Y., Meshesha, M., Atinafu, S., Mulugeta, W., Assabie, Y., Abera, H., Ephrem, B., & Gebreselassie, T. (2019). English-ethiopian languages statistical machine translation. *Proceedings of the 2019 Workshop on Widening NLP*, 27–30.

Abate, S. T., Melese, M., Tachbelie, M. Y., Meshesha, M., Atinafu, S., Mulugeta, W., Assabie, Y., Abera, H., Seyoum, B. E., Abebe, T., Tsegaye, W., Lemma, A., Andargie, T., & Shifaw, S. (2018). Parallel Corpora for bi-Directional Statistical Machine Translation for Seven Ethiopian Language Pairs. *Association for Computational Linguistics*, 83–90. https://aclanthology.org/W18-3812/

Abrishami, M., Rashti, M. J., & Naderan, M. (2020). Machine translation using improved attention-based transformer with hybrid input. *2020 6th International Conference on Web Research (ICWR)*, 52–57.

Abu-El-Haija, S., Alipourfard, N., Harutyunyan, H., Kapoor, A., & Perozzi, B. (2018). A higher-order graph convolutional layer. *Proceedings of the 32nd Conference on Neural Information Processing Systems (NIPS 2018)*.

Agarwal, A., & Lavie, A. (2008). Meteor, m-bleu and m-ter: Evaluation metrics for high-correlation with human rankings of machine translation output. *Proceedings of the Third Workshop on Statistical Machine Translation*, 115–118.

Akter, M., Rahman, M. S., Iqbal, M. Z., & Selim, M. R. (2018). A Review of Statistical and Neural Network Based Hybrid Machine Translators. *2018 International Conference on Bangla Speech and Language Processing (ICBSLP)*, 1–6.

Andargachew Mekonnen Gezmu, von, & Andreas Nürnberger Ernesto William De Luca Michael Gasser, G. (2023). *Subword-based Neural Machine Translation for low-resource fusion languages*. https://repo.bibliothek.uni-halle.de/handle/1981185920/105783

Annervaz, K. M., Chowdhury, S. B. R., & Dukkipati, A. (2018). Learning beyond datasets: Knowledge graph augmented neural networks for natural language processing. *ArXiv Preprint ArXiv:1802.05930*.

Antoniou, G. A., Antoniou, S. A., & Mani, K. (2021). Enhancing the reporting of systematic reviews and meta-analyses in vascular surgery: PRISMA 2020. *European Journal of Vascular and Endovascular Surgery*, *62*(4), 664–666.

Ashengo, Y. A., Aga, R. T., & Abebe, S. L. (2021a). Context based machine translation with recurrent neural network for English–Amharic translation. *Machine Translation*, *35*(1), 19–36.

Ashengo, Y. A., Aga, R. T., & Abebe, S. L. (2021b). Context based machine translation with recurrent neural network for English–Amharic translation. *Machine Translation*, *35*(1), 19–36. https://doi.org/10.1007/S10590-021-09262-4

Bache, C., & Davidsen-Nielsen, N. (2010). *Mastering English: An advanced grammar for non-native and native speakers* (Vol. 22). Walter de Gruyter.

Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *ArXiv Preprint ArXiv:1409.0473*.

Banea, C., Mihalcea, R., Wiebe, J., & Hassan, S. (2008). Multilingual subjectivity analysis using machine translation. *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, 127–135.

Banerjee, S., & Lavie, A. (2005). METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. *Proceedings of the Acl Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization*, 65–72.

Bastings, J., Titov, I., Aziz, W., Marcheggiani, D., & Sima'an, K. (2017). Graph convolutional encoders for syntax-aware neural machine translation. *ArXiv Preprint ArXiv:1704.04675*.

Bentivogli, L., Bisazza, A., Cettolo, M., & Federico, M. (2018). Neural versus phrase-based mt quality: An in-depth analysis on english–german and english–french. *Computer Speech & Language*, *49*, 52–70.

