🚀 GAUDI INITIALIZATION
==
✅ Habana frameworks loaded
✅ Using device: hpu
✅ HPU device count: 8
==

📊 LOADING DATA

Enter path to the directory containing CoNLL-U files (e.g., /path/to/UD_English-EWT):
training_data

✅ Found training file: training_data/bg_btb-ud-train.conllu
✅ Found validation (dev) file: training_data/bg_btb-ud-dev.conllu
✅ Found test file: training_data/bg_btb-ud-test.conllu
✅ Loaded 8,907 training sentences.
✅ Loaded 1,115 validation sentences.
✅ Loaded 1,116 test sentences.
Train: 8,907 | Val: 1,115 | Test: 1,116
Vocabulary: 9,684 words | 17 tags

🏗️ BUILDING MODEL

Train DataLoader batches: 5
Validation DataLoader batches: 1
Test DataLoader batches: 1
✅ Model on hpu
✅ Parameters: 3,615,761

🎯 TRAINING ON GAUDI
==

Epoch 1/25

Training: 100%
5/5 [01:13<00:00, 15.57s/it, loss=2.2685, acc=0.2048]
Evaluating: 100%
1/1 [00:08<00:00, 8.94s/it, loss=1.9663, acc=0.3796]
Train Loss: 2.4915 | Train Acc: 0.2048
Val Loss: 1.9663 | Val Acc: 0.3796
Time: 82.1s
✨ New best validation accuracy: 0.3796. Model saved to best_pos_tagger_model_FOLDER.pt

Epoch 2/25

Training: 100%
5/5 [00:03<00:00, 1.57it/s, loss=1.2899, acc=0.4785]
Evaluating: 100%
1/1 [00:00<00:00, 3.59it/s, loss=1.1445, acc=0.6266]
Train Loss: 1.5900 | Train Acc: 0.4785
Val Loss: 1.1445 | Val Acc: 0.6266
Time: 3.6s
✨ New best validation accuracy: 0.6266. Model saved to best_pos_tagger_model_FOLDER.pt

Epoch 3/25

Training: 100%
5/5 [00:03<00:00, 1.61it/s, loss=1.0631, acc=0.6319]
Evaluating: 100%
1/1 [00:00<00:00, 3.62it/s, loss=0.8794, acc=0.7124]
Train Loss: 1.1354 | Train Acc: 0.6319
Val Loss: 0.8794 | Val Acc: 0.7124
Time: 3.5s
✨ New best validation accuracy: 0.7124. Model saved to best_pos_tagger_model_FOLDER.pt

Epoch 4/25

Training: 100%
5/5 [00:03<00:00, 1.56it/s, loss=0.8708, acc=0.6962]
Evaluating: 100%
1/1 [00:00<00:00, 3.41it/s, loss=0.7238, acc=0.7619]
Train Loss: 0.9286 | Train Acc: 0.6962
Val Loss: 0.7238 | Val Acc: 0.7619
Time: 3.6s
✨ New best validation accuracy: 0.7619. Model saved to best_pos_tagger_model_FOLDER.pt

Epoch 5/25

Training: 100%
5/5 [00:03<00:00, 1.52it/s, loss=0.7202, acc=0.7438]
Evaluating: 100%
1/1 [00:00<00:00, 3.68it/s, loss=0.5947, acc=0.8051]

Train Loss: 0.7701 | Train Acc: 0.7438
Val Loss: 0.5947 | Val Acc: 0.8051
Time: 3.8s
✨ New best validation accuracy: 0.8051. Model saved to best_pos_tagger_model_FOLDER.pt

Epoch 6/25

Training: 100%
5/5 [00:03<00:00, 1.73it/s, loss=0.5807, acc=0.7891]
Evaluating: 100%
1/1 [00:00<00:00, 3.85it/s, loss=0.4951, acc=0.8413]
Train Loss: 0.6302 | Train Acc: 0.7891
Val Loss: 0.4951 | Val Acc: 0.8413
Time: 3.4s
✨ New best validation accuracy: 0.8413. Model saved to best_pos_tagger_model_FOLDER.pt

