2025 3rd International Conference on Advances in Computation, Communication and Information Technology (ICAICCIT)

A Transformer-Based Framework for Domain-Sensitive Amharic–English Neural Machine Translation with Character-Aware Subword Encoding

1st Syed Rizwana
Dept. of Computer Science and Engineering
Narasaraopeta Engineering College
Narasaraopet, India
syedrizwananrt@gmail.com

2nd Mallela Sai Jyothi
Dept. of Computer Science and Engineering
Narasaraopeta Engineering College
Narasaraopet, India
malleasaijyothi@gmail.com

3rd Duddu Divya
Dept. of Computer Science and Engineering
Narasaraopeta Engineering College
Narasaraopet, India
Duddudivya06@gmail.com

4th Dasari Srujana Sankeerthi
Dept. of Computer Science and Engineering
Narasaraopeta Engineering College
Narasaraopet, India
apsaappushaik@gmail.com

5th V. Lakshmi
Department of Computer Science and Engineering
GRIET,
Telangana, India
lakshmi1682@grietcollege.com

6th Yerraparu Sravani Devi
Dept. of Computer Science and Engineering
G. Narayanamma Institute of Technology & Science (Women)
Telangana, India
y.sravanidevi@gnits.ac.in

7th Dodda Venkata Reddy
Department of Computer Science and Engineering
Narasaraopeta Engineering College
Narasaraopet, India
doddavenkatareddy@gmail.com

***Abstract*—An NMT framework that has been adapted to the domain of low-resourced and morphologically complex languages has been developed for translating from Amharic to English. Character-aware subword tokenization via SentencePiece and the Tanzil corpus allows the system to work around rare and compound words. A trained Transformer encoder-decoder model with multihead attention and feed-forward subnets achieved 59.03 BLEU score on the Tanzil corpus compared to the 26.08 RNN with attention baseline. Integration of domain parallel data with underwent sub- word modeling and the design and development of a low- resource reproducible Transformer pipeline in addition to the consolidation of methodologically relevant parallel data for Amharic to English translation serves as primary and novel contributions of this research. Through the translation, domain adaptation and subword level segmentation pair and the results for the translated text confirms the Amharic to English translation underwent a boosted performance which speaks to the level of improvement of the translation model. This serves as a groundwork as an entry point for further research on new areas pertaining to the neural architectures of machine translations of Semitic and other languages of complicated morphology.**

***Index Terms*—Neural Machine Translation, Transformer, Amharic, Religious Texts, Character-Level Embedding, BLEU Score.**

I. Introduction

Recent advancements in Neural Machine Translation (NMT) have revolutionized the paradigm of language translation from rule-based and phrase-based statistical approaches to fully data-driven neural structures [1], [2]. One of the critical steps in this progress was the publication of the Transformer model by Vaswani et al. in 2017 [3], which employed a self-attention mechanism over recurrent frameworks. This design innovation enabled efficient parallel processing of sequences and led to major breakthroughs in a number of natural language processing applications. However, the quality of these models rests squarely on abundant good-quality bilingual data, which reduces their potential benefit in linguistically poor regions [4], [5].

One such low-resource environment is Ethiopia's official working language, Amharic. Being a morphologically rich Semitic language written with the Ge'ez script and an agglutinative grammar, Amharic poses particular difficulties for machine translation systems. The morphological richness gives rise to a huge vocabulary space and high out-of-vocabulary (OOV) occurrence rate, causing problems with standard word-level tokenization [6], [7]. Moreover, general-purpose systems are most likely to underperform when adapted to specific domains such as religious literature, in which semantic closeness and stylistic appropriateness must be preserved [8].

Due to the limitations presented above, interest has been growing in subword-level modeling approaches and character-level representations for improving the quality of translations in morphologically complex languages. SentencePiece, one of the most widely used unsupervised tokenizers, enables learning of subword units from raw text data without the need for language-specific preprocessing tools [9], [10]. Other Semitic and African languages such as Oromo and Tigrinya have existing work which proves that subword tokenization significantly alleviates OOV errors and enhances model generalization in low-resource settings [11], [12].

