Testing method

We concocted a set of 29 unique poem excerpts (Appendix A), 15 of which were machine-generated and 14 human-written. We worked with five sets of time periods: Pre-1550, 1550-1780, 1781-1900, 1901-1950, 1950-present. The accuracy of the results is hindered by the more ambiguous transitionary periods in between them. The time period mix is as follows: 6 excerpts from each time period and 5 excerpts from the pre-1550 interval. To ensure uniformity, all the samples used in testing were generated with the key length of 2 words.

To ensure that the results are as meticulous as possible we reached out to a Shakespearian instructor in the English department and asked her to identify and separate the excerpts from machine-generated poems from the excerpts from actual poems and sort them into their respective time periods. On evaluating the effectiveness of differentiating the Markov poems to the actual poems, we found that the instructor was 82.76% efficient. We found the overall time period detection efficiency was 62.1%. The time period detection was never more than one period.

The breakdown of the number of words learned and periodic efficiency is as follows:

Table 1- Periodic efficiency and number of words learned.

Time period	No. of words learned	Efficiency (%)
Pre-1550	93886	83.4%
1550-1780	331050	50%
1781-1900	460829	83.4%
1901-1950	253699	83.4%
1950- present	724382	16.67%

While this is counter-intuitive to the idea of larger data sets being more linguistically accurate, most errors occurred in the separation of the periods 1901-1950 and 1950-present. Prior to taking the test, the instructor warned us that the linguistic and poetic style of the 20th century is virtually indistinguishable without the provision of a context. The accuracy is 100% when both these time periods are taken as one interval. Our revised accuracies do not still exhibit a proportional increase. The reason for this is more linguistic than mathematical. The peculiarity of early modern English in our samples makes it very distinguishable. The periods thereafter reflect an increase in accuracy with an increase in sample size. Also note that while the first few periods are in intervals of 100-200 years, the last two periods are only in intervals of fifty years each.

As a result, we think that the differences in the accuracy of the professor's guesses about the time periods of the poetry stem not from differences in the sizes of the training data, but are likely representative instead of a decrease in the distinctiveness of poems between the first and second half of the 1900s.