Epoch 7/25

Training: 100%
5/5 [00:02<00:00, 1.76it/s, loss=0.4798, acc=0.8245]
Evaluating: 100%
1/1 [00:00<00:00, 4.02it/s, loss=0.4156, acc=0.8664]
Train Loss: 0.5228 | Train Acc: 0.8245
Val Loss: 0.4156 | Val Acc: 0.8664
Time: 3.2s
✨ New best validation accuracy: 0.8664. Model saved to best_pos_tagger_model_FOLDER.pt

Epoch 8/25

Training: 100%
5/5 [00:02<00:00, 1.78it/s, loss=0.4104, acc=0.8552]
Evaluating: 100%
1/1 [00:00<00:00, 4.02it/s, loss=0.3547, acc=0.8862]
Train Loss: 0.4352 | Train Acc: 0.8552
Val Loss: 0.3547 | Val Acc: 0.8862
Time: 3.2s
✨ New best validation accuracy: 0.8862. Model saved to best_pos_tagger_model_FOLDER.pt

Epoch 9/25

Training: 100%
5/5 [00:02<00:00, 1.71it/s, loss=0.3402, acc=0.8793]
Evaluating: 100%
1/1 [00:00<00:00, 3.91it/s, loss=0.3067, acc=0.9004]
Train Loss: 0.3628 | Train Acc: 0.8793
Val Loss: 0.3067 | Val Acc: 0.9004
Time: 3.3s
✨ New best validation accuracy: 0.9004. Model saved to best_pos_tagger_model_FOLDER.pt

Epoch 10/25

Training: 100%
5/5 [00:03<00:00, 1.68it/s, loss=0.2824, acc=0.8992]
Evaluating: 100%
1/1 [00:00<00:00, 4.04it/s, loss=0.2839, acc=0.9107]
Train Loss: 0.3033 | Train Acc: 0.8992
Val Loss: 0.2839 | Val Acc: 0.9107
Time: 3.3s
✨ New best validation accuracy: 0.9107. Model saved to best_pos_tagger_model_FOLDER.pt

Epoch 11/25

Training: 100%
5/5 [00:02<00:00, 1.76it/s, loss=0.2816, acc=0.9087]
Evaluating: 100%
1/1 [00:00<00:00, 4.06it/s, loss=0.2613, acc=0.9193]
Train Loss: 0.2796 | Train Acc: 0.9087
Val Loss: 0.2613 | Val Acc: 0.9193
Time: 3.2s
✨ New best validation accuracy: 0.9193. Model saved to best_pos_tagger_model_FOLDER.pt

Epoch 12/25

Training: 100%
5/5 [00:02<00:00, 1.81it/s, loss=0.2325, acc=0.9204]
Evaluating: 100%
1/1 [00:00<00:00, 4.01it/s, loss=0.2391, acc=0.9254]
Train Loss: 0.2430 | Train Acc: 0.9204
Val Loss: 0.2391 | Val Acc: 0.9254
Time: 3.1s
✨ New best validation accuracy: 0.9254. Model saved to best_pos_tagger_model_FOLDER.pt

Epoch 13/25

Training: 100%
5/5 [00:02<00:00, 1.71it/s, loss=0.2079, acc=0.9304]
Evaluating: 100%
1/1 [00:00<00:00, 3.98it/s, loss=0.2231, acc=0.9300]
Train Loss: 0.2116 | Train Acc: 0.9304
Val Loss: 0.2231 | Val Acc: 0.9300
Time: 3.3s
✨ New best validation accuracy: 0.9300. Model saved to best_pos_tagger_model_FOLDER.pt

Epoch 14/25

Training: 100%
5/5 [00:03<00:00, 1.74it/s, loss=0.1811, acc=0.9404]
Evaluating: 100%
1/1 [00:00<00:00, 4.10it/s, loss=0.2088, acc=0.9350]
Train Loss: 0.1826 | Train Acc: 0.9404
Val Loss: 0.2088 | Val Acc: 0.9350
Time: 3.3s
✨ New best validation accuracy: 0.9350. Model saved to best_pos_tagger_model_FOLDER.pt

Epoch 15/25

Training: 100%
5/5 [00:02<00:00, 1.76it/s, loss=0.1604, acc=0.9466]
Evaluating: 100%
1/1 [00:00<00:00, 4.10it/s, loss=0.2004, acc=0.9377]
Train Loss: 0.1648 | Train Acc: 0.9466
Val Loss: 0.2004 | Val Acc: 0.9377
Time: 3.1s
✨ New best validation accuracy: 0.9377. Model saved to best_pos_tagger_model_FOLDER.pt