In addition to preprocessing improvements, domain-specialized fine-tuning of NMT models is currently a primary method of bridging the performance gap between specialized and general-purpose translation. Domain-specific training on targeted corpora—legal, biomedical, religious—improves semantic correctness and contextual understanding [13], [14]. However, collecting parallel data in such domains remains problematic, especially for languages like Amharic where public resources are scarce. To address this limitations, this work introduces an entire NMT pipeline targeted at Amharic religious translation into English. Our approach employs a Transformer-based encoder–decoder model trained on a carefully filtered and cleaned subset of Tanzil, a publicly available corpus of aligned religious texts in Amharic and English.

In addition to preprocessing improvements, domain-specialized fine-tuning of NMT models is currently a primary method of bridging the performance gap between specialized and general-purpose translation. Domain-specific training on targeted corpora—legal, biomedical, religious—improves semantic correctness and contextual understanding [13], [14].

Collecting parallel data in specialized domains remains challenging, particularly for low-resource languages such as Amharic. This study introduces a neural machine translation pipeline for Amharic-to-English religious text, employing a Transformer encoder–decoder trained on a cleaned subset of the Tanzil corpus. The system incorporates SentencePiece subword tokenization, dropout regularization, and tuned hyperparameters for low-resource conditions. Experimental results show a BLEU score of 59.03, significantly surpassing a baseline RNN with attention (26.08).

The key contributions of this paper are as follows:

- Construction of a domain-aligned parallel corpus for Amharic–English translation in the religious context.
- Introduction of subword-level tokenization to address vocabulary and morphological challenges.
- Development of a reproducible Transformer model tailored for domain-sensitive translation in low-resource environments.

Section II presents a review of related literature. While Section IV includes the discussion of the experiment results and the comparative analysis, Section III outlines the datasets and the suggested methodology. Lastly, Section V presents the conclusion, highlighting the main findings and proposing avenues for further investigation.

II. Related Work

Neural machine translation studies for low-resource and morphologically rich languages have grown significantly with focus on subword tokenization, domain adaptation, and architecture simplification. Transformer model [3] is the present workhorse of NMT, but its performance in low-resource scenarios is limited without adequate data preparation and tokenization methods [4], [5].

Tokenization is a critical element in NMT, particularly for very inflectionally diverse languages like Amharic. SentencePiece [8], a language-independent subword tokenizer, has achieved great success in low-resource translation with the use of the versatility of character-level models and the effectiveness of subword-level segmentation. Such techniques reduce the OOV problem and have a better capture of morphological patterns. Charformer and hybrid tokenization models have also demonstrated the power of character-aware representation to enhance translation robustness [13], [14].

Amharic-to-English translation research in particular is limited, even though there has been tremendous effort. Gezmu et al. [4] experimented with varying Transformer configurations and segmentation methods for Amharic-to-English translation with BLEU scores above 32 using subword methods. Belay and Assabie [7] extended the results further with a fine-tuned multilingual pre-trained model after homophone normalization. The research proved the power of domain-specific adaptation to improve fluency as well as semantic quality.

Morphological modeling techniques have also been explored. Gezmu and Nu¨rnberger [6] introduced MorphoSeg, a morpheme-level segmentation scheme that was tailored specifically for Amharic. Their experiments demonstrated morpheme-aware models to be superior to wordpiece or BPE segmentation, particularly with synthetic data from the Contemporary Amharic Corpus (CACO). This underlines the importance of morphology-sensitive preprocessing for Semitic language translation.

In other general African language settings, cross-lingual embeddings and multilingual transfer learning have been used to enhance the performance of translation in low-resource settings [9], [10]. Although these techniques also require access to large-scale multilingual data or pre-trained models, which may not be present in low-resource deployments.

Dropout regularization, so prevalent in deep learning, has also been used to successfully prevent overfitting in NMT models, especially in low-resource environments [15]. While techniques such as Elastic Weight Consolidation [16] and cross-lingual consistency training [17] have been proposed, this paper limits regularization to standard dropout for simplicity and reproducibility.

Domain-specific NMT has been effective when applied to corpora of specialized text. For instance, Amrhein and Sennrich enhanced coherence in legal translation with the use of domain-aligned data [18]. Similarly, well-curated data on

Quranic translations have enabled the conduct of meaningful research in religious domain NMT for Arabic and Persian. However, to our knowledge, no previous research addresses Amharic-English translation under the religious domain using a character-aware Transformer model.

