Jonathan Mandl 211399175

Danielle Hodaya Shrem 208150433

**Part 1 – A Simple Window-Based Tagger**

**Architecture**

For this part, we used the same hyperparameter configuration to both our POS and NER tasks, as it delivered strong performance on each.

Our neural network consists of a single hidden layer with 250 neurons and tanh activation.

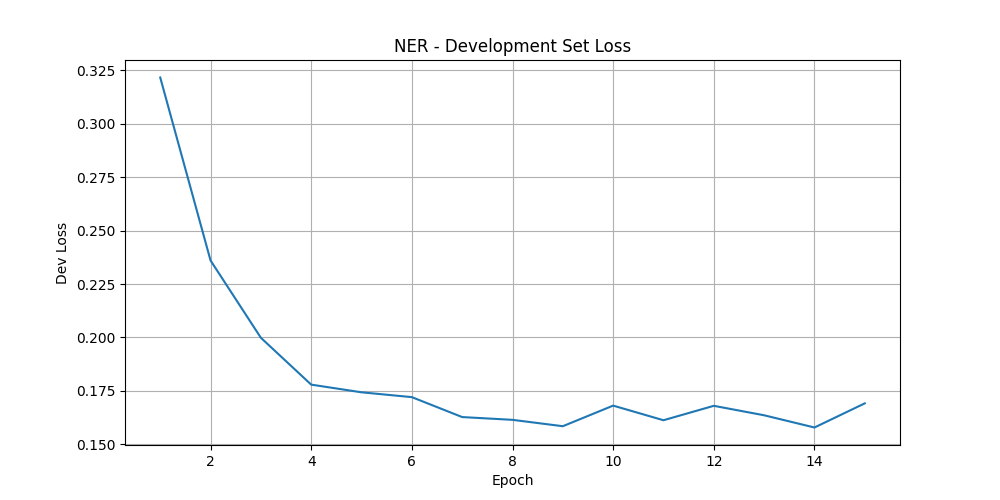
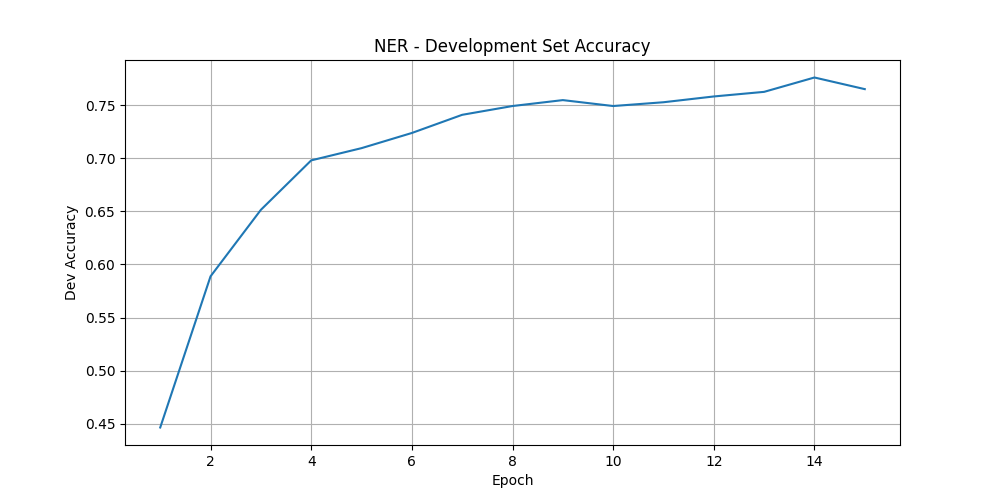
Our chosen hyperparameters were:

* **Learning rate:** 0.001
* **Epochs:** 15
* **Batch size:** 64

**Model choices**

To handle words that appear in the development set but not in training, we added a special <UNK> token to our embedding matrix. During training, we randomly masked 15% of tokens—replacing them with the <UNK> token - so that the model learns a useful representation for unknown words. To handle context windows that include sentence boundaries, we padded each input with <PAD> tokens on both sides (equal to the context size) and likewise included <PAD> in our embedding matrix.

**Task -NER**



**Task - POS**

