WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Syntactic Reordering for Myanmar-to-English Sentence Translation with Adjective Phrases Using Neural Machine Translation

Aye Thida Win[1], Phyo Phyo Wai[2], Yimon Aung[3]

[1]Associate Professor, Polytechnic University (Dawei), Myanmar
[2]Associate Professor, Computer University (Magway), Myanmar
[3]Associate Professor, Polytechnic University (Dawei), Myanmar

A B S T R A C T

Myanmar and English differ significantly in syntactic structure, especially in adjective-noun ordering: Myanmar follows [Adjective + Noun] order, whereas English often uses [Determiner + Adjective + Noun]. This paper presents a neural machine translation (NMT) approach enhanced with syntactic reordering rules and attention-based models to accurately translate Myanmar sentences containing adjectives into grammatically correct English sentences. A combination of rule-based preprocessing and Transformer-based learning ensures the syntactic reordering is preserved. Evaluation results show significant improvements in BLEU scores compared to baseline NMT models. We also analyze linguistic challenges and demonstrate solutions with detailed examples.

Keywords: Myanmar NLP, Syntax Reordering, Adjective Phrases, Neural Machine Translation, Rule-Based Analysis, Transformer

1. Introduction

In the field of Natural Language Processing (NLP), machine translation (MT) between structurally different languages remains a persistent challenge, particularly when one of the languages is low-resource and morphologically rich like **Myanmar (Burmese)**. Myanmar, a Sino-Tibetan language, has a flexible syntactic structure, heavily relies on postpositions and particles, and generally follows a **Subject-Object-Verb (SOV)** word order. In contrast, **English**, a Germanic language, exhibits a **Subject-Verb-Object (SVO)** word order with strict syntactic constraints, especially concerning adjective and noun phrase formation. A particularly notable linguistic difference between Myanmar and English is the **placement and sequencing of adjectives in noun phrases**. In Myanmar, adjectives typically precede the noun directly with little morphological variation and often in flexible order. For example:

- မြင့်မားသော အဆောက်အဦး)Lit: tall building)
- လှပသော မိန်းမ)Lit: beautiful woman)

While English also uses pre-nominal adjectives (e.g., "a tall building"), the **ordering and structure** are more rigid. English follows a defined **adjective order hierarchy** (opinion > size > age > shape > colour > origin > material > purpose > noun), which must be respected to produce fluent and natural output. Moreover, in English, articles and determiners (e.g., "a", "the") are mandatory in many contexts, while Myanmar sentences often omit equivalent forms. These differences often result in **word-order divergence** during machine translation. A direct or naive neural translation from Myanmar to English can produce output that is grammatically incorrect, semantically distorted, or unnatural. For instance, a phrase such as:

- **Myanmar:** ကြီးမားသော အိမ်အနားမှာ
- **Raw English Translation**: "Near the big house" may become distorted if the adjective "big" is misaligned or misplaced in the translation process.

This issue becomes more complex when **multiple adjectives** are stacked or when **adjective clauses** are embedded, as Myanmar syntax permits recursive nesting and flexible placement of modifiers. For example:

- မြင်ကွင်းလှသော ကြီးမားသည့် အိမ်)Lit: beautiful-view big house) should translate as: "the big house with a beautiful view" or "a large house with a beautiful view", depending on context and modifier scope.

Given this challenge, this research addresses the problem of accurately translating **Myanmar sentences containing adjectives** into **English sentences with correct syntactic order**, focusing particularly on adjective-noun phrase reordering. The core hypothesis of this paper is that **syntactic reordering as a pre-processing step**, combined with **neural machine translation (NMT)**, can significantly improve the fluency and grammatical correctness of English output.

This study contributes in several ways:

1. **Linguistic Analysis**: We provide a detailed comparison of adjective ordering in Myanmar and English, supported by syntactic rules.
2. **Reordering Module**: We design a rule-based syntactic reordering module that pre-processes Myanmar input sentences, transforming adjective-noun structures into a canonical order more suitable for translation.
3. **Neural Model Integration**: We integrate the reordered input into a Transformer-based NMT model and train it on an adjective-focused Myanmar-English parallel corpus.
4. **Evaluation and Analysis**: We conduct a comprehensive evaluation using BLEU scores, human judgment, and error analysis to assess the improvement in translation quality, particularly in adjective phrases.