Asefa and Assabie [22] proposed a Transformer-based Amharic-to-English NMT model using character-aware embeddings and combined regularization (dropout, label smoothing, L2). Their approach improves translation accuracy for low-resource, morphologically rich languages.

Reproducibility and benchmarking are valuable aspects of NMT work. Use of the same evaluation methods such as BLEU and SacreBLEU [21], and open-source preprocessing pipelines, ensures fair comparison and reproducibility of results. Reproducibility of a comparison between a Transformer and an RNN + Attention model is based on these best practices.

Overall, while all three areas of subword modeling, morphological segmentation, and domain-specific translation have seen enhancements, few studies combine all three for low-resource Semitic languages. Our suggested framework is the first to merge these techniques within an integrated, high-quality pipeline specifically tailored for Amharic-English religious text translation.

III. MATERIALS AND METHODS

The proposed Amharic–English NMT system adopts a Transformer-based encoder–decoder architecture, optimized for low-resource and morphologically rich translation tasks. This section elaborates on the dataset, preprocessing, tokenization, model architecture, and training procedure, as shown in Fig. 1.

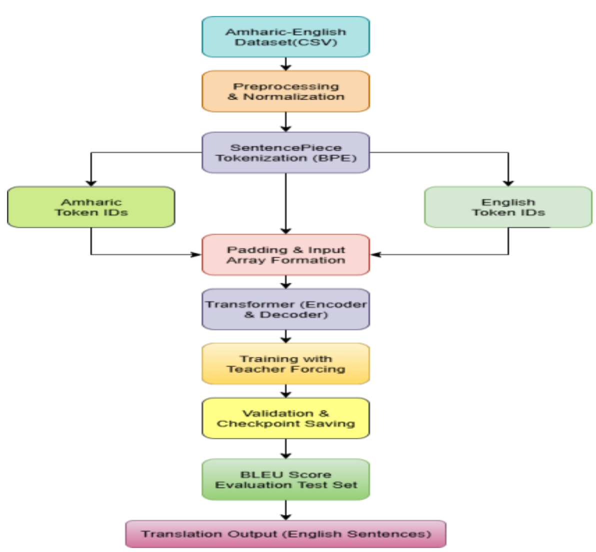


Fig. 1. Workflow of the proposed Amharic-to-English translation system using Transformer architecture.

A. Experimental Setup

Experiments were conducted on Google Colab using a free Tesla T4 GPU (16 GB VRAM) with 12 GB RAM. The environment utilized Python 3.10 along with TensorFlow 2.18, SentencePiece, and NumPy libraries. A custom Transformermodel has been trained on the Tanzil Amharic–English data set after cleaning it. It was divided into 80% training data, 10% validation data, and 10% test data. Each language has been trained with a dedicated, independently constructed SentencePiece tokenizer with a vocabulary size of 8,000.The model has an architecture of 2 encoder-decoder layers, 8 attention heads, 128 hidden units, and 1 feedforward layer of 512 units. Training utilized the Adam optimizer with a learning rate of 10^{-4}. The model trained with a batch size of 64 and employed early stopping. The model's performance was assessed using BLEU scoring on the test set illuminated by the learning.

B. Dataset Description

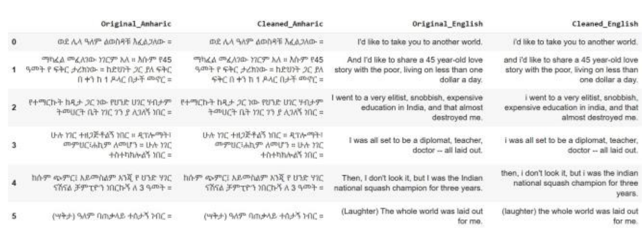
The OPUS Tanzil corpus [11] was used, containing 25,197 aligned Amharic–English sentence pairs derived from Quranic text. After filtering duplicates, misalignments, and malformed entries, 23,790 clean sentence pairs were retained. The final dataset split was as follows:

- **Training**: 19,032 pairs (80%)
- **Validation**: 2,379 pairs (10%)
- **Test**: 2,379 pairs (10%)

All data was tokenized using SentencePiece and padded to a fixed sequence length before being fed to the model.

