



Jeevya AROUN
Nicolas NGAUV

CBOW

Model Implementation

Q°1) How do you think the cosine similarity scores will/should change after training ? In the plot, how would you expect the vectors to move ? Which word vectors should end up closer/farther ?

Initially, the embeddings for the words are randomly initialized, meaning there is no inherent semantic structure to how the words are represented. As a result, the cosine similarity scores between different word vectors are expected to be relatively random and not reflective of any meaningful relationships.

After training, we expect the embeddings to capture the semantic relationships between words based on the context they appear in. The cosine similarity scores should change to reflect these learned relationships.

First, we would see an increase of similarity for semantically related words, especially with words that often appear in similar contexts, indicating that their vectors are closer together in the embedding space. For example, 'king' and 'queen' might end up with a high cosine similarity score because they are often found together in contexts like royalty or leadership.

Then, we would notice a decrease of similarity for semantically unrelated words like words that appear in very different contexts, indicating that their vectors are farther apart. For instance, 'king' and 'woman' might have a lower cosine similarity score as they rarely share similar contexts.

For the vectors in the plot, we will see words like 'king' and 'queen', 'man' and 'woman', 'lord' and 'lady' should move closer together. These pairs of words are semantically related and likely to share contexts (likely the same grade of title, but different gender). Maybe we could also see 'lady' and 'queen' and 'woman', 'lord' and 'king' and 'man' moving closer together (same gender but different grade of title).

In one hand we have terms discriminated by social class, and in the other hand we have terms discriminated by gender.

We also should see words that are semantically distinct move farther apart. For example, if 'king' and 'woman' were initially close due to random initialization, they should move farther apart after training because they do not share similar contexts.

So, in terms of visual representation, after training, we could see clusters of semantically similar words. For example, 'king', 'queen', 'lord', and 'lady' might form a cluster representing royalty and nobility.

Different clusters should be well-separated, reflecting different semantic fields. For instance, a cluster of words related to royalty or nobility ('king', 'lady', ...) should be distinct from a cluster related to common class (like 'man' and 'woman').

Q°2) Using the same list of words as before, the next cells show the new similarities and plot the new word vectors. What has changed ? Do the results seem to verify your hypotheses ? How do you think the results could be improved ?

