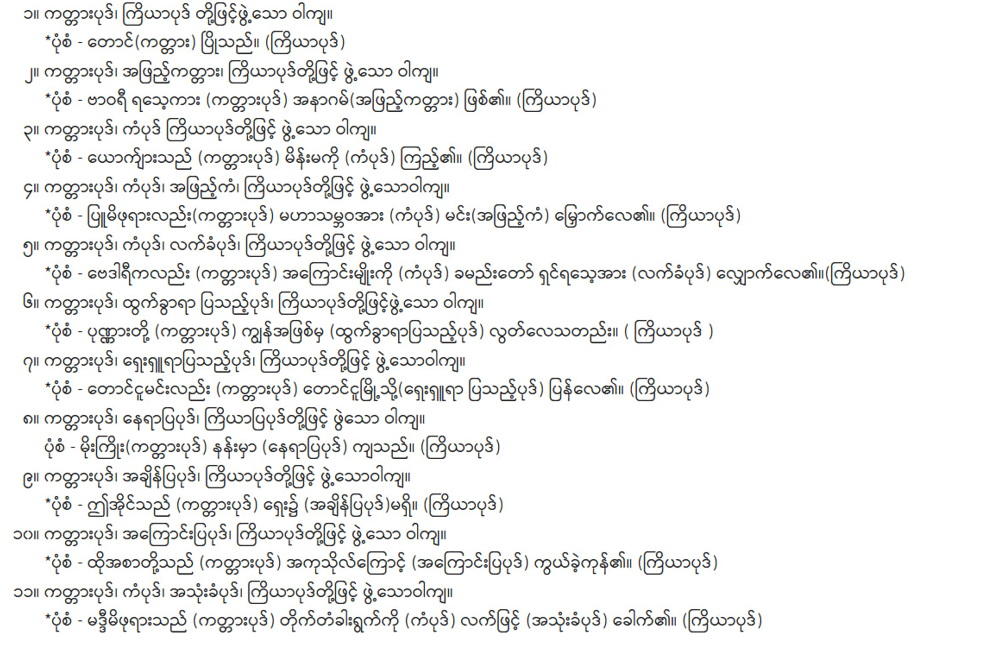
1. Problem Statement

The Burmese language is an underrepresented language in the field of NLP. Its script presents unique challenges due to ambiguities in complex grammar rules, and the lack of standardized word segmentation. These challenges are especially evident in digital communication, where users often make usage and typing errors. Unlike widely studied languages like English, Burmese lacks reliable tools for automatic sentence correction.

This project aims to develop a deep learning-based model that can automatically transform incorrect Burmese sentences into correct ones. The correction will focus on grammar structure, and word segmentation, with the scope limited to simple sentence patterns (ဝါကျရိုး) from the figure shown below, forms limited to affirmative, negative, question, with formal past, present, and future tenses. Solving this problem will improve digital text processing in Burmese and contribute to the advancement of Burmese NLP tools, enabling better educational and linguistic applications, especially for Burmese as a second language learners.





**Figure 1** – Simple Sentence Patterns Scope (Wikipedia, 2023)

1. Example of input, output

Input – correct/incorrect Burmese sentence

Output – correct Burmese sentence

1. ကျောင်း မှ စောင့်နေပါ။ -> ကျောင်းက စောင့်နေပါ။
2. သူတိုသည် ဒီမိုက‌ရေစီရေးအတွက် တိုက်ကြသည်။ -> သူတိုသည် ဒီမိုကရေစီရရှိရေးအတွက် တိုက်ကြသည်။
3. သူအိမ်ကို ထွက်သွားတယ်။ ->ﾠသူအိမ်မှ ထွက်သွားတယ်။
4. ကျောင်းသွား ပြင်ဆင်တယ်။-> ကျောင်းသွားဖို့ ပြင်ဆင်တယ်။
5. စာဖတ်မယ် ပြီးကျောင်းသွားမယ်။ -> စာဖတ်ပြီး ကျောင်းသွားမယ်။
6. ငါ နံနက်ကို လာမယ်။ -> ငါ နံနက်မှာ လာမယ်။
7. သူ အားနည်းခြင်းဖြင့် လဲကျတယ်။ -> သူ အားနည်းခြင်းကြောင့် လဲကျတယ်။
8. သူလည်း မနက်စာလည်း မစားဘူး။ -> သူ မနက်စာလည်း မစားဘူး။
9. သူ မ သွားမယ်။ -> သူ မသွားဘူး။
10. မနေ့က သူ လာတယ်။ -> မနေ့က သူ လာခဲ့တယ်။
11. Dataset – minimum 1000 rows
12. Expected outcome

Model performance goal: >=60%

Evaluation metrics:

* 1. Accuracy score
  2. BLEU/ROUGE

**Reference**

Wikipedia. (2023, May 10). ဝါကျ: တည်းဖြတ်မှု ရာဇဝင်. Retrieved August 2, 2025, from https://my.wikipedia.org/wiki/%E1%80%9D%E1%80%AB%E1%80%80%E1%80%BB#%E1%80%80%E1%80%BC%E1%80%AD%E1%80%9A%E1%80%AC%E1%80%90%E1%80%85%E1%80%BA%E1%80%81%E1%80%AF%E1%80%9B%E1%80%BE%E1%80%AD%E1%80%9E%E1%80%B1%E1%80%AC\_%E1%80%9D%E1%80%AB%E1%80%80%E1%80%BB%E1%80%96%E1%80%BD%E1%80%B2%E1%80%B7%E1%80%94%E1%80%8A%E1%80%BA%E1%80%B8%E1%80%99%E1%80%BB%E1%80%AC%E1%80%B8