C. Data Preprocessing

The preprocessing pipeline included removal of empty or malformed sentence pairs, trimming, normalization, and character filtering. Amharic punctuation like ":" and "::" was removed, while English text was lowercased with appropriate spacing. Sentences shorter than 2 or longer than 100 tokens were discarded. Only sentences with valid Ethiopic and English alphabetic characters were retained. The data was randomized to eliminate ordering bias is shown in Fig. 2.



	Original_Amharic	Cleaned_Amharic	Original_English	Cleaned_English
0	ወደ ሌላ ዓለም ልወስዳቹ እፈልጋለው ።	ወደ ሌላ ዓለም ልወስዳቹ እፈልጋለው ።	I'd like to take you to another world.	i'd like to take you to another world.
1	ማካፈል መፈለገው ነገርም አለ ። እሱም የ45 ዓመት የ ፍቅር ታሪክነው ። ከደሀነት ጋር ያለ ፍቅር በ ቀን ከ 1 ዶላር በታች መኖር ።	ማካፈል መፈለገው ነገርም አለ ። እሱም የ45 ዓመት የ ፍቅር ታሪክነው ። ከደሀነት ጋር ያለ ፍቅር በ ቀን ከ 1 ዶላር በታች መኖር ።	And I'd like to share a 45 year-old love story with the poor, living on less than one dollar a day.	and i'd like to share a 45 year-old love story with the poor, living on less than one dollar a day.
2	የተማርኩት ከዲታ ጋር ነው የሀንድ ሀገር ሃብታም ትመሀርት ቤት ነገር ገን ያ ሊገለኝ ነበር ።	የተማርኩት ከዲታ ጋር ነው የሀንድ ሀገር ሃብታም ትመሀርት ቤት ነገር ገን ያ ሊገለኝ ነበር ።	I went to a very elitist, snobbish, expensive education in India, and that almost destroyed me.	i went to a very elitist, snobbish, expensive education in india, and that almost destroyed me.
3	ሁሉ ነገር ተዘጋጅቶልኝ ነበር ። ዲፕሎማት፣ መምህር፣ሐኪም ለመሆን ። ሁሉ ነገር ተስተካክሎልኝ ነበር ።	ሁሉ ነገር ተዘጋጅቶልኝ ነበር ። ዲፕሎማት፣ መምህር፣ሐኪም ለመሆን ። ሁሉ ነገር ተስተካክሎልኝ ነበር ።	I was all set to be a diplomat, teacher, doctor -- all laid out.	i was all set to be a diplomat, teacher, doctor -- all laid out.
4	ከሱም ጭምር፤ አይመስልም እንጂ የ ሀንድ ሃገር ናሽናል ቻምፒዮን ነበርኩኝ ለ 3 ዓመት ።	ከሱም ጭምር፤ አይመስልም እንጂ የ ሀንድ ሃገር ናሽናል ቻምፒዮን ነበርኩኝ ለ 3 ዓመት ።	Then, I don't look it, but I was the Indian national squash champion for three years.	then, i don't look it, but i was the indian national squash champion for three years.
5	(ሣቅታ) ዓለም በጠቃላይ ተሰታኝ ነበር ።	(ሣቅታ) ዓለም በጠቃላይ ተሰታኝ ነበር ።	(Laughter) The whole world was laid out for me.	(laughter) the whole world was laid out for me.

Fig. 2. Illustration of text transformation before and after preprocessing.

D. Tokenization

SentencePiece was used for subword tokenization, with separate models for Amharic and English (vocab size =

8,000). This mitigates the out-of-vocabulary issue and enhances generalization. Each sentence was converted into subword units before being input into the model is shown in Fig. 3.