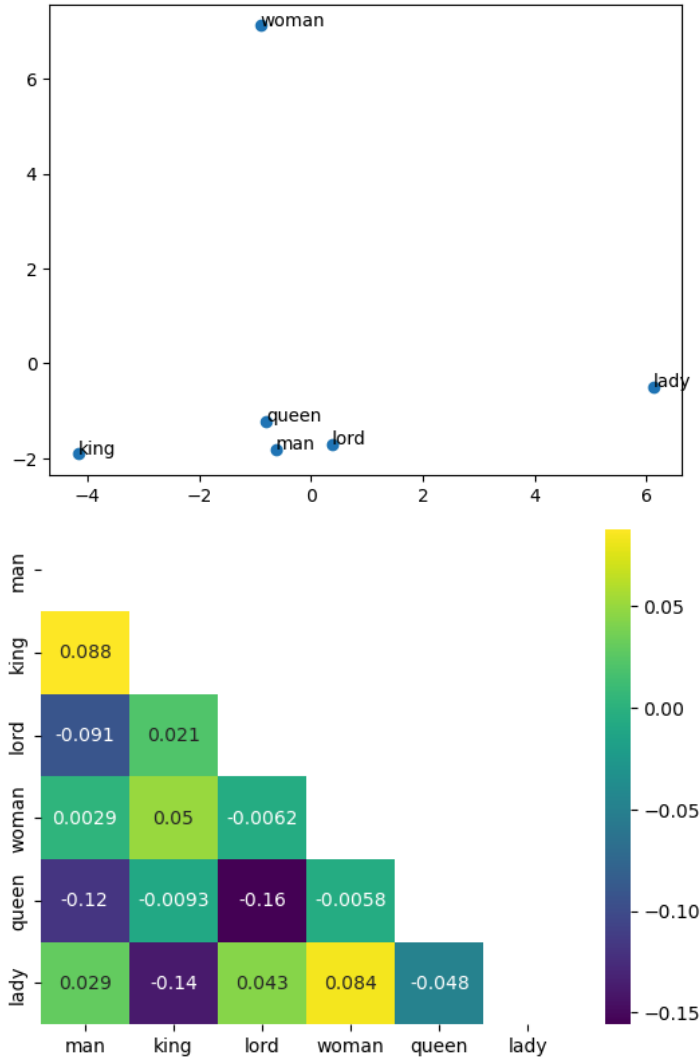
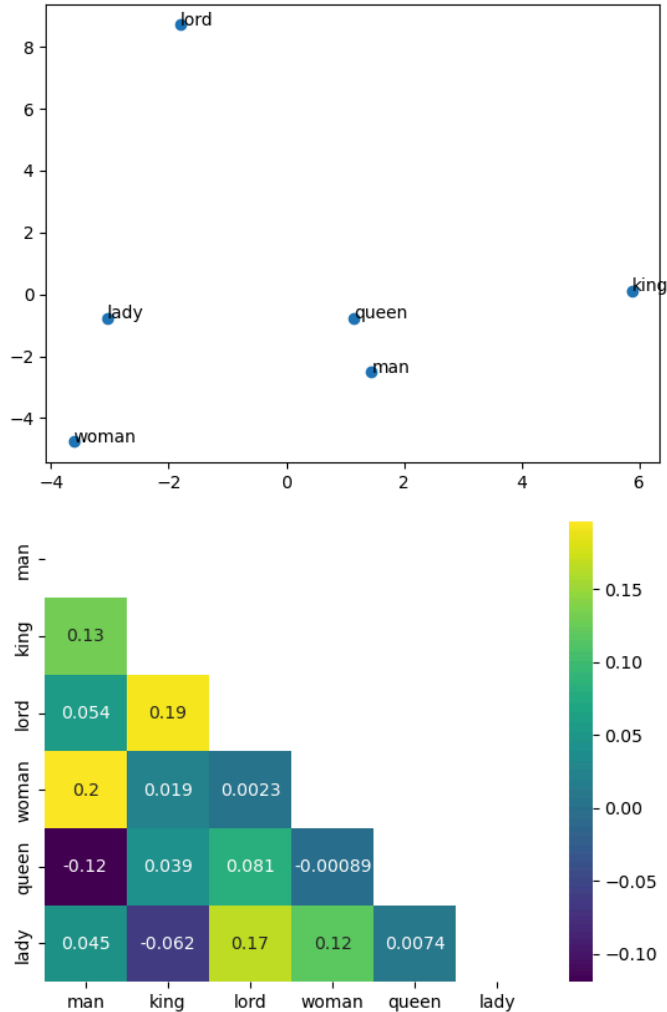
Some words moved closer together ("lady" and "woman" for example) and others moved farther apart ("queen" and "lord"), we can also notice that some cosine similarity scores increased ("king" and "man" for example) and others decreased ("woman" and "king").

Unfortunately, the results don't seem to verify our hypotheses... All the words moved, but we can't see any clusters clearly in our results...