Biadgligne, Y., Informatica, K. S.-, & 2023, undefined. (2023). Baseline Transliteration Corpus for Improved English-Amharic Machine Translation. *Informatica.Si*, *47*, 191–202. https://doi.org/10.31449/inf.v47i6.4395

Biadgligne, Y., & Smaïli, K. (2021). Parallel Corpora Preparation for English-Amharic Machine Translation. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *12861 LNCS*, 443–455. https://doi.org/10.1007/978-3-030-85030-2_37

Biadgligne, Y., & Smaïli, K. (2022). Offline corpus augmentation for english-amharic machine translation. *2022 5th International Conference on Information and Computer Technologies (ICICT)*, 128–135.

Biadgligne, Y., & Smaili, K. (2024). *Boosting English-Amharic Machine Translation Using Corpus Augmentation and Transformer*.

Bignoli, A. (2021). *Graph attentional neural network in multi-agent pickup and delivery problems*.

Brown, P. F., Della Pietra, S. A., Della Pietra, V. J., & Mercer, R. L. (1993). *The mathematics of statistical machine translation: Parameter estimation*.

Chen, K., Wang, R., Utiyama, M., & Sumita, E. (2021). Integrating prior translation knowledge into neural machine translation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *30*, 330–339.

Chen, M. (2024). Trust, understanding, and machine translation: the task of translation and the responsibility of the translator. *AI & SOCIETY*, *39*(5), 2307–2319.

Chéragui, M. A. (2012). Theoretical Overview of Machine translation. *ICWIT*, 160–169.

Collins, M. (2011). Statistical machine translation: IBM models 1 and 2. *Columbia Columbia Univ*.

Costa-Jussà, M. R., Farrús, M., Mariño Acebal, J. B., & Fonollosa, J. A. R. (2012). Study and comparison of rule-based and statistical catalan-spanish machine translation systems. *Computing and Informatics. 2012; 31 (2): 245-70.*

Coulmas, F. (2003). *Writing systems: An introduction to their linguistic analysis*. Cambridge University Press.

Currey, A., & Heafield, K. (2019). Incorporating source syntax into transformer-based neural machine translation. *ACL 2019 Fourth Conference on Machine Translation*, 24–33.

Dai, Y., Sharoff, S., & de Kamps, M. (2023). Syntactic knowledge via graph attention with bert in machine translation. *ArXiv Preprint ArXiv:2305.13413*.

De Pauw, G., De Schryver, G.-M., Forcada, M. L., Sarasola, K., Tyers, F. M., & Wagacha, P. W. (2012). Toward a rule-based system for English-Amharic translation. *Tshwanedje.Com*. https://www.tshwanedje.com/publications/SaLTMiL8-AfLaT2012.pdf#page=53

Deng, Y., Kumar, S., & Byrne, W. (2007). Segmentation and alignment of parallel text for statistical machine translation. *Natural Language Engineering*, *13*(3), 235–260.

Denkowski, M., & Lavie, A. (2011). Meteor 1.3: Automatic metric for reliable optimization and evaluation of machine translation systems. *Proceedings of the Sixth Workshop on Statistical Machine Translation*, 85–91.

Destaw Belay, T., Lambebo Tonja, A., Kolesnikova, O., Muhie Yimam, S., Ali Ayele, A., Bogale Haile, S., Sidorov, G., & Gelbukh, A. (2022). The effect of normalization for bi-directional amharic-english neural machine translation. *Ieeexplore.Ieee.Org*. https://doi.org/10.5281/zenodo.3734260

Dorr, B. J., Jordan, P. W., & Benoit, J. W. (1999). A survey of current paradigms in machine translation. In *Advances in computers* (Vol. 49, pp. 1–68). Elsevier.

Duan, S., Zhao, H., & Wang, R. (2020). Graph-to-Sequence Neural Machine Translation. *ArXiv Preprint ArXiv:2009.07489*.

Dwivedi, S. K., & Sukhadeve, P. P. (2010). Machine translation system in indian perspectives. *Journal of Computer Science*, *6*(10), 1111.

El Maazouzi, Z., El Mohajir, B. E., & Al Achhab, M. (2017). A systematic reading in statistical translation: From the statistical machine translation to the neural translation models. *Journal of Information and Communication Technology*, *16*(2), 408.

