1) a) Let $y = (y_1, y_2 \cdots y_n)$

$$P_{Y|F}(y|f) = P_{f^{-1}(y_1)} \cdot \prod_{i=2}^{n} M_{f^{-1}(y_i), f^{-1}(y_{i-1})}$$

b) $$P_{F|Y}(f|y) = \frac{P_{Y|F}(y|f) \cdot P_F(f)}{P_Y(y)} = \frac{P_{Y|F}(y|f) \cdot \frac{1}{|A|!}}{\sum_f P_{Y|F}(y|f) \cdot \frac{1}{|A|!}} = \frac{P_{Y|F}(y|f)}{\sum_{f'} P_{Y|F}(y|f')}$$

$$\hat{f}_{MAP}(y) = \arg\max_f P_{F|Y}(f|y) = \arg\max_f P_{Y|F}(y|f)$$

c) The posterior is non-convex in f, so standard optimization techniques will not suffice. In addition, the space of all possible f's is much too large to search exhaustively.

2) a) $$\frac{\binom{|A|}{2}}{|A|!}$$

b) Our objective posterior is $P_{F|Y}(f|y) = \frac{P_{Y|F}(y|f)}{\sum_{f'} P_{Y|F}(y|f')}$. With this as our target dist, the Markov chain constructed with MH algorithm can be done as follows:

Our partition function is $Z_p = \sum_{f'} P_{Y|F}(y|f')$, so let $P_0(f) = P_{Y|F}(y|f)$.

Now, define our proposal function as:

$$v(f|f') = \begin{cases} \frac{1}{\binom{|A|}{2}} & \text{if } f \text{ and } f' \text{ differ at exactly 2 letters} \\ 0 & \text{o.w.} \end{cases}$$

Since $v(f|f') = v(f'|f)$, our acceptance function is now:

$$a(f \to f') = \min\left(1, \frac{P_0(f')\, v(f|f')}{P_0(f)\, v(f'|f)}\right) = \min\left(1, \frac{P_0(f')}{P_0(f)}\right).$$

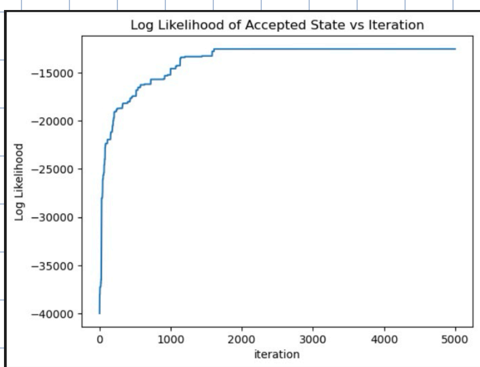
With these functions defined, to get the Markov Chain with the posterior as the stationary dist, follow algorithm 20.2 in the course notes.

2) c) One possible decoding algorithm is as follows:

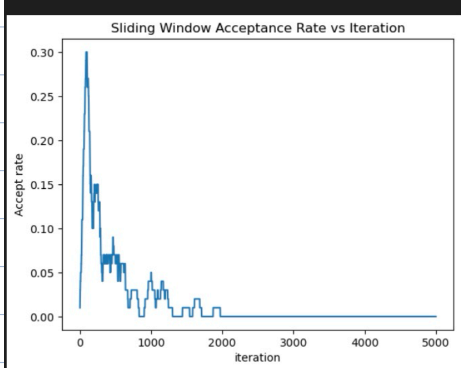
1. Pick an initial permutation f_i
2. Define $f^* = f_i$, $L^* = P_{Y|f}(y|f_i)$, and $f_p = f_i$
3. Generate f' according to $V(f'|f_p)$ as defined in the previous part
4. Sample a Bernoulli RV with parameter $a(f \to f')$ as defined in the previous part
5. If we get 1, we accept f' as the next state
   If $P_{Y|f}(y|f') > L^*$, update $L^* = P_{Y|f}(y|f')$ and $f^* = f'$. Set $f_p = f'$

   If we get 0, we reject f' and take f as the next state. keep $f_p = f$.
6. Go back to step 3, or terminate if a predetermined # of steps is reached. If terminate, take f^* as our predicted permutation.

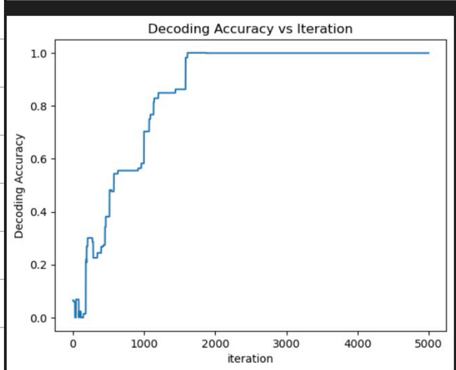
3) a)



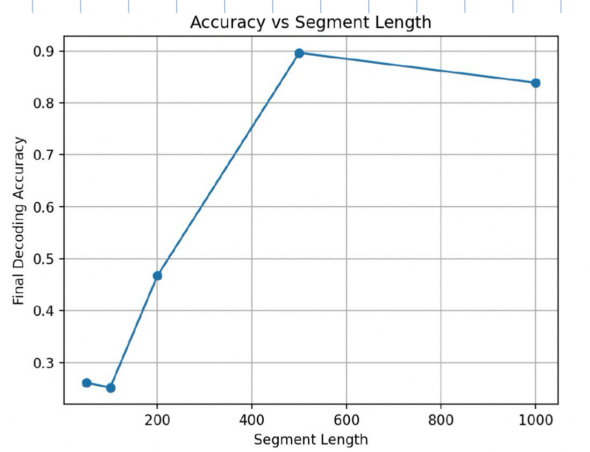
b)



c)



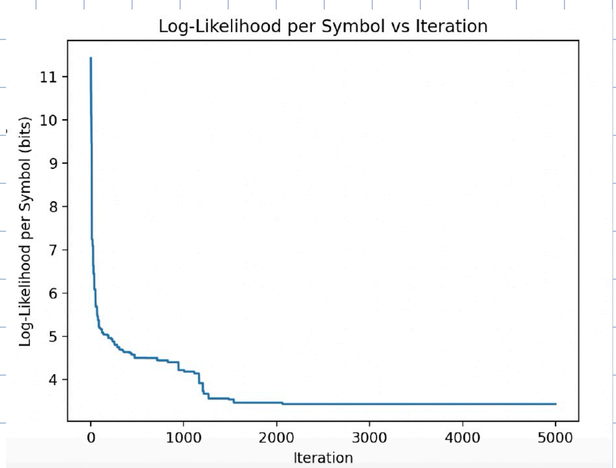
3) d)



Smaller segments with lower accuracy. They do not get enough of the overall context to see how the letters go together.

We are effectively disrupting the signal and breaking up dependencies.

e)



The bits/symbol decreases to around 3.4. The entropy for the English text we are dealing with is 4.1 bits/symbol. Notice this is higher than our value.

Actual English has smaller entropy at 2.3 because there are further relationships than letter to letter.