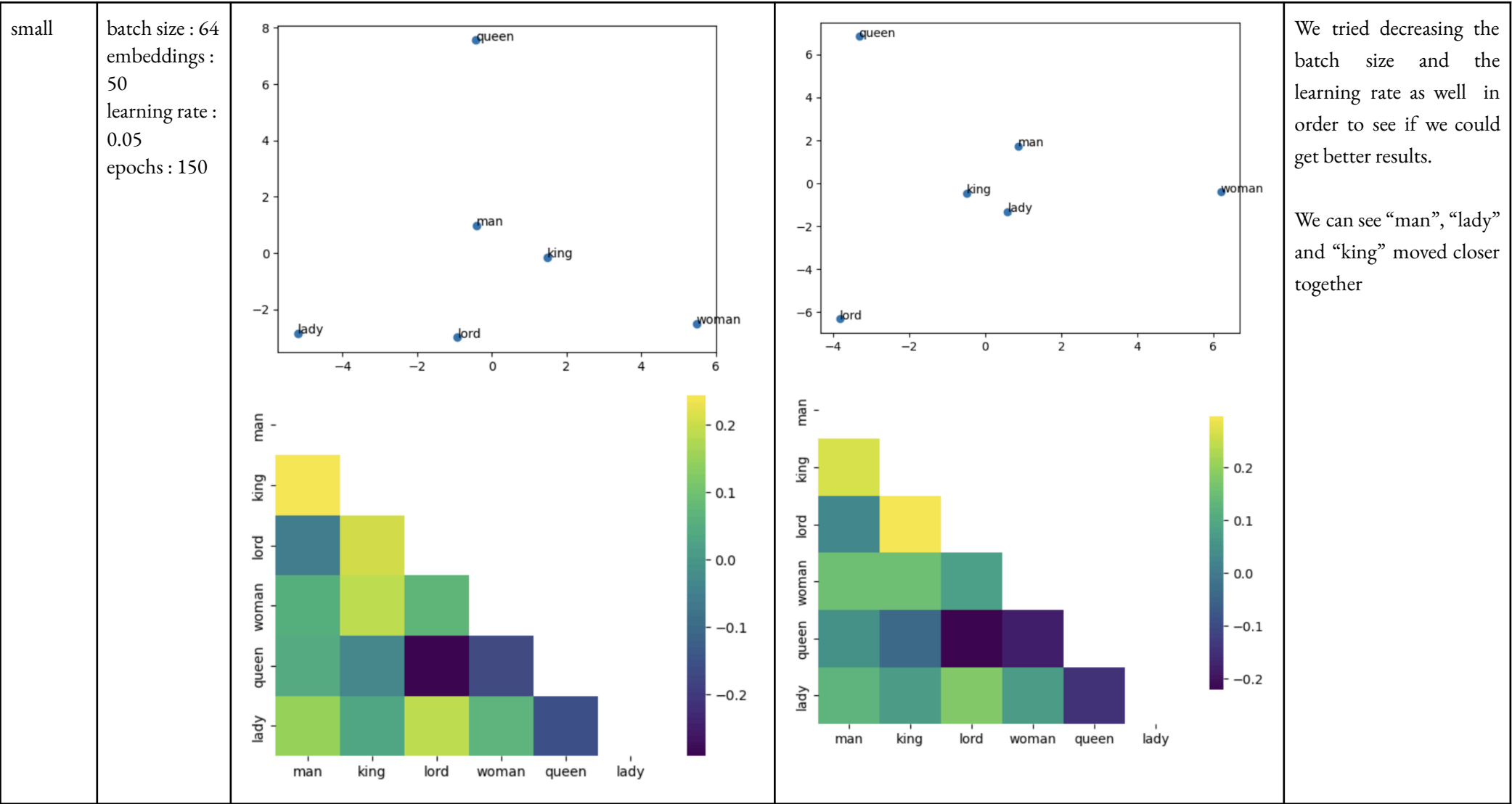
In order to improve our results, we tried modifying different hyperparameters. For example, we tried decreasing the learning rate so that the model could converge more slowly. We also tried changing the embedding dimension so that the model could maybe capture more nuances. The number of epochs was also increased to give the model more opportunities to learn better. In addition we varied the batch size from 64 to 512 to see how it would affect our results. These changes were made on the attempts on the small corpus because running the script on larger ones was very time consuming and so we couldn't try many variations.