Epoch 16/25

Training: 100%
5/5 [00:02<00:00, 1.79it/s, loss=0.1369, acc=0.9517]
Evaluating: 100%
1/1 [00:00<00:00, 4.12it/s, loss=0.1907, acc=0.9403]

Train Loss: 0.1465 | Train Acc: 0.9517
Val Loss: 0.1907 | Val Acc: 0.9403
Time: 3.1s
✨ New best validation accuracy: 0.9403. Model saved to best_pos_tagger_model_FOLDER.pt

Epoch 17/25

Training: 100%
5/5 [00:03<00:00, 1.63it/s, loss=0.1407, acc=0.9545]
Evaluating: 100%
1/1 [00:00<00:00, 3.67it/s, loss=0.1841, acc=0.9421]
Train Loss: 0.1382 | Train Acc: 0.9545
Val Loss: 0.1841 | Val Acc: 0.9421
Time: 3.4s
✨ New best validation accuracy: 0.9421. Model saved to best_pos_tagger_model_FOLDER.pt

Epoch 18/25

Training: 100%
5/5 [00:03<00:00, 1.61it/s, loss=0.1331, acc=0.9582]
Evaluating: 100%
1/1 [00:00<00:00, 3.37it/s, loss=0.1812, acc=0.9439]
Train Loss: 0.1276 | Train Acc: 0.9582
Val Loss: 0.1812 | Val Acc: 0.9439
Time: 3.5s
✨ New best validation accuracy: 0.9439. Model saved to best_pos_tagger_model_FOLDER.pt

Epoch 19/25

Training: 100%
5/5 [00:03<00:00, 1.60it/s, loss=0.1125, acc=0.9606]
Evaluating: 100%
1/1 [00:00<00:00, 3.73it/s, loss=0.1761, acc=0.9457]
Train Loss: 0.1182 | Train Acc: 0.9606
Val Loss: 0.1761 | Val Acc: 0.9457
Time: 3.5s
✨ New best validation accuracy: 0.9457. Model saved to best_pos_tagger_model_FOLDER.pt

Epoch 20/25

Training: 100%
5/5 [00:03<00:00, 1.58it/s, loss=0.1012, acc=0.9638]
Evaluating: 100%
1/1 [00:00<00:00, 3.32it/s, loss=0.1759, acc=0.9470]
Train Loss: 0.1094 | Train Acc: 0.9638
Val Loss: 0.1759 | Val Acc: 0.9470
Time: 3.6s
✨ New best validation accuracy: 0.9470. Model saved to best_pos_tagger_model_FOLDER.pt

Epoch 21/25

Training: 100%
5/5 [00:03<00:00, 1.54it/s, loss=0.1100, acc=0.9651]
Evaluating: 100%
1/1 [00:00<00:00, 3.57it/s, loss=0.1751, acc=0.9472]
Train Loss: 0.1070 | Train Acc: 0.9651
Val Loss: 0.1751 | Val Acc: 0.9472
Time: 3.7s
✨ New best validation accuracy: 0.9472. Model saved to best_pos_tagger_model_FOLDER.pt

Epoch 22/25

Training: 100%
5/5 [00:03<00:00, 1.55it/s, loss=0.1023, acc=0.9663]
Evaluating: 100%
1/1 [00:00<00:00, 3.64it/s, loss=0.1729, acc=0.9469]
Train Loss: 0.1025 | Train Acc: 0.9663
Val Loss: 0.1729 | Val Acc: 0.9469
Time: 3.6s
Validation accuracy did not improve. Epochs without improvement: 1

Epoch 23/25

Training: 100%
5/5 [00:03<00:00, 1.60it/s, loss=0.0982, acc=0.9674]
Evaluating: 100%
1/1 [00:00<00:00, 3.52it/s, loss=0.1699, acc=0.9474]
Train Loss: 0.0983 | Train Acc: 0.9674
Val Loss: 0.1699 | Val Acc: 0.9474
Time: 3.5s
✨ New best validation accuracy: 0.9474. Model saved to best_pos_tagger_model_FOLDER.pt