Eriguchi, A., Tsuruoka, Y., & Cho, K. (2017). Learning to parse and translate improves neural machine translation. *ArXiv Preprint ArXiv:1702.03525*.

Ganesh, S., Dhotre, V., Patil, P., & Pawade, D. (2023). A Comprehensive Survey of Machine Translation Approaches. *2023 6th International Conference on Advances in Science and Technology (ICAST)*, 160–165.

Gashaw, I., & Shashirekha, H. L. (2019). Amharic-Arabic Neural Machine Translation. *Arxiv.Org*. http://arxiv.org/abs/1912.13161

Gashaw, I., & Shashirekha, H. L. (2020a). CONSTRUCTION OF AMHARIC-ARABIC PARALLEL TEXT CORPUS FOR NEURAL MACHINE TRANSLATION. *Researchgate.Net*, *11*(1). https://doi.org/10.5121/ijaia.2020.11107

Gashaw, I., & Shashirekha, H. L. (2020b). *Machine Learning Approaches for Amharic Parts-of-speech Tagging*. http://arxiv.org/abs/2001.03324

Getachew, S., & Yayeh, Y. (2023). Gex'ez-English Bi-Directional Neural Machine Translation Using Transformer. *2023 International Conference on Information and Communication Technology for Development for Africa (ICT4DA)*, 160–164.

Gezmu, A., Low, A. N.-A. T. on A. and, & 2023, undefined. (2023). Morpheme-Based Neural Machine Translation Models for Low-Resource Fusion Languages. *Dl.Acm.Org*, *22*(231). https://doi.org/10.1145/3610773

Gezmu, A. M., Nürnberger, A., & Bati, T. B. (2021). Neural Machine Translation for Amharic-English Translation. *ICAART (1)*, 526–532.

Gezmu, A. M., Nürnberger, A., & Bati, T. B. (2022). Extended Parallel Corpus for Amharic-English Machine Translation. *2022 Language Resources and Evaluation Conference, LREC 2022*, 6644–6653.

Gezmu, A., Nürnberger, A., arXiv:2104.03543, T. B. preprint, & 2021, undefined. (2022). Extended parallel corpus for Amharic-English machine translation. *Arxiv.Org*. https://doi.org/10.24352/ub.ovgu-2018-145

Gong, L., Li, Y., Guo, J., Yu, Z., & Gao, S. (2022). Enhancing low-resource neural machine translation with syntax-graph guided self-attention. *Knowledge-Based Systems*, *246*, 108615.

Gu, J., Awadalla, H. H., Devlin, J., & Li, V. O. K. (2018). Universal neural machine translation for extremely low resource languages. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, 344–354.

Gupta, A., Venkatapathy, S., & Sangal, R. (2010). METEOR-Hindi: automatic MT evaluation metric for hindi as a target. *Proceedings of ICON-2010: 8th International Conference on Natural Language Processing, Macmillan Publishers. India*.

Hadgu, A. T., Aregawi, A., & Beaudoin, A. (2022). Lesan–machine translation for low resource languages. *NeurIPS 2021 Competitions and Demonstrations Track*, 297–301.

Hadgu, A. T., Beaudoin, A., & Aregawi, A. (2020a). Evaluating Amharic Machine Translation. *Arxiv.Org*. http://arxiv.org/abs/2003.14386

Hadgu, A. T., Beaudoin, A., & Aregawi, A. (2020b). Evaluating amharic machine translation. *ArXiv Preprint ArXiv:2003.14386*.

Hao, J., Tian, Y., Zhang, Q., & Zhang, G. (2022). Mongolian-Chinese Machine Translation Based on Graph Neural Network. *Journal of Physics: Conference Series*, *2400*(1), 012050.

Hashimoto, K., & Tsuruoka, Y. (2017). Neural machine translation with source-side latent graph parsing. *ArXiv Preprint ArXiv:1702.02265*.

Hassan, H., & Darwish, K. (2014). Statistical machine translation. In *Natural Language Processing of Semitic Languages* (pp. 199–219). Springer.

Hearne, M., & Way, A. (2011). Statistical machine translation: a guide for linguists and translators. *Language and Linguistics Compass*, *5*(5), 205–226.