	Language	Original Sentence	Tokenized Output
0	Amharic	ያንን እርሱ ገሀነምን ገቢ የሆነውን ሰው ቢሆን እንጅ	_ያንን _እርሱ _ገሀነምን _ገቢ _የሆነውን _ሰው _ቢሆን _እንጅ
1	English	except him who shall roast in the blazing fire .	_except _him _who _shall _roast _in _the _blaz...
2	Amharic	በእዝነትህም ከከሓዲዎች ሕዝቦች አድነን	_በእዝነትህም _ከከሓዲዎች _ሕዝቦች _አድነን
3	English	and deliver us , through your mercy , from the...	_and _deliver _us _, _through _your _mercy _, ...
4	Amharic	(አላህን) የሚፈራ ሰው በእርግጥ ይገሠጻል	_(_አላህን _) _የሚፈራ _ሰው _በእርግጥ _ይገሠጻል
5	English	he who fears will mind ,	_he _who _fears _will _mind _,
6	Amharic	መከራይቱ በወደቀች ጊዜ	_መከራይቱ _በወደቀች _ጊዜ
7	English	when the event (the resurrection) comes	_when _the _event _(_the _resurrection _) _comes

Fig. 3. Tokenization samples for Amharic and English text using SentencePiece.

E. Model Architecture

We propose a Transformer-based model specifically designed for morphologically rich and resource-scarce translation is shown in Fig. 4. The architecture leverages :

- Subword tokenization via SentencePiece
- Two-layer Transformer encoder and decoder
- Multi-head attention with 8 heads
- Feed-forward networks (512 units, ReLU activation)
- Dropout (rate = 0.1) and layer normalization
- Dense + Softmax output layer

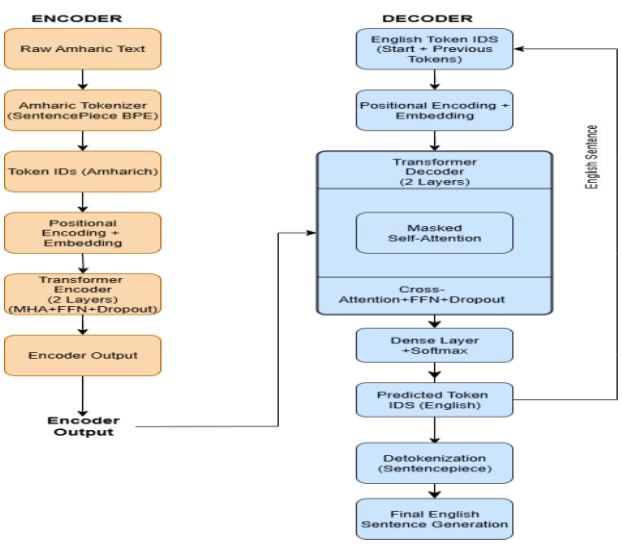


Fig. 4. Proposed Transformer-based Amharic–English NMT system.

1) Input Representation: Subword tokens are embedded and combined with positional encodings, then passed to the encoder.

2) Encoder Design: Each encoder block consists of:

- Multi-head self-attention (8 heads)
- Feed-forward network (512 units)
- Layer normalization and residual connections
- Dropout (0.1)

3) Decoder Design: The decoder incorporates:

- Masked self-attention
- Cross-attention with encoder outputs
- Similar FFN, dropout, and normalization layers
- Teacher forcing during training and greedy decoding at inference

4) Output and Training: Output tokens are predicted using a dense layer with softmax activation. Training used sparse categorical cross-entropy loss, with Adam optimizer (lr = 10^{-4}), batch size 64, and early stopping to avoid overfitting.

F. Regularization

Layer normalization and dropout (0.1) were applied to enhance generalization. The system was designed for reproducibility with a preference for simplicity over architectural complexity.

IV. Results and Discussion

A. Model Evaluation Metrics

The Bilingual Evaluation Understudy (BLEU) score was used as the primary metric for translation quality. BLEU-4 scoring with Smoothing Method 4 was adopted to mitigate penalties from short or morphologically rich outputs, as often encountered in low-resource languages. The BLEU score is computed as in (1),

$$BLEU = BP \cdot exp(\sum_{n=1}^{N} w_n \log\Big(p_n\Big)) \tag{1}$$

where p_n denotes the modified n-gram precision, w_n is the weight assigned to each n-gram (typically $1/N$), Using this metric, the model achieved a BLEU score of 59.03 on the test set, indicating strong generalization across Amharic sentences within religious contexts. Validation loss was monitored throughout training, with early stopping applied once the loss plateaued, thereby reducing overfitting.

