

Jonathan Mandl 211399175

Danielle Hodaya Shrem 208150433

Part 4 – Adding Sub-word Units

Architecture

In this section, our model is a simple feed-forward neural network with a single hidden layer of 250 neurons and tanh activation.

We used the following hyperparameter configuration for the NER task:

- Learning rate: 0.001
- Epochs: 10
- Batch size: 64

For the POS task, we used the following hyperparameter configuration:

- Learning rate: 0.001
- Epochs: 5
- Batch size: 64

Sub-word units choices

We lower-cased all the words in our vocabulary before extracting prefixes and suffixes for each word to ensure we can use the lower-case vocabulary of the pre-trained embeddings. All three embedding matrices—words, prefixes, and suffixes—were initialized randomly. If pretrained vectors were provided, they replace the corresponding rows in the word, prefix, or suffix matrix.

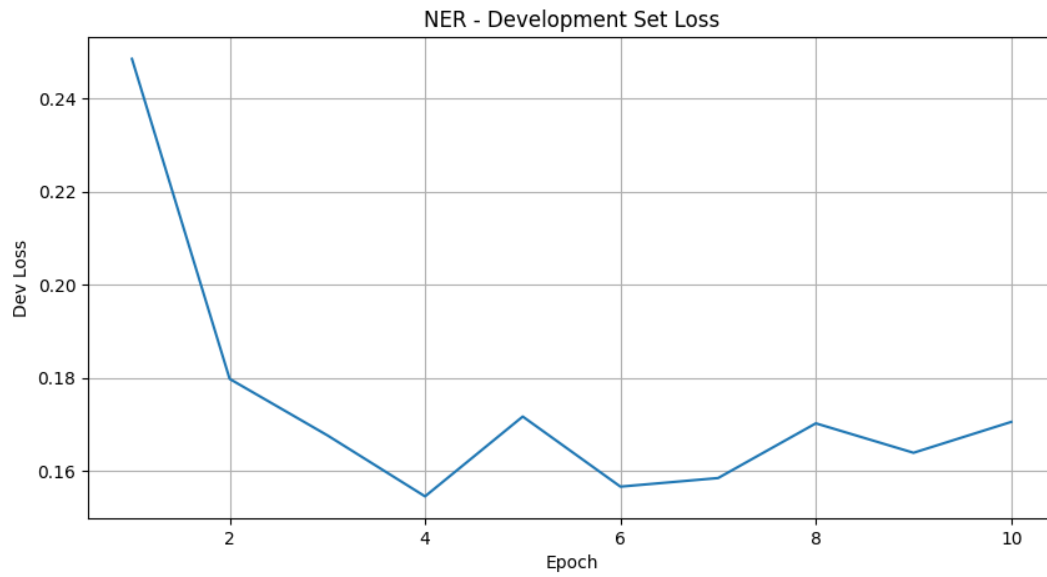
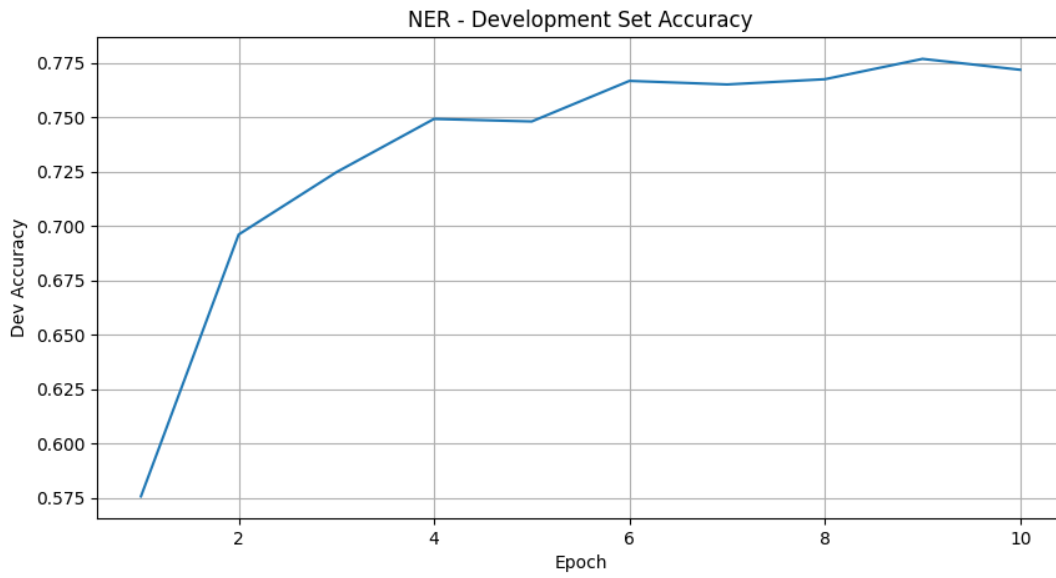
To handle prefixes, suffixes and words that appear in the development set but not in training, we added a special <UNK> token to our embedding matrices of prefixes, suffixes and words. During training, we randomly masked 15% of prefix/suffix/word tokens—replacing them with <UNK> token—so that the model learns a useful representation for unknown tokens.

