Name: Kenny Patel

Roll no. : 19074010

Lab Task-6: Isolated Word Morphological Generation

Morphological Generation

Morphology is the study of formation of words, i.e. how words are build up from smaller pieces or how they are structured and formed.

Morphemes are the the most important part of Morphology. Word forms are analysed as arrangements of Morphemes. Hence, Morphemes are the smallest linguistic unit with grammatical function.

Eg: unhappiness : un - prefix ; happy - stem ; ness - suffix

un, happy, ness --> Morphemes

Morphological Generation is the inverse of morphological anslysis, namely the process of converting the internal representation of a word to its surface form.

For eg, if our internal representations of the words mice and move are:

- Root: MOUSE PartOfSpeech:Noun Number:Plural
- Root: MOVE PartOfSpeech:Verb Tense:Past then morphological generation would convert these to the character strings mice and moved.

Directory Structure

- 19074010-Kenny-Patel (Main)
    - Cross-lingual-Inflection
        - Hindi-Bengali
            - Data
            - Model
            - Jupyter Notebook
        - German-Middle High German
            - Same as Hindi-Bengali
    - Single-Language-Morph-Generation
        - Hindi
            - Data
            - Result
            - Model
        - Bengali
            - Same as Hindi
        - German
            - Same as Hindi
        - Middle High German
            - Same as Hindi

Cross Lingual Inflection

For this, I cloned the following repo:

https://github.com/sigmorphon/crosslingual-inflection-baseline.git and trained the data on Hindi dataset(available in data folder) and tested it on low resource Bengali data.

Same was done with German and Middle High German German being High resource language.

- Command used: `!bash scripts/run-task1-tag.sh soft hindi--bengali`

Results:

- Hindi-Bengali :
  - Accuracy : 48%
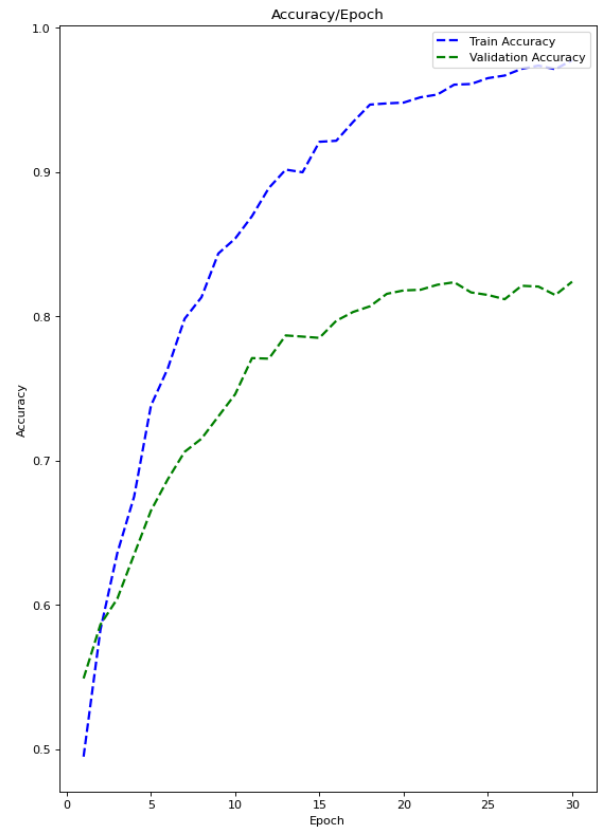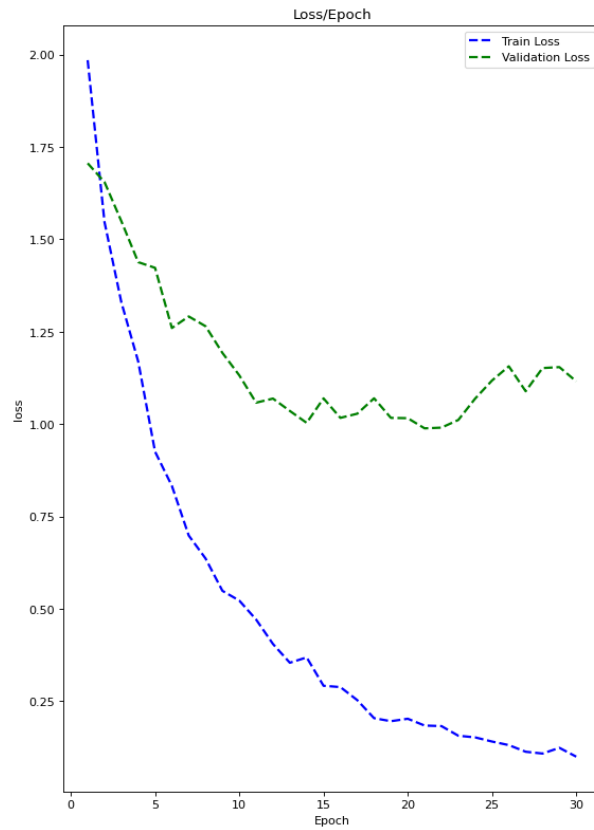- German-Middle High German :
  - Accuracy : ~ 46%