The motivation behind this research is both practical and linguistic. On the one hand, improving adjective phrase translation enhances the **readability, accuracy, and naturalness** of machine-generated English text from Myanmar. On the other hand, this task sheds light on broader issues in **cross-linguistic syntactic divergence**, which is crucial for building robust NLP systems for underrepresented languages. As Myanmar continues to digitize and globalize, the need for high-quality language tools becomes critical in fields such as **education, diplomacy, legal affairs, and content localization**. We believe that tackling core linguistic problems—like syntactic reordering in adjective-rich sentences—is essential to unlocking the full potential of Myanmar-to-English translation systems.

2. Background and Related Work

Myanmar (Burmese) is an agglutinative, morphologically rich, and low-resource language with a flexible Subject-Object-Verb (SOV) structure. In contrast, English follows a Subject-Verb-Object (SVO) structure and relies heavily on function words and fixed word order. These contrasts lead to challenges in machine translation, particularly with syntactic phenomena such as **adjective-noun ordering**. Myanmar allows multiple adjectives to appear before a noun in a flexible order, whereas English enforces a strict hierarchical sequence (Dixon, 1982). This syntactic divergence often results in erroneous or awkward English output if not handled correctly during translation. Early Rule-Based Machine Translation (RBMT) systems used handcrafted linguistic rules, but they lacked scalability (Hutchins & Somers, 1992). Statistical Machine Translation (SMT), such as phrase-based SMT (Koehn et al., 2003), improved robustness by learning probabilistic alignments from parallel corpora. However, SMT systems typically underperform in long-distance dependency modeling and reordering, especially with structurally divergent language pairs like Myanmar-English (Win & Than, 2021). Neural Machine Translation (NMT) introduced significant advances, particularly with the attention-based sequence-to-sequence models (Bahdanau et al., 2015; Luong et al., 2015). More recently, the Transformer model (Vaswani et al., 2017) has become the de facto architecture in NMT. However, for low-resource languages, Transformers alone cannot fully capture deep syntactic structures without supplementary guidance (Neubig & Hu, 2018). Studies also show that NMT models often produce grammatically incorrect translations when dealing with adjective ordering (Bentivogli et al., 2016). English follows a well-known adjective order, generally described as Opinion > Size > Age > Shape > Color > Origin > Material > Purpose > Noun (Larson, 1998). Violations of this order result in unnatural-sounding phrases (e.g., *a red round small ball*). In contrast, Myanmar permits relatively free placement of adjectives without affecting semantic acceptability, which further complicates direct translation without reordering. To address syntactic divergence, researchers have proposed **pre-translation syntactic reordering**. For instance, Xia and McCord (2004) used tree transformations to align source structures with English. Habash and Sadat (2006) applied morphological and syntactic reordering to improve Arabic-English translation. Similarly, Collins et al. (2005) explored syntax-based SMT with reordering for German-English pairs. Such techniques significantly improved translation quality when combined with statistical or neural models. In Myanmar-English translation, only limited syntactic resources are available. Myint et al. (2018) explored SMT alignments but did not incorporate structural reordering. To the best of our knowledge, no prior work has explicitly addressed adjective phrase reordering in Myanmar-English NMT. To overcome data scarcity, several hybrid approaches have been proposed. Currey et al. (2017) demonstrated that copying source sentences into the target side during training can improve low-resource translation. Lakew et al. (2018) proposed morphology-aware NMT for under-resourced African languages. Multilingual approaches, such as those by Guzmán et al. (2019), have also proven effective for improving translation quality in low-resource settings by leveraging cross-lingual transfer. Our work builds on these foundations by proposing a **rule-based syntactic reordering module** tailored for Myanmar adjective structures. We integrate this module with a Transformer-based NMT system, enhancing the model's ability to produce grammatically correct English translations with properly ordered adjective phrases.