Models Evaluation

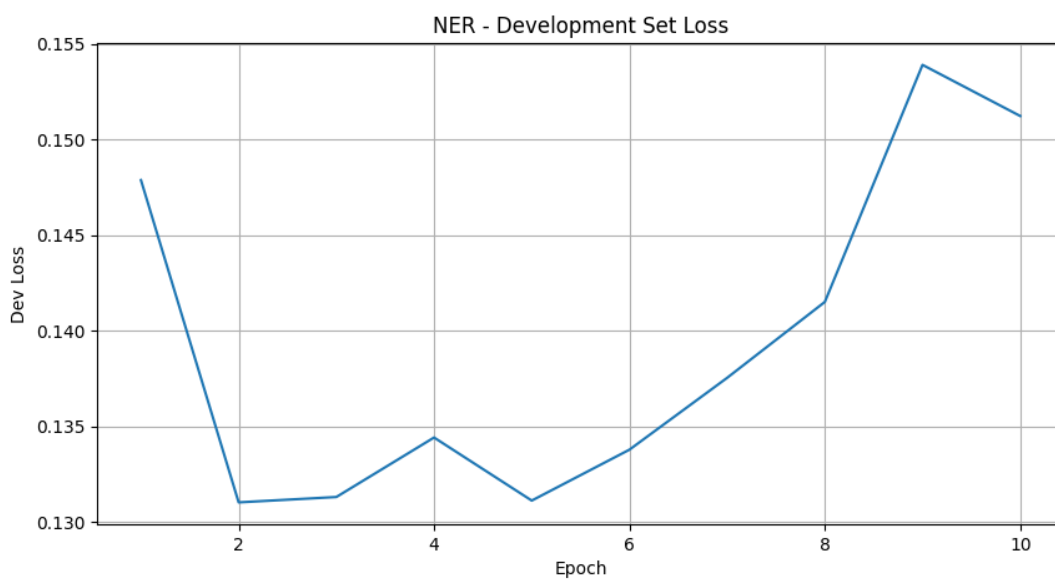
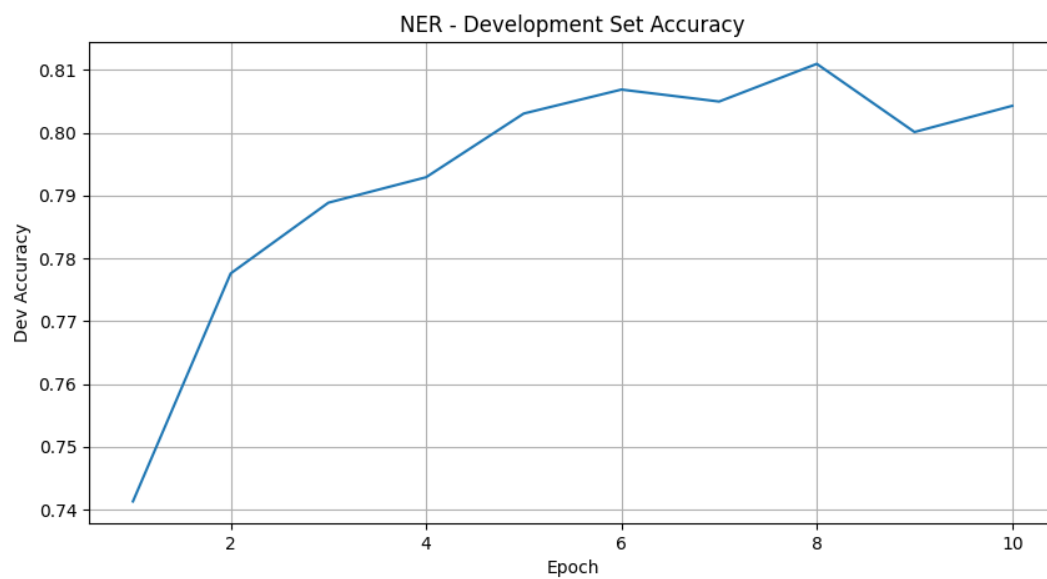
Our evaluation showed that neither subword units nor pre-trained embeddings alone improved on the baseline accuracies of 0.77 for NER and 0.95 for POS. However, their effects were complementary. When combined, subword-units and pre-trained embeddings yielded better performance. This trend was consistent across both tasks: NER accuracy rose to 0.80, and POS accuracy increased to nearly 0.96.