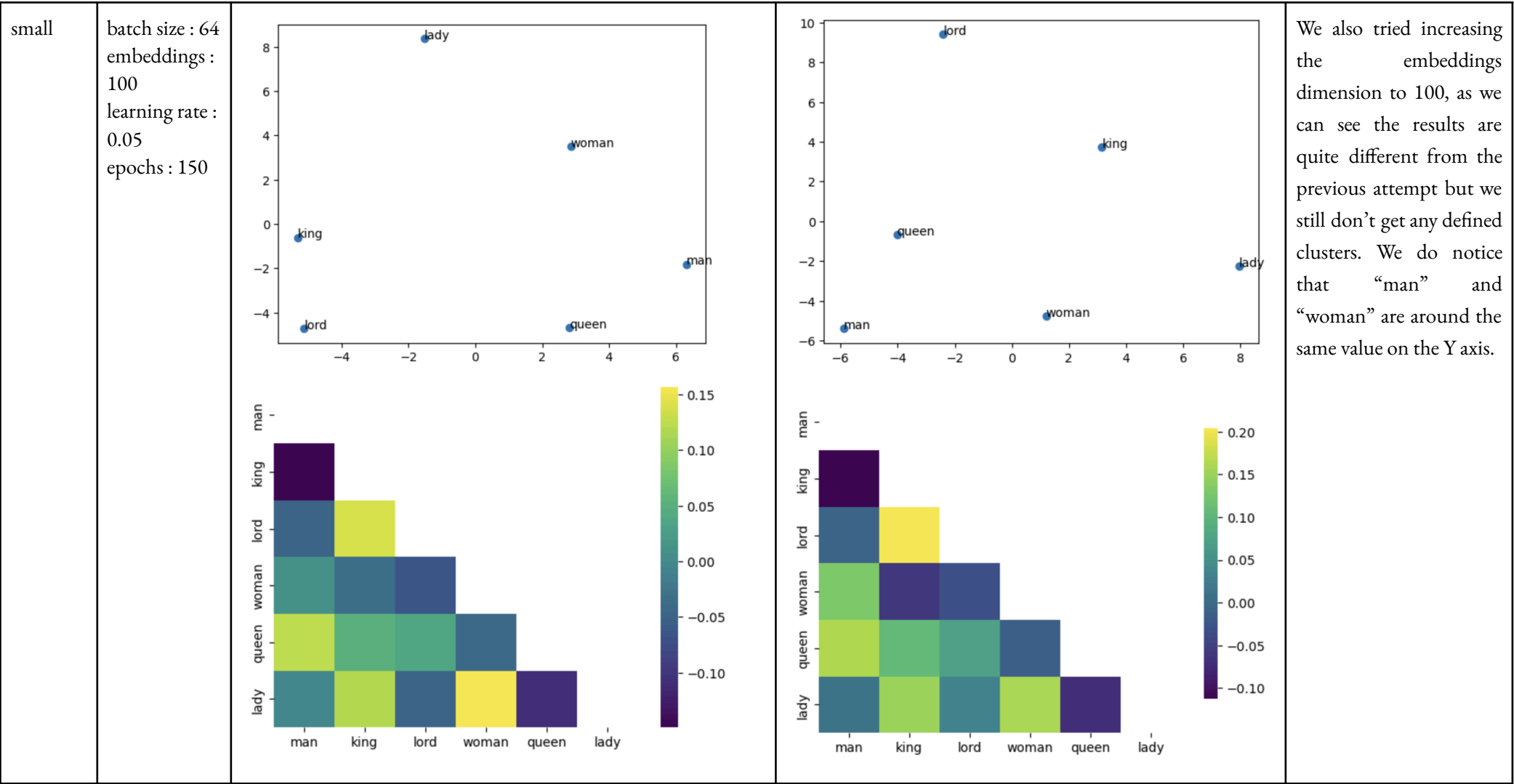
After experimenting with different hyperparameters, we observed that the overall loss decreases significantly as we increase the batch size, add more epochs, and raise the learning rate.