Epoch 24/25

Training: 100%
 5/5 [00:03<00:00, 1.59it/s, loss=0.0981, acc=0.9689]
Evaluating: 100%
 1/1 [00:00<00:00, 4.76it/s, loss=0.1714, acc=0.9488]
Train Loss: 0.0953 | Train Acc: 0.9689
Val Loss: 0.1714 | Val Acc: 0.9488
Time: 3.4s
✨ New best validation accuracy: 0.9488. Model saved to best_pos_tagger_model_FOLDER.pt

Epoch 25/25

Training: 100%
 5/5 [00:02<00:00, 1.70it/s, loss=0.0878, acc=0.9699]
Evaluating: 100%
 1/1 [00:00<00:00, 3.83it/s, loss=0.1722, acc=0.9495]
Train Loss: 0.0904 | Train Acc: 0.9699
Val Loss: 0.1722 | Val Acc: 0.9495
Time: 3.3s
✨ New best validation accuracy: 0.9495. Model saved to best_pos_tagger_model_FOLDER.pt

📊 FINAL TEST EVALUATION
==
✅ Loaded best model from best_pos_tagger_model_FOLDER.pt for final evaluation.

Evaluating: 100%
 1/1 [00:09<00:00, 9.42s/it, loss=0.1747, acc=0.9484]
Test Loss: 0.1747
Test Accuracy: 0.9484 (94.84%)

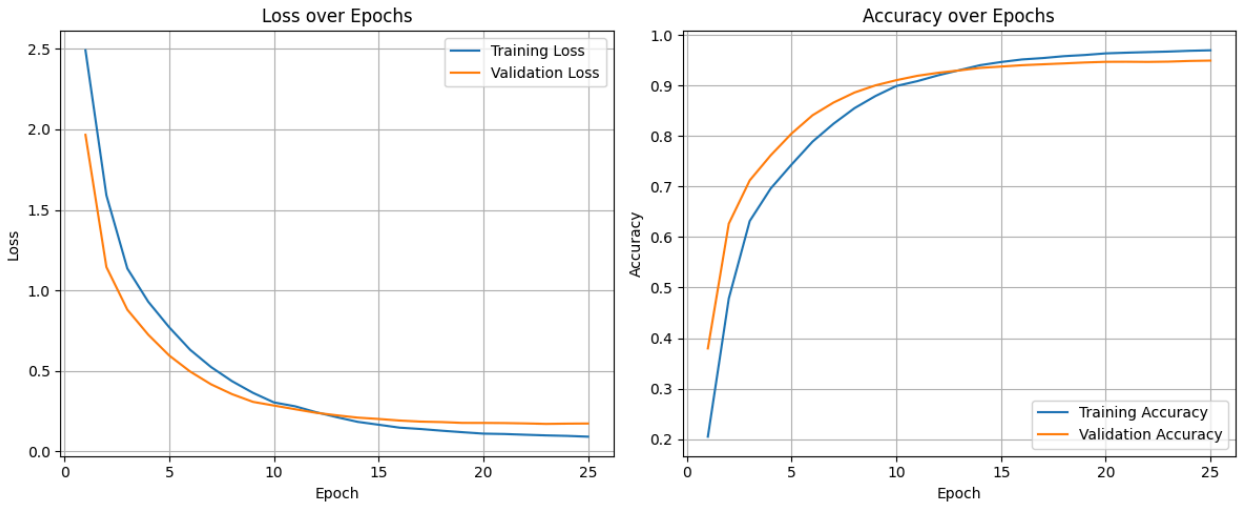
🎉 TRAINING COMPLETE ON GAUDI HPU!
Best Val Acc: 0.9495
Final Test Acc: 0.9484
Final Test Accuracy Classification: 🎉 Great! Very strong performance.