Hu, Y., Chen, C., Yang, C.-H. H., Li, R., Zhang, D., Chen, Z., & Chng, E. S. (2024). Gentranslate: Large language models are generative multilingual speech and machine translators. *ArXiv Preprint ArXiv:2402.06894*.

Hutchins, J. (2001). Machine translation and human translation: in competition or in complementation. *International Journal of Translation*, *13*(1–2), 5–20.

IBRAHIM, A. A., Mohammed, T. A., & DARA, O. N. (2024). *Predicting Big data Drug Interactions and associated side effects by Using Artificial Neural Networks (ANN) over Traditional Graph Convolutional Networks (GCNs)*.

Imani, A., Severini, S., Sabet, M. J., Yvon, F., & Schütze, H. (2022). Graph-based multilingual label propagation for low-resource part-of-speech tagging. *ArXiv Preprint ArXiv:2210.09840*.

Iyer, V., Malik, B., Stepachev, P., Chen, P., Haddow, B., & Birch, A. (2024). Quality or quantity? On data scale and diversity in adapting large language models for low-resource translation. *ArXiv Preprint ArXiv:2408.12780*.

Jassem, K., & Dwojak, T. (2019). Statistical versus neural machine translation–a case study for a medium size domain-specific bilingual corpus. *Poznan Studies in Contemporary Linguistics*, *55*(2), 491–515.

Joshi, A., Dabre, R., Kanojia, D., Li, Z., Zhan, H., Haffari, G., & Dippold, D. (2024). Natural Language Processing for Dialects of a Language: A Survey. *ACM Computing Surveys*, *1*(6), 149. https://doi.org/10.1145/3712060/ASSET/65FEC2BE-7302-4101-8B8E-A03788E31EE9/ASSETS/GRAPHIC/CSUR-2024-0027-F03.JPG

Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., Cowan, B., Shen, W., Moran, C., & Zens, R. (2007). Moses: Open source toolkit for statistical machine translation. *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions*, 177–180.

Koehn, P., & Knowles, R. (2017). Six challenges for neural machine translation. *ArXiv Preprint ArXiv:1706.03872*.

Kore, M., Technol, V. G.-Int. J. Comput. Sci. T., & 2017, undefined. (2017). Machine transliteration for English to Amharic proper nouns. *Academia.Edu*. https://www.academia.edu/download/53884037/IJCST-V5I4P5.pdf

Kumar, L. A., Renuka, D. K., Rajakumari, R. G., Ahmed, R. A., Nihmathullah, M. M., Kumar, S. N., & Rohith, V. (2024). Spoken Language Translation using Transformer model. *2024 6th International Conference on Computational Intelligence and Networks (CINE)*, 1–6.

Lakew, S. M., Negri, M., & Turchi, M. (2020). *Low Resource Neural Machine Translation: A Benchmark for Five African Languages*. http://arxiv.org/abs/2003.14402

Lavie, A., & Denkowski, M. J. (2009). The METEOR metric for automatic evaluation of machine translation. *Machine Translation*, *23*(2), 105–115.

Li, Q., Wong, D. F., Chao, L. S., Zhu, M., Xiao, T., Zhu, J., & Zhang, M. (2018). Linguistic knowledge-aware neural machine translation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *26*(12), 2341–2354.

Lin, D., Sun, S., Ding, J., Ke, X., Gu, H., Huang, X., Song, C., Zhang, X., Yi, L., & Wen, J. (2022). PlatoGL: Effective and scalable deep graph learning system for graph-enhanced real-time recommendation. *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, 3302–3311.

Lin, G. H., & Chien, P. S. C. (2009). Machine Translation for Academic Purposes. *Online Submission*.

Lison, P., & Tiedemann, J. (2016). *Opensubtitles2016: Extracting large parallel corpora from movie and tv subtitles*.

Liu, B., & Wu, L. (2022). Graph neural networks in natural language processing. In *Graph neural networks: foundations, frontiers, and applications* (pp. 463–481). Springer.

Lopez, A. (2008). Statistical machine translation. *ACM Computing Surveys (CSUR)*, *40*(3), 1–49.

