Natuurlijke Taalmodellen & Interfaces
Tejaswini Deoskar
Part A (Step 1)

Lucas van Berkel, 10747958 Ruben Gerritse, 10760326 Gideon Mooijen, 10686290

1. The 10 most frequent sequences.

	n = 1	n = 2	n = 3	
1	the	of the	I do not	
2	to	to be	I am sure	
3	and	in the	in the world	
4	of	I am	She could not	
5	a	of her	would have been	
6	her	to the	I dare say	
7	I	it was	as soon as	
8	was	had been	a great deal	
9	in	she had	it would be	
10	it	to her	could not be	

2. The sum of all frequencies of all sequences for that n.

	n = 1	n=2	n = 3
1	617091	617091	617091

3. To use the program, run the following command in your terminal.

python BerkelGerritseMooijen_1.py [PATH TO CORPUS] [VALUE N] [VALUE M]