B. Training and Validation Analysis

The training followed the sequence-to-sequence framework capturing the essence of the teacher forcing strategy. In sequence training, the Adam optimizer is advocate using a learning rate of 1×10^{-4}, a batch size of 64, and for 15 epochs. The loss function employed is sparse categorical cross-entropy. The data was processed using TensorFlow's `tf.data.Dataset` API with the goal of prefetching for the sake of effective operation.

They generated the decoder input by shifting target sequences to the right. The decoder input is then given by target sequences aligned beside shifted sequences. The .keras format contained the most productive check points from validation performance.

Training was executed on Google Colab. The subject was endowed with the NVIDIA Tesla T4 GPU (16 GB VRAM). The duration for every epoch was 12 minutes, and by the 15 epoch, validation loss was 0.0605. This is shown in Fig. 5 for the sake of illustrating convergence behavior.

As shown in Figs. 5 and 6, both the loss and BLEU score plots confirm model stability and consistent improvement. The absence of divergence between training and validation curves suggests effective generalization despite the limited dataset.

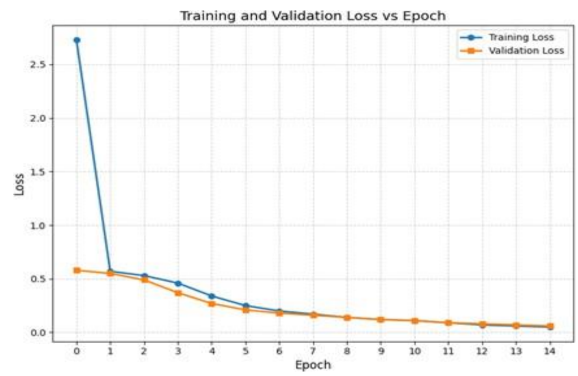


Fig. 5. Training and validation loss vs. epochs.

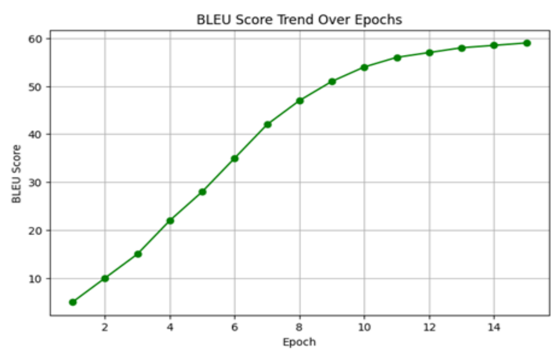


Fig. 6. BLEU score progression across epochs.

C. Comparative Performance

The Transformer model was evaluated against a baseline RNN + Attention model and the model by Asefa and Assabie(2025) [22]. Table I shows the BLEU scores and Fig. 7 shows the Comparison of baseline and proposed models in BLEU, loss, and accuracy.

TABLE I
BLEU Score Comparison

Model	BLEU Score
RNN + Attention	26.08
Proposed Transformer	59.03
Asefa , Assabie (2025)	40.59

Results show that the proposed Transformer outperforms both the baseline and Asefa and Assabie's model, demonstrating the effectiveness of the domain-specific adaptations and subword-tokenization strategies.

D. Qualitative Translation Evaluation

In addition to quantitative metrics, manual evaluation of translations revealed that the Transformer generated fluent, contextually accurate sentences—especially for domain-specific religious phrases,Sample Amharic–English translation outputs is shown in Fig. 8.

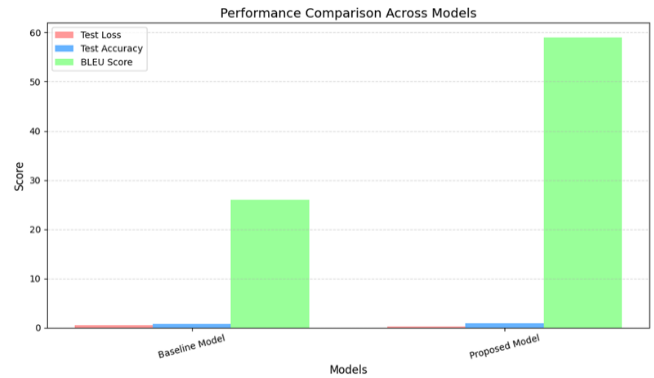


Fig. 7. Comparison of baseline and proposed models in BLEU, loss, and accuracy.