Luong, M.-T., Pham, H., & Manning, C. D. (2015). Effective approaches to attention-based neural machine translation. *ArXiv Preprint ArXiv:1508.04025*.

Mahesh, K., & Nirenburg, S. (1996). Meaning representation for knowledge sharing in practical machine translation. *Proe. the FLAIRS-96 Track on Information Interchange*.

Marcheggiani, D., Bastings, J., & Titov, I. (2018). Exploiting semantics in neural machine translation with graph convolutional networks. *ArXiv Preprint ArXiv:1804.08313*.

Marcheggiani, D., & Titov, I. (2017). Encoding sentences with graph convolutional networks for semantic role labeling. *ArXiv Preprint ArXiv:1703.04826*.

Mekonnen Gezmu, A., Nürnberger, A., & Bayu Bati, T. (2021). Extended Parallel Corpus for Amharic-English Machine Translation. *ArXiv E-Prints*, arXiv-2104.

Melese, M., Besacier, L., & Meshesha, M. (2017). Amharic-english speech translation in tourism domain. *Proceedings of the Workshop on Speech-Centric Natural Language Processing*, 59–66.

Moussallem, D., Wauer, M., & Ngomo, A.-C. N. (2018). Machine translation using semantic web technologies: A survey. *Journal of Web Semantics*, *51*, 1–19.

Negia, M. K., Tamiru, R. M., & Meshesha, M. (2023). Morpheme Based Amharic-Kistanigna Bi-directional Machine Translation Using Deep Learning. *Communications in Computer and Information Science*, *1800 CCIS*, 245–257. https://doi.org/10.1007/978-3-031-31327-1_14

Nguyen, L. H. B., Pham, V., & Dinh, D. (2020). Integrating AMR to neural machine translation using graph attention networks. *2020 7th NAFOSTED Conference on Information and Computer Science (NICS)*, 158–162.

Nguyen, T., Nguyen, L., Tran, P., & Nguyen, H. (2021). Improving Transformer-Based Neural Machine Translation with Prior Alignments. *Complexity*, *2021*(1), 5515407.

Nigusie, G. (2024). Suffix inflectional morphology generation for Amharic text. *5th Workshop on African Natural Language Processing*.

Nirenburg, S., Somers, H. L., & Wilks, Y. (2003). *Readings in machine translation*. MIT Press.

Nzeyimana, A. (2024). Low-resource neural machine translation with morphological modeling. *ArXiv Preprint ArXiv:2404.02392*.

Och, F. J., & Ney, H. (2002). Discriminative training and maximum entropy models for statistical machine translation. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, 295–302.

Papineni, K., Roukos, S., Ward, T., & Zhu, W.-J. (2002). Bleu: a method for automatic evaluation of machine translation. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, 311–318.

Popescu-Belis, A. (2019). Context in neural machine translation: A review of models and evaluations. *ArXiv Preprint ArXiv:1901.09115*.

Popovic, M., & Ney, H. (2006). POS-based Word Reorderings for Statistical Machine Translation. *LREC*, 1278–1283.

Prasada, P., & Rao, M. V. P. (2024). Reinforcement of low-resource language translation with neural machine translation and backtranslation synergies. *Int J Artif Intell ISSN*, *2252*(8938), 3479.

Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). *Improving language understanding by generative pre-training*. https://www.mikecaptain.com/resources/pdf/GPT-1.pdf

Ragni, V., & Nunes Vieira, L. (2022). What has changed with neural machine translation? A critical review of human factors. *Perspectives*, *30*(1), 137–158.

Riezler, S., Vasserman, A., Tsochantaridis, I., Mittal, V. O., & Liu, Y. (2007). Statistical machine translation for query expansion in answer retrieval. *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, 464–471.

Ruiz, L., Gama, F., & Ribeiro, A. (2020). Gated graph recurrent neural networks. *IEEE Transactions on Signal Processing*, *68*, 6303–6318.

Saengthongpattana, K., Kriengket, K., Porkaew, P., & Supnithi, T. (2019). Thai-english and english-thai translation performance of transformer machine translation. *2019 14th International Joint Symposium on Artificial Intelligence and Natural Language Processing (ISAI-NLP)*, 1–5.

