

Checkpoint 2.1:

First 5: [4.9, 4.1, 5.7, 4.3, 4.3]

Last 5: [4.2, 4.4, 4.2, 4.2, 3.7]

Checkpoint 2.2:

Accuracy plot saved to results/training_plot.png

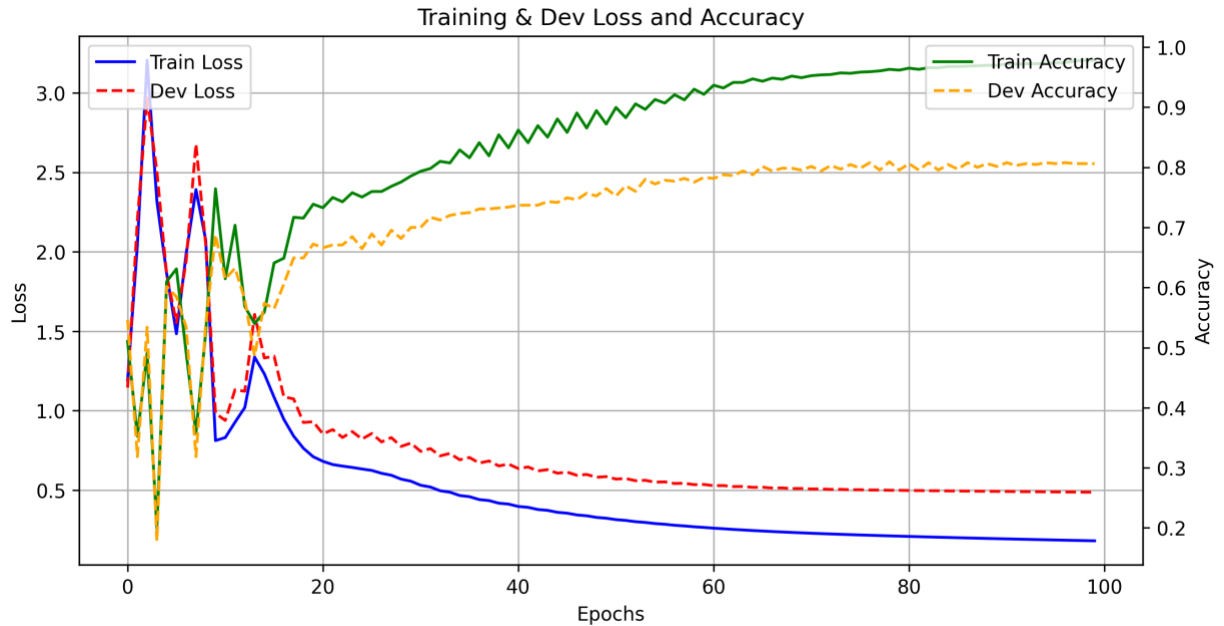


Checkpoint 2.3:

	1e-05	0.001	0.1
0.1	0.591	0.591	0.589
1	0.72	0.723	0.592
10	0.806	0.731	0.318

Best hyperparameters: lr=10, l2_penalty=1e-05

Accuracy plot saved to results/best_training_plot.png



Checkpoint 2.4:

```
[('The', 'O'), ('horse', 'O'), ('raced', 'N'), ('past', 'O'), ('the', 'O'), ('barn', 'N'), ('fell', 'O'), ('.', 'O')]
[('For', 'V'), ('3', 'O'), ('years', 'N'), ('.', 'O'), ('we', 'N'), ('attended', 'V'), ('S.B.U.', 'N'), ('in', 'O'), ('the', 'O'), ('CS', 'N'), ('program', 'N'), ('.', 'O')]
[('Did', 'O'), ('you', 'N'), ('hear', 'O'), ('Sam', 'N'), ('tell', 'V'), ('me', 'N'), ('to', 'O'), ('"', 'O'), ('chill', 'N'), ('out', 'O'), ('"', 'O'), ('yesterday', 'N'), ('?', 'O'), ('#rude', 'O')]
```

Qualitative Observations:

1. Performance on the test data is below average, likely because 50% of the tokens are out-of-bag vocabulary.
2. The model has learned that the token "the" (not "The") is generally followed by a noun, as it correctly predicted the tags for "barn" and "CS" but not for "horse".
3. It has also learned that tokens with capital letters are generally nouns, as it correctly predicted the tags for "S.B.U.", "CS", and "Sam", despite them being out-of-bag vocabulary.