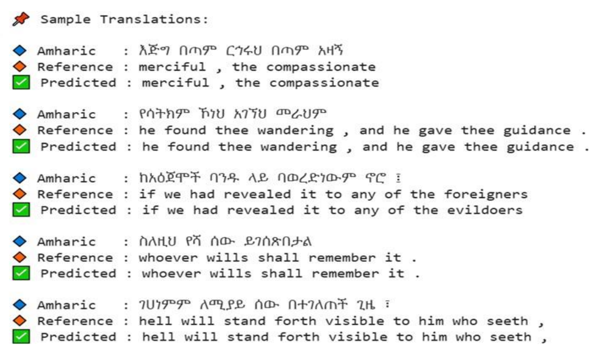


Fig. 8. Sample Amharic–English translation outputs.

The model maintained semantic and stylistic consistency and produced grammatically correct outputs aligned with domain vocabulary the obtained outputs of amharic to english translation .

E. Discussion and Insights

The results highlight the importance of subword-level tokenization and domain adaptation. SentencePiece effectively reduced out-of-vocabulary (OOV) issues and handled complex Amharic morphology. Although greedy decoding was used, translation fluency was sufficient due to strong corpus alignment. Compared to the RNN baseline, the Transformer architecture showed clear advantages in handling context and structure, affirming its suitability for low-resource neural translation tasks.

V. Conclusion

This study developed an Amharic-to-English neural machine translation (NMT) system based on the Transformer encoder-decoder architecture, with an emphasis on domain adaptation and performance under low-resource settings. To tackle the morphological complexity and sparse vocabulary characteristic of Amharic, subword-level tokenization was implemented using SentencePiece, enabling more effective learning in sparse-data scenarios.

The model underwent training on a refined portion of the Tanzil corpus and then underwent rigorous preprocessing along with the adoption of domain-related methods. It then evaluated the central RNN+Attention against a BLEU score of 59.03 on the test set and outperformed it quite easily. Key architectural choices such as positional encodings, dropout regularization, and multi-head attention proved effective in capturing the long-range dependencies of religious texts. Moreover, convergence and generalization were facilitated with the use of teacher forcing and early stopping techniques.

Despite these strengths, the decoding method used—greedy decoding—was somewhat limiting, as it may yield suboptimal translations in certain contexts. Learned constraints were applied during inference, but not directly integrated into the model. Furthermore, regularization strategies such as label smoothing, attention dropout, or scheduled sampling were not explored due to computational constraints.

Future work will explore these methods, alongside beam search decoding to enhance output fluency. The inclusion of multilingual pretraining or transfer learning from related Semitic languages is also a promising avenue for performance gains. Incorporating external linguistic features like POS tags, syntactic trees, and morphological analyzers could further improve alignment and grammatical quality.

While the current study focused on the religious domain, future directions include integrating domain-specific corpora from education, healthcare, and news. This expansion will enable a broader assessment of the model's adaptability and support more accessible Amharic machine translation for educational, governmental, and humanitarian applications.

REFERENCES

[1] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," in *Proc. Int. Conf. Learn. Representations (ICLR)*, 2015. [Online]. Available: https://arxiv.org/abs/1409.0473

[2] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint* arXiv:1301.3781, 2013. [Online]. Available: https://arxiv.org/abs/1301.3781

[3] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017. [Online]. Available: https://papers.nips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

[4] G. Gezmu, A. Nürnberger, and S. Yamamoto, "Subword-based transformer NMT models for Amharic-English," *arXiv preprint* arXiv:2203.00567, 2022. [Online]. Available: https://arxiv.org/abs/2203.00567

[5] M. Artetxe and H. Schwenk, "Massively multilingual sentence embeddings for zero-shot cross-lingual transfer and beyond," *Transactions of the Association for Computational Linguistics*, vol. 7, pp. 597–610, 2019. [Online]. Available: https://aclanthology.org/Q19-1042.pdf

[6] G. Gezmu and A. Nürnberger, "MorphoSeg: Morphologically-aware segmentation for Amharic NMT," in *Proc. Int. Conf. Language Resources and Evaluation (LREC)*, 2022. [Online]. Available: https://aclanthology.org/2022.lrec-1.124.pdf