Sakre, M. M. (2019). Machine translation status and its effect on business. *Journal of the ACS*, *10*, 33–48.

Shadiev, R., Sun, A., & Huang, Y. (2019). A study of the facilitation of cross-cultural understanding and intercultural sensitivity using speech-enabled language translation technology. *British Journal of Educational Technology*, *50*(3), 1415–1433.

Simard, M., Goutte, C., & Isabelle, P. (2007). Statistical phrase-based post-editing. *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference*, 508–515.

Simard, M., & Isabelle, P. (2009). Phrase-based machine translation in a computer-assisted translation environment. *Proceedings of Machine Translation Summit XII: Papers*.

Singh, M. T., Borgohain, R., & Gohain, S. (2014). An English-assamese machine translation system. *International Journal of Computer Applications*, *93*(4).

Smirnov, A. V, Teslya, N., Shilov, N., Frank, D., Minina, E., & Kovacs, M. (2022). Comparative Analysis of Neural Translation Models based on Transformers Architecture. *ICEIS (1)*, 586–593.

Specia, L. (2010). Translating from complex to simplified sentences. *International Conference on Computational Processing of the Portuguese Language*, 30–39.

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. *Advances in Neural Information Processing Systems*, *27*.

Talapatra, A., Boluki, S., Honarmandi, P., Solomou, A., Zhao, G., Ghoreishi, S. F., Molkeri, A., Allaire, D., Srivastava, A., & Qian, X. (2019). Experiment design frameworks for accelerated discovery of targeted materials across scales. *Frontiers in Materials*, *6*, 82.

Tan, G. (2025). NAH-GNN: A graph-based framework for multi-behavior and high-hop interaction recommendation. *PloS One*, *20*(4), e0321419.

Tan, Z., Wang, S., Yang, Z., Chen, G., Huang, X., Sun, M., & Liu, Y. (2020). Neural machine translation: A review of methods, resources, and tools. *AI Open*, *1*, 5–21.

Tato, A. (2018). *Improving adam optimizer*.

Tedla, Y., on, K. Y.-2016 I. C., & 2016, undefined. (2016). The effect of shallow segmentation on English-Tigrinya statistical machine translation. *Ieeexplore.Ieee.Org*. https://ieeexplore.ieee.org/abstract/document/7875939/

Terumasa, E. (2007). Rule based machine translation combined with statistical post editor for japanese to english patent translation. *Proceedings of the Workshop on Patent Translation*.

Tesfaye, S. (2020). *A Hybrid approach for Machine Translation from Ge'ez to Amharic language*. http://repository.smuc.edu.et/handle/123456789/6235

Teshome, E. (2013). *Bidirectional English-Amharic machine translation: an experiment using constrained corpus*. http://thesisbank.jhia.ac.ke/6064/

Teshome, M., Besacier, L., Taye, G., 2015, D. T.-A., & 2015, undefined. (2015). Phoneme-based English-Amharic statistical machine translation. *Ieeexplore.Ieee.Org*. https://ieeexplore.ieee.org/abstract/document/7331921/

Teshome, M. G., & Besacier, L. (2012a). Preliminary experiments on English-Amharic statistical machine translation. *Spoken Language Technologies for Under-Resourced Languages*.

Teshome, M. G., & Besacier, L. (2012b). Preliminary experiments on English-Amharic statistical machine translation. *Spoken Language Technologies for Under-Resourced Languages*.

Teshome, M. G., Besacier, L., Taye, G., & Teferi, D. (2015a). Phoneme-based English-Amharic statistical machine translation. *AFRICON 2015*, 1–5.

Teshome, M. G., Besacier, L., Taye, G., & Teferi, D. (2015b). Phoneme-based English-Amharic statistical machine translation. *AFRICON 2015*, 1–5.

Tomita, M., & Carbonell, J. G. (1986). Another stride towards knowledge-based machine translation. *Coling 1986 Volume 1: The 11th International Conference on Computational Linguistics*.

Torregrosa, D., Pasricha, N., Masoud, M., Chakravarthi, B. R., Alonso, J., Casas, N., & Arcan, M. (2020). Aspects of terminological and named entity knowledge within rule-based machine translation models for under-resourced neural machine translation scenarios. *ArXiv Preprint ArXiv:2009.13398*.