3. Core Foundations of the Proposed System

This research builds upon two key pillars that are tightly integrated to address the specific challenges posed by Myanmar-to-English translation of adjective-containing sentences:

3.1. Linguistic Reordering Rules Specific to Myanmar Syntactic Structures

Myanmar and English differ fundamentally in the way adjectives are structured and ordered in a sentence. Myanmar adjectives are more flexible in positioning and are often context-driven, while English requires a strict, rule-governed ordering of adjectives before a noun. In this paper, we define and implement **rule-based syntactic transformations** that are specific to Myanmar grammar in order to prepare the source sentences for more effective translation. This component is **linguistically motivated** and follows these guiding principles:

- **Adjective Position Identification**: Using Myanmar Part-of-Speech (POS) tagging and syntactic chunking, adjectives preceding nouns are detected.
- **Semantic Classification**: Each adjective is mapped to a semantic class (e.g., opinion, size, color) using a predefined lexicon and semantic rule base.
- **Adjective Ordering Rule Enforcement**: The order of adjectives is rearranged to match the canonical English sequence (opinion > size > age > shape > color > origin > material > purpose).
- **Syntactic Reordering Template**: Myanmar phrases such as "အလှပြည့်သော ကြီးမားသည့် အိမ်" (Lit: beautiful large house) are reordered to reflect "a large beautiful house" in English, following proper adjective hierarchy.

This component acts as a **preprocessing module** that modifies the Myanmar sentence structure into a more "English-friendly" word order before passing it to the translation model. By doing so, it reduces the syntactic learning burden on the neural model, especially in low-resource settings.

3.2. Transformer-Based Neural Machine Translation with Reordering-Aware Input Pre processing

The second foundation of the system leverages the power of **Transformer-based Neural Machine Translation (NMT)** models, known for their capacity to capture long-range dependencies and global sentence structure through self-attention mechanisms (Vaswani et al., 2017). However, because Myanmar-English translation involves high syntactic divergence and limited parallel corpora, we enhance the Transformer model using a **reordering-aware preprocessing strategy**:

- **Input Sentence Reordering**: The reordered Myanmar sentences (output from the rule-based module) are used as source input for training and inference, ensuring that the model sees syntactically pre-aligned inputs.
- **Standard Transformer Architecture**: We use a standard 6-layer Transformer encoder-decoder model with multi-head attention and positional encoding.
- **Alignment Enhancement via Attention**: The reordering helps the model learn clearer alignments between source and target tokens, particularly in the adjective-noun structures, which are often mistranslated when presented in raw Myanmar order.
- **Training on Augmented Data**: In addition to raw sentence pairs, reordered versions of sentences are added to the training data to increase exposure to syntactic variants and improve generalization.

This hybrid approach enables the Transformer model to **focus more on lexical selection, fluency, and coherence**, while offloading the burden of deep structural realignment to the rule-based reordering component.

4. Proposed Framework

To address the syntactic divergence between Myanmar and English, particularly in adjective phrase translation, we propose a **hybrid translation framework** that combines **linguistic rule-based syntactic reordering** with a **Transformer-based Neural Machine Translation (NMT)** model. The framework is designed to optimize syntactic alignment prior to translation while leveraging the contextual understanding capabilities of deep learning models. The overall architecture consists of the following stages:

1. **Input Sentence Parsing**: Myanmar sentences are first passed through a Part-of-Speech (POS) tagger and syntactic chunker to identify the structure, including adjective-noun phrases.
2. **Linguistic Reordering Module**: Based on linguistic rules derived from English adjective order grammar, Myanmar adjective sequences are reordered to match English expectations.
3. **Preprocessed Sentence Input to NMT**: The reordered Myanmar sentence is then fed into a Transformer-based NMT system trained to map syntactically aligned Myanmar sentences to fluent English output.
4. **Post processing and Evaluation**: The output sentence is post-processed, evaluated using BLEU scores, and compared to baseline outputs.

The framework ensures that syntactic reordering is handled explicitly and systematically **before** translation, enabling the neural model to focus on semantics and fluency, rather than structural transformation.

5. Methodology

This section outlines the detailed design and implementation of the two major components in the proposed system: the rule-based reordering module and the Transformer-based translation model.

