The `viterbi` function was tested with a set of manually tagged sentences, as well as a larger dataset from the Brown corpus and the simple dataset. The performance varied significantly across different tests, largely due to the handling of unseen words and the structure of the test sentences.

Hard coded sentences (from pd7):
"cat/N chase/V dog/N",
"cat/N watch/V chase/NP",
"chase/NP get/V watch/N",
"chase/NP watch/V dog/N and/CNJ cat/N",
"dog/N watch/V cat/N watch/V dog/N",
"cat/N watch/V watch/N and/CNJ chase/NP",
"dog/N watch/V and/CNJ chase/V chase/NP"

Test Sentence Results:

1. Test Sentence 1: "cat watch dog"
   - Expected: N V N
   - Result: N V N

This sentence was tagged correctly, which indicates that the training data included these words with the corresponding tags.

2. Test Sentence 2: "cat chase dog and watch dog"
   - Expected: N V N CNJ V N
   - Result: N V N CNJ V N

The tags were as expected. The training data seems to have included these word-to-tag and tag-to-tag relationships.

3. Test Sentence 3: "cat dog watch chase and"
   - Expected: N N V V CNJ
   - Result: N V N V CNJ

The tagging for "dog" was incorrect. This was likely because the model has not learned to recognize "dog" as a noun following another noun as well as “watch” as a verb following a verb. This was likely because the data that this instance of the viterbi function was ran on a significantly small dataset, with only seven sentences to be monitored. Thus, the model likely has insufficient training data, supported by the fact that a “cat” - “dog” transition was never in the data set.

Overall Testing Performance:

Simple Data Accuracy: 86.49%

Brown Data Accuracy: 96.47% - 35109/36394 (consistent with instructions in ps5)

These results suggest that the model performs well with the Brown dataset, which is larger and more varied, providing a better foundation for the algorithm to learn from. The lower accuracy on the simple dataset could be due to its limited size and diversity, which affects the model's ability to handle less common patterns and unseen words.

Unseen-word Penalty and Parameters

The unseen-word penalty (set to -100.0) is important for the performance of the `viterbi` function. If the penalty is too harsh, the model may incorrectly tag known words following an unknown word due to the propagated low score. If too lenient, it might lead to over-predicting unseen words with incorrect tags. I used the value of -100.0 because it was the value used in the PS5 instructions, and is also a consistent baseline for this value.

Moreover, the model's performance is sensitive to the training data's size and quality. A more comprehensive dataset leads to more accurate predictions, as seen with the Brown dataset. With smaller datasets, it is less consistent and accurate.