Nolan Shikanai

N-Grams Models Homework

4. Othello is not a good model for Don Quijote. Othello is old timey English and Don Quijote is in Spanish. Same time period, but entirely different languages. Testing Othello on Don Quijote gives me a perplexity of 2.870898496046415. Testing Othello against the interviews from the 1970's give me a perplexity of 1.0033968166224285. Testing against the Two Gentlemen of Verona gives me 1.003322116981076. I think my perplexity numbers are off, but the numbers are somewhat representative. Don Quijote is way higher than any other perplexity because it's in an entirely different language. Both the english ones are low, but the Two Gentlemen of Verona is lower, because that language is closer to Othello than interviews from the 1970's.

5. The perplexity is higher than before because it's more strict without smoothing. If there isn't a match in the trigrams it will think that the sentence is not possible.