5.1. Linguistic Reordering Rules Module

5.1.1 Motivation

Adjective order in English follows a strict hierarchy (e.g., “a beautiful large old red Japanese wooden house”), while Myanmar allows flexible adjective positioning. Translating directly from Myanmar without reordering leads to output such as “a red Japanese wooden house beautiful large” — grammatically incorrect and semantically confusing.

5.1.2 Syntactic Analysis and Phrase Identification

Using a Myanmar-specific POS tagger and syntactic chunker, the system identifies all **[Adjective + Noun]** and **multi-adjective chains**. For example:

- Input sentence: လှပပြီး ကြီးမားသော တောင်တန်းများ
- Identified structure: ADJ1 = လှပ (beautiful), ADJ2 = ကြီးမား (large), NOUN = တောင်တန်းများ (mountains)

5.1.3 Adjective Classification and Ordering

We assign each adjective a **semantic category** (Opinion, Size, Age, etc.) using a handcrafted Myanmar–English adjective lexicon.

Table 1: Adjective classification and ordering of Myanmar and English

Myanmar Adjective	**English Equivalent**	**Class**
လှပ	beautiful	Opinion
ကြီးမား	large	Size
ရှေးဟောင်း	ancient	Age

The system then applies English adjective ordering rules and reorders the Myanmar adjectives accordingly.

5.1.4 Sentence Reconstruction

The reordered phrase is reinserted into the original sentence. For example:

- Original Myanmar: လှပပြီး ကြီးမားသော တောင်တန်းများ
- Reordered: ကြီးမားပြီး လှပသော တောင်တန်းများ
- Translation-ready: “the large beautiful mountains”

This reordering is **language-neutral**, meaning it improves NMT quality without changing the meaning of the Myanmar sentence.

6. Transformer-Based Neural Machine Translation (NMT)

6.1. Model Architecture

We adopt the standard **Transformer** architecture (Vaswani et al., 2017) implemented using the OpenNMT-py framework. The model consists of:

- **6-layer encoder** and **6-layer decoder**
- **Multi-head attention (8 heads)**
- **Positional encoding**
- **Dropout regularization (0.3)**

6.2. Data Preparation

A parallel corpus of 10,000 manually aligned Myanmar-English sentence pairs was created. The corpus includes a high proportion of adjective-rich sentences from:

- Myanmar literature

- News articles
- Open-source language datasets

For training, we prepare **two sets of inputs**:

1. **Original Myanmar Sentences**
2. **Reordered Myanmar Sentences** (from the Reordering Module)

The model is trained on the reordered inputs to learn **better alignment** between reordered Myanmar syntax and correct English output.

Table 2: Training Parameters

Parameter	**Value**
Optimizer	Adam
Learning rate	0.0005
Batch size	64
Epochs	30
Tokenization	SentencePiece (BPE)
Evaluation Metric	BLEU, Human Review

7. Integration of Components

The complete pipeline is shown below the two-stage system design makes it easier to explain and interpret errors, fine-tune rule-based components separately, and improves the **grammatical correctness** of translated English sentences—especially in the context of adjective-noun structures.

Myanmar Input Sentence

↓

POS Tagging + Chunking

↓

Linguistic Reordering (adjective normalization)

↓

Reordered Myanmar Sentence

↓

Transformer NMT Model

↓

English Output Sentence (with correct adjective order)

Fig 1. System Architecture of System

8. Evaluation and Analysis

The motivation behind this paper is both practical and linguistic. A comprehensive evaluation using BLEU scores, human judgment, and error analysis to assess the improvement in translation quality, particularly in adjective phrases. On the one hand, improving adjective phrase translation enhances the readability, accuracy, and naturalness of machine-generated English text from Myanmar. On the other hand, this task sheds light on broader issues in cross-linguistic syntactic divergence, which is crucial for building robust NLP systems for underrepresented languages. As Myanmar continues to digitize and globalize, the need for high-quality language tools becomes critical in fields such as education, diplomacy, legal affairs, and content localization. We believe that tackling core linguistic problems like syntactic reordering in adjective-rich sentences is essential to unlocking the full potential of Myanmar-to-English translation systems. We use BLEU, TER, and human judgment to evaluate fluency and correctness.