Tracey, J., & Strassel, S. (2020). Basic language resources for 31 languages (plus English): The LORELEI representative and incident language packs. *Proceedings of the 1st Joint Workshop on Spoken Language Technologies for Under-Resourced Languages (SLTU) and Collaboration and Computing for Under-Resourced Languages (CCURL)*, 277–284.

Tran, Q.-D. (2024). Exploring Low-Resource Machine Translation: Case Study of Lao-Vietnamese Translation. *2024 International Conference on Multimedia Analysis and Pattern Recognition (MAPR)*, 1–6.

Tu, M., Huang, J., He, X., & Zhou, B. (2020). Graph sequential network for reasoning over sequences. *ArXiv Preprint ArXiv:2004.02001*.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, *30*.

Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., & Bengio, Y. (2017). Graph attention networks. *Stat*, *1050*(20), 10–48550.

Vo, H. N. K., Le, D. D., Phan, T. M. D., Nguyen, T. S., Pham, Q. N., Tran, N. O., Nguyen, Q. D., Vo, T. M. H., & Quan, T. (2024). Revitalizing bahnaric language through neural machine translation: challenges, strategies, and promising outcomes. *Proceedings of the AAAI Conference on Artificial Intelligence*, *38*(21), 23360–23368.

Wang, Q., Li, B., Xiao, T., Zhu, J., Li, C., Wong, D. F., & Chao, L. S. (2019). Learning deep transformer models for machine translation. *ArXiv Preprint ArXiv:1906.01787*.

Wibawa, K. A. (2019). Fragmentation of the thinking structure of translation in solving mathematical modelling problems. *Southeast Asian Mathematics Education Journal*, *9*(1), 25–36.

Wilks, Y. (2008). *Machine translation: Its scope and limits*. Springer Science & Business Media.

Wu, L., Chen, Y., Shen, K., Guo, X., Gao, H., Li, S., Pei, J., & Long, B. (2023). Graph neural networks for natural language processing: A survey. *Foundations and Trends® in Machine Learning*, *16*(2), 119–328.

Xia, M., Huang, G., Liu, L., & Shi, S. (2019). Graph based translation memory for neural machine translation. *Proceedings of the AAAI Conference on Artificial Intelligence*, *33*(01), 7297–7304.

Xu, K., Wu, L., Wang, Z., Feng, Y., Witbrock, M., & Sheinin, V. (2018). Graph2seq: Graph to sequence learning with attention-based neural networks. *ArXiv Preprint ArXiv:1804.00823*.

Xu, M., Li, L., Wong, D. F., Liu, Q., & Chao, L. S. (2021). Document graph for neural machine translation. *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 8435–8448.

Yang, Y., Li, X., Jiang, T., Kong, J., Ma, B., Zhou, X., & Wang, L. (2017). Improving adversarial neural machine translation with prior knowledge. *2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 1373–1377.

Zhang, J., & Zong, C. (2020). Neural machine translation: Challenges, progress and future. *Science China Technological Sciences*, *63*(10), 2028–2050.

Zhang, T., Huang, H., Feng, C., & Wei, X. (2020). Similarity-aware neural machine translation: reducing human translator efforts by leveraging high-potential sentences with translation memory. *Neural Computing and Applications*, *32*(23), 17623–17635.

Brief Biodata and Publications

Name: Muluken Hussen Asebel

From: Haramaya University

Designation: PhD Candidate, Adama Science and Technology University

Academic details: BSc and MSc in Computer Science from University of Gondar

Journals:

1 Hussen Asebel, M., Getu Assefa, S., & Abebe Haile, M. Exploring the evolution and future prospects of Amharic to English machine translation: a systematic review. Frontiers in Artificial Intelligence, 8, 1456245.

2 Asebel, M. H., Assefa, S. G., Haile, M. A., 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.

3 Chala, T. D., Guta, A. C., & Asebel, M. H. (2022). Design and development of a text-to speech synthesizer for Afan Oromo. SN Computer Science, 3(5), 420.