[7] B. Belay and Y. Assabie, "Improving Amharic–English machine translation using homophone normalization," *arXiv preprint* arXiv:2302.10934, 2023. [Online]. Available: https://arxiv.org/abs/2302.10934

[8] T. Kudo and J. Richardson, "SentencePiece: A simple and language independent subword tokenizer and detokenizer for neural text processing," in *Proc. Conf. Empirical Methods in Natural Language Processing (EMNLP)*, 2018. [Online]. Available: https://aclanthology.org/D18-2012.pdf

[9] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, and V. Stoyanov, "Unsupervised cross-lingual representation learning at scale," in *Proc. Annu. Meeting Assoc. Comput. Linguistics (ACL)*, 2020. [Online]. Available: https://aclanthology.org/2020.acl-main.747.pdf

[10] S. Baniata, A. Al-Khalifa, A. Al-Saqqa, M. Al-Kabi, and A. Tawalbeh, "Arabic dialects to modern standard Arabic NMT with transformer and subword units," in *Proc. Annu. Meeting Assoc. Comput. Linguistics (ACL)*, 2023. [Online]. Available: https://aclanthology.org/2023.acl-long.25.pdf

[11] Tanzil Project, "Tanzil: Quranic translation corpus," [Online]. Available: https://opus.nlpl.eu/Tanzil.php

[12] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997. [Online]. Available: https://www.bioinf.jku.at/publications/older/2604.pdf

[13] Y. Zhang, H. Chen, Y. Liu, and M. Zhou, "Character-level representations improve translation for morphologically rich languages," in *Proc. Annu. Meeting Assoc. Comput. Linguistics (ACL)*, 2020. [Online]. Available: https://aclanthology.org/2020.acl-main.430.pdf

[14] Y. Liu, J. Gu, N. Goyal, X. Li, S. Edunov, M. Ghazvininejad, M. Lewis, and L. Zettlemoyer, "Multilingual denoising pre-training for neural machine translation," *Transactions of the Association for Computational Linguistics*, vol. 8, pp. 726–742, 2020. [Online]. Available: https://aclanthology.org/2020.tacl-1.47.pdf

[15] H. Zhou, X. Ma, W. Huang, C. Zhuang, and H. Yang, "Understanding dropout in deep learning," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2022. [Online]. Available: https://arxiv.org/abs/2205.13055

[16] D. Variš and O. Bojar, "EWC for low-resource neural machine translation," *arXiv preprint* arXiv:2301.00251, 2023. [Online]. Available: https://arxiv.org/abs/2301.00251

[17] J. Gao, M. Li, H. Wang, and L. Ma, "Cross-lingual consistency regularization for multilingual neural machine translation," in *Proc. Conf. Empirical Methods in Natural Language Processing (EMNLP)*, 2023. [Online]. Available: https://aclanthology.org/2023.emnlp-main.115.pdf

[18] M. Wu, J. Zhang, Y. Sun, and T. Yang, "UniDrop: Structured dropout with uniform sampling," in *Proc. Int. Conf. Learn. Representations (ICLR)*, 2023. [Online]. Available: https://openreview.net/forum?id=xKzcY0kAQ9G

[19] A. Fan, E. Grave, and A. Joulin, "LayerDrop: Structured dropout for transformer models," *arXiv preprint* arXiv:1909.11556, 2019. [Online]. Available: https://arxiv.org/abs/1909.11556

[20] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, and A. Rush, "Transformers: State-of-the-art natural language processing," in *Proc. Conf. Empirical Methods in Natural Language Processing (EMNLP): System Demonstrations*, pp. 38–45, 2020. [Online]. Available: https://aclanthology.org/2020.emnlp-demos.6.pdf

[21] T. Kocmi and O. Bojar, "An exploration of sentence embeddings for neural machine translation," *arXiv preprint* arXiv:2001.03156, 2020. [Online]. Available: https://arxiv.org/abs/2001.03156

[22] S. H. Asefa and Y. Assabie, "Transformer-based Amharic-to-English machine translation with character embedding and combined regularization techniques," *IEEE Access*, vol. 13, pp. 1089–1106, 2025. doi: 10.1109/ACCESS.2025.10813169. [Online]. Available: https://ieeexplore.ieee.org/document/10813169