Figures

NER - no pre-trained embeddings

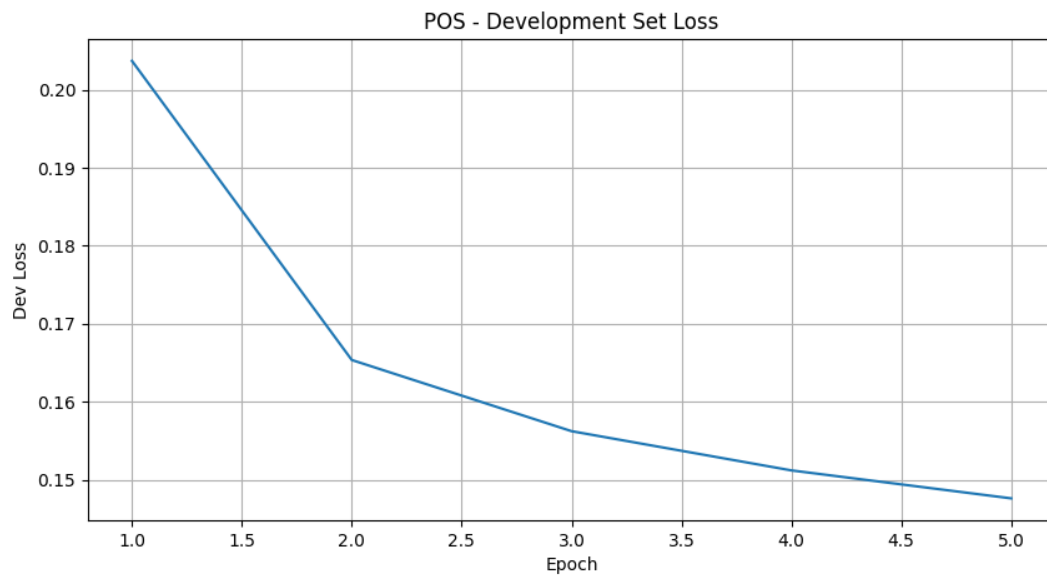
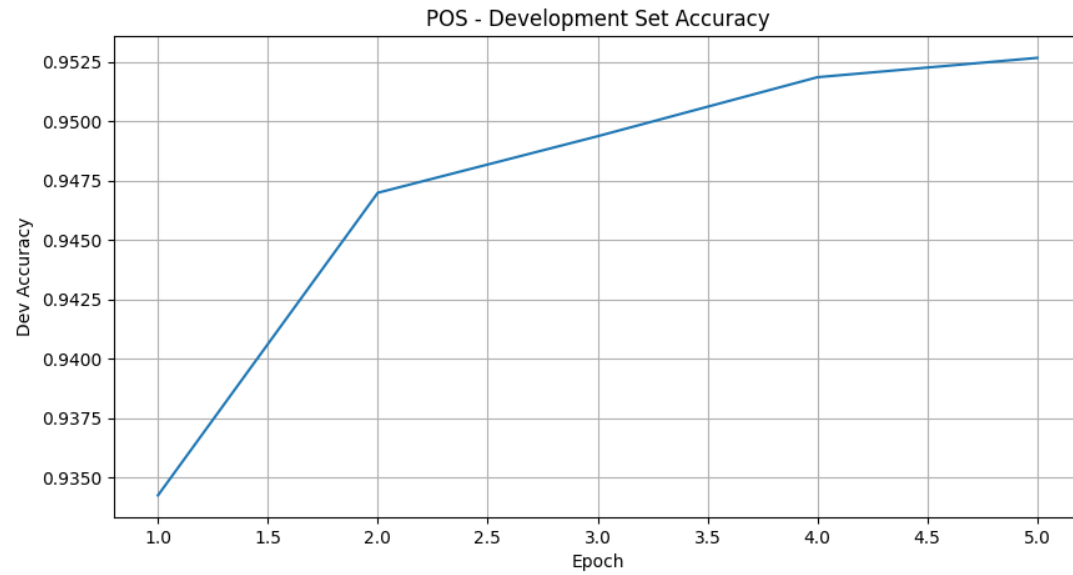


NER - with pre-trained embeddings



Task – POS

POS -no pre-trained embeddings



POS - with pre-trained embeddings