Single Language Morphological Generation

For this task, I cloned from this repo and run the given python file taking four different languages: Hindi, Bengali, German, and Middle High German
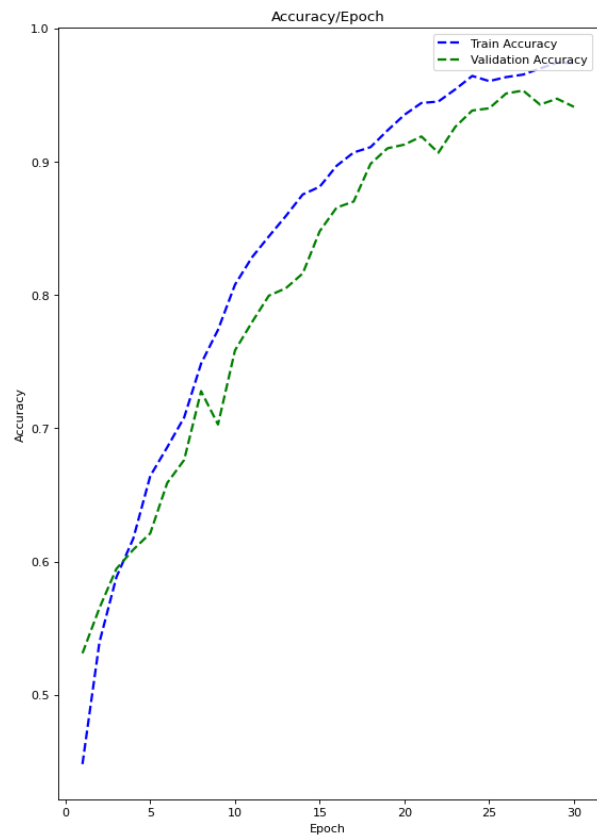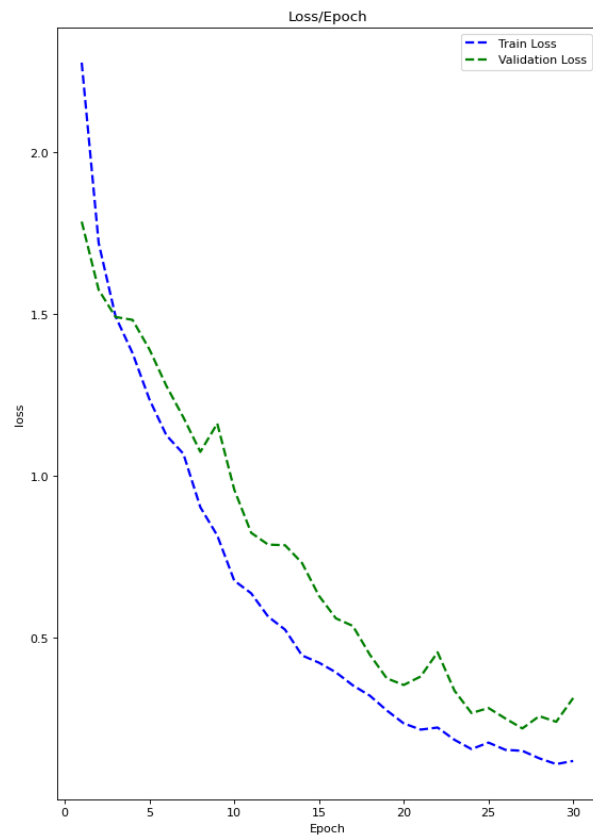
- Command Used : `!python morph_inflection_generation.py`
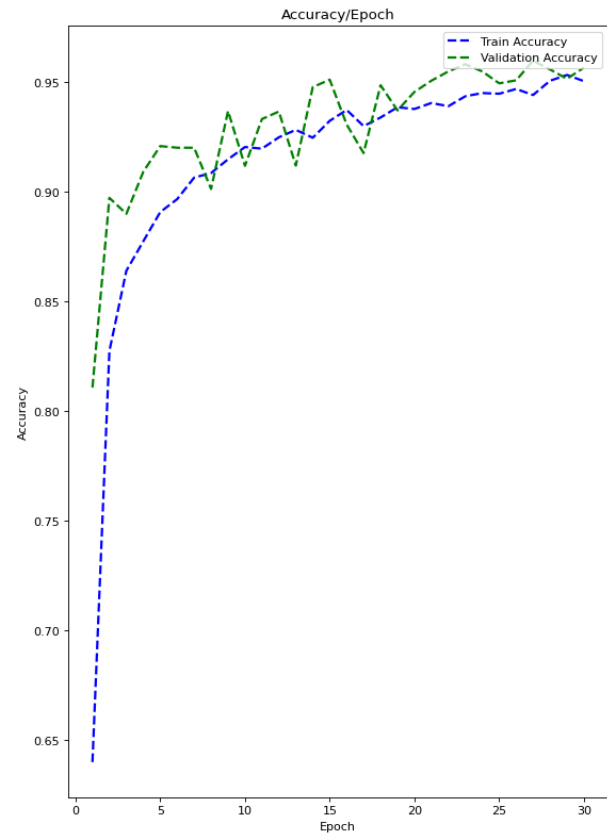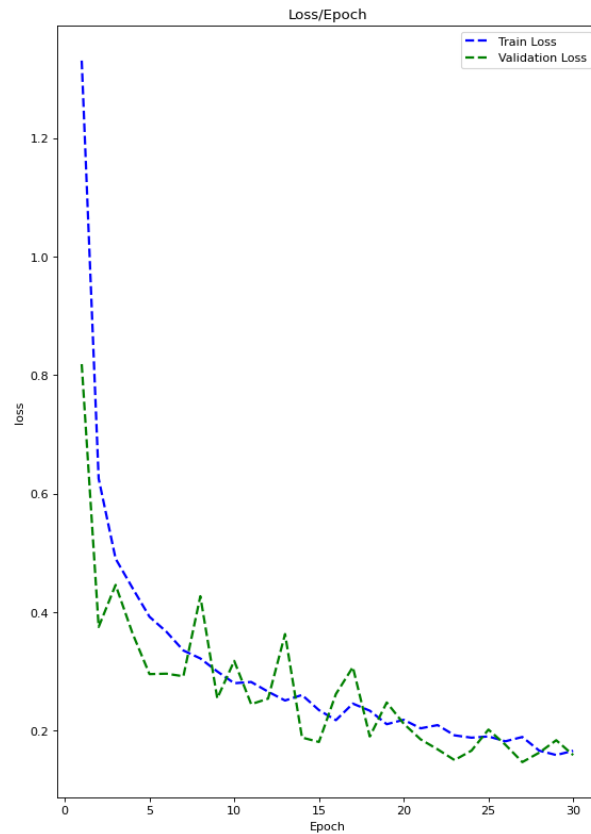
Results:

- Hindi :
  - Avg_TrainLoss: 0.099868
  - Avg_TrainAccuracy: 0.977354
  - Avg_ValLoss: 1.115763
  - Avg_ValAccuracy: 0.823852

- Bengali :
  - Avg_TrainLoss: 0.118441
  - Avg_TrainAccuracy: 0.973970
  - Avg_ValLoss: 0.312782
  - Avg_ValAccuracy: 0.941111

- German :
  - Avg_TrainLoss: 0.166123
  - Avg_TrainAccuracy: 0.950300
  - Avg_ValLoss: 0.158930
  - Avg_ValAccuracy: 0.956516

- Middle High German
  - Avg_TrainLoss: 0.281561
  - Avg_TrainAccuracy: 0.904929
  - Avg_ValLoss: 0.383734
  - Avg_ValAccuracy: 0.878462

Loss/Epoch

Accuracy/Epoch