📚 Vocabulary Summary:

 Words: 9,684
 POS tags: 17

Sample words: ['В', 'дискусията', ',', 'предполагам', 'ще', 'се', 'важни', 'въпроси', '.', 'път', '...']
POS tags: ['ADP', 'NOUN', 'PUNCT', 'VERB', 'AUX', 'PRON', 'ADJ', 'PART', 'ADV', 'INTJ', 'DET', 'PROPN', 'CCONJ', 'NUM', 'SCONJ', 'X']

📈 Metrics plot saved to training_metrics.png



📊 POS Tagging Results (Random Sample from Test Set):

--- Sample Sentence 1 ---

Word	Predicted Tag	Actual Tag	Correct?
Неотдавна	ADV	ADV	✅
мъжът	NOUN	NOUN	✅
с	ADP	ADP	✅
83	NUM	NUM	✅
имена	NOUN	NOUN	✅
бе	AUX	AUX	✅
арестуван	VERB	VERB	✅
в	ADP	ADP	✅
дома	NOUN	NOUN	✅
си	PRON	PRON	✅
в	ADP	ADP	✅
<unk>	NOUN	PROPN	❌
по	ADP	ADP	✅
обвинения	NOUN	NOUN	✅
в	ADP	ADP	✅
<unk>	ADJ	ADJ	✅
<unk>	NOUN	NOUN	✅
.	PUNCT	PUNCT	✅

--- Sample Sentence 2 ---

Word	Predicted Tag	Actual Tag	Correct?
Сред	ADP	ADP	✅
20	NUM	NUM	✅
000	NUM	NUM	✅
<unk>	NOUN	ADJ	❌
на	ADP	ADP	✅
Балканите	PROPN	PROPN	✅
военнослужещи	NOUN	NOUN	✅
може	VERB	VERB	✅
да	AUX	AUX	✅
се	PRON	PRON	✅
очакват	VERB	VERB	✅
по	ADP	ADP	✅
2	NUM	NUM	✅
-	PUNCT	PUNCT	✅
3	NUM	NUM	✅
случая	NOUN	NOUN	✅
всяка	DET	DET	✅
година	NOUN	NOUN	✅
,	PUNCT	PUNCT	✅
каза	VERB	VERB	✅
<unk>	PROPN	PROPN	✅
.	PUNCT	PUNCT	✅

--- Sample Sentence 3 ---

Word	Predicted Tag	Actual Tag	Correct?
<unk>	NOUN	NOUN	✅
на	ADP	ADP	✅
<unk>	NOUN	NOUN	✅
,	PUNCT	PUNCT	✅
<unk>	NOUN	PRON	❌
<unk>	VERB	VERB	✅
<unk>	NOUN	NOUN	✅
му	PRON	PRON	✅
,	PUNCT	PUNCT	✅
<unk>	VERB	VERB	✅
.	PUNCT	PUNCT	✅

--- Sample Sentence 4 ---

Word	Predicted Tag	Actual Tag	Correct?

Слънцето | NOUN | NOUN | ✅
бе | AUX | AUX | ✅
<unk> | VERB | VERB | ✅
зад | ADP | ADP | ✅
сините | ADJ | ADJ | ✅
далечни | ADJ | ADJ | ✅
<unk> | NOUN | NOUN | ✅
на | ADP | ADP | ✅
<unk> | NOUN | NOUN | ✅
. | PUNCT | PUNCT | ✅

--- Sample Sentence 5 ---
Word | Predicted Tag | Actual Tag | Correct?

Да | PART | PART | ✅
, | PUNCT | PUNCT | ✅
<unk> | VERB | NOUN | ❌
<unk> | NOUN | VERB | ❌
. | PUNCT | PUNCT | ✅

🧠 Analysis of Sample POS Tagging Results:

Total words analyzed in samples: 66
Correct predictions in samples: 61
Incorrect predictions in samples: 5
Sample Accuracy: 92.42%

--- Performance on Unknown (<unk>) Words ---
Total <unk> words in samples: 16
Correct <unk> predictions: 11
Incorrect <unk> predictions: 5
Accuracy on <unk> words: 68.75%
Observation: The model shows some ability to guess tags for unknown words, which is good.

--- Common Tagging Errors (Predicted -> Actual) ---
- Predicted 'NOUN' but was actually 'PROPN': 1 times
- Predicted 'NOUN' but was actually 'ADJ': 1 times
- Predicted 'NOUN' but was actually 'PRON': 1 times
- Predicted 'VERB' but was actually 'NOUN': 1 times
- Predicted 'NOUN' but was actually 'VERB': 1 times

Observation: Specific tag confusions indicate areas for model refinement or more diverse training data for those tags.

--- Overall Sample Quality Assessment ---
Conclusion: The sample predictions show a high level of accuracy and consistency. The model generalizes well to unseen sentences.