Table 3: Analysis of System

Model	**BLEU**	**TER**	**Human Score (1-5)**
Baseline NMT	22.5	51.3	2.9
Reordering + NMT	**31.8**	**41.7**	**4.1**

Example 1

Input (Myanmar): လှပသော အိမ်

Rule Output: [ADJ=beautiful] + [NOUN=house]

English: A beautiful house

Example 2

Input (Myanmar): ကြီးမားသော အနုပညာသမားကြီး

Rule Output: [ADJ=great, ADJ=big] + [NOUN=artist]

English: A great big artist

9. Conclusion and Future Work

Combining rule-based linguistic knowledge with deep learning improves translation quality in low-resource, high-divergence language pairs. Reordering before translation helps the model learn better alignments and reduces the burden on the decoder to fix structure. This framework could be adapted to other agglutinative and SOV languages with syntactic differences from English. We presented a hybrid syntactic reordering approach for translating adjective-containing Myanmar sentences into English. By integrating rule-based preprocessing with NMT, we achieved better fluency and structural accuracy. Future work includes:

- Expanding to more syntactic constructions (e.g., relative clauses)
- Training with larger corpora
- Applying BERT-style pre-training for Myanmar.

References

1. Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. *International Conference on Learning Representations (ICLR)*.
2. Bentivogli, L., Bisazza, A., Cettolo, M., & Federico, M. (2016). Neural versus phrase-based machine translation quality: A case study. *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 257–267.
3. Collins, M., Koehn, P., & Kučerová, I. (2005). Clause restructuring for statistical machine translation. *Proceedings of ACL 2005*, 531–540.
4. Currey, A., Miceli Barone, A. V., & Heafield, K. (2017). Copied monolingual data improves low-resource neural machine translation. *Proceedings of the Second Conference on Machine Translation*, 148–156.
5. Dixon, R. M. W. (1982). *Where have all the adjectives gone? And other essays in semantics and syntax.* Mouton Publishers.
6. Guzmán, F., Chen, P., Ott, M., et al. (2019). The FLORES evaluation datasets for low-resource machine translation: Nepali–English and Sinhala–English. *EMNLP 2019*, 6073–6081.
7. Habash, N., & Sadat, F. (2006). Arabic preprocessing schemes for statistical machine translation. *Proceedings of NAACL-HLT 2006*, 49–52.
8. Hutchins, J., & Somers, H. (1992). *An Introduction to Machine Translation.* Academic Press.
9. Koehn, P., Och, F. J., & Marcu, D. (2003). Statistical phrase-based translation. *NAACL 2003*, 48–54.
10. Larson, R. K. (1998). Events and modification in nominals. *Proceedings of Semantics and Linguistic Theory (SALT)*, 145–168.
11. Lakew, S. M., Gawlitza, L., & Negri, M. (2018). Transfer learning in multilingual neural machine translation with dynamic vocabulary. *Proceedings of the 15th International Conference on Spoken Language Translation (IWSLT)*.
12. Luong, M.-T., Pham, H., & Manning, C. D. (2015). Effective approaches to attention-based neural machine translation. *Proceedings of EMNLP 2015*, 1412–1421.

13. Myint, N. N., Tun, T. T., & Win, A. (2018). Statistical machine translation from Myanmar to English using phrase-based approach. *Journal of Computer Science and Technology*, 18(3), 79–89.

14. Neubig, G., & Hu, J. (2018). Rapid adaptation of neural machine translation to new languages. *Proceedings of EMNLP 2018*, 875–880.

15. Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention is all you need. *Advances in Neural Information Processing Systems (NeurIPS)*, 5998–6008.

16. Win, S. T., & Than, T. T. (2021). POS tagging and chunking for Myanmar language using BiLSTM-CRF. *Proceedings of the 2021 Workshop on Asian Language Resources (ALR)*.

17. Xia, F., & McCord, M. (2004). Improving a statistical MT system with automatically learned rewrite patterns. *Proceedings of COLING 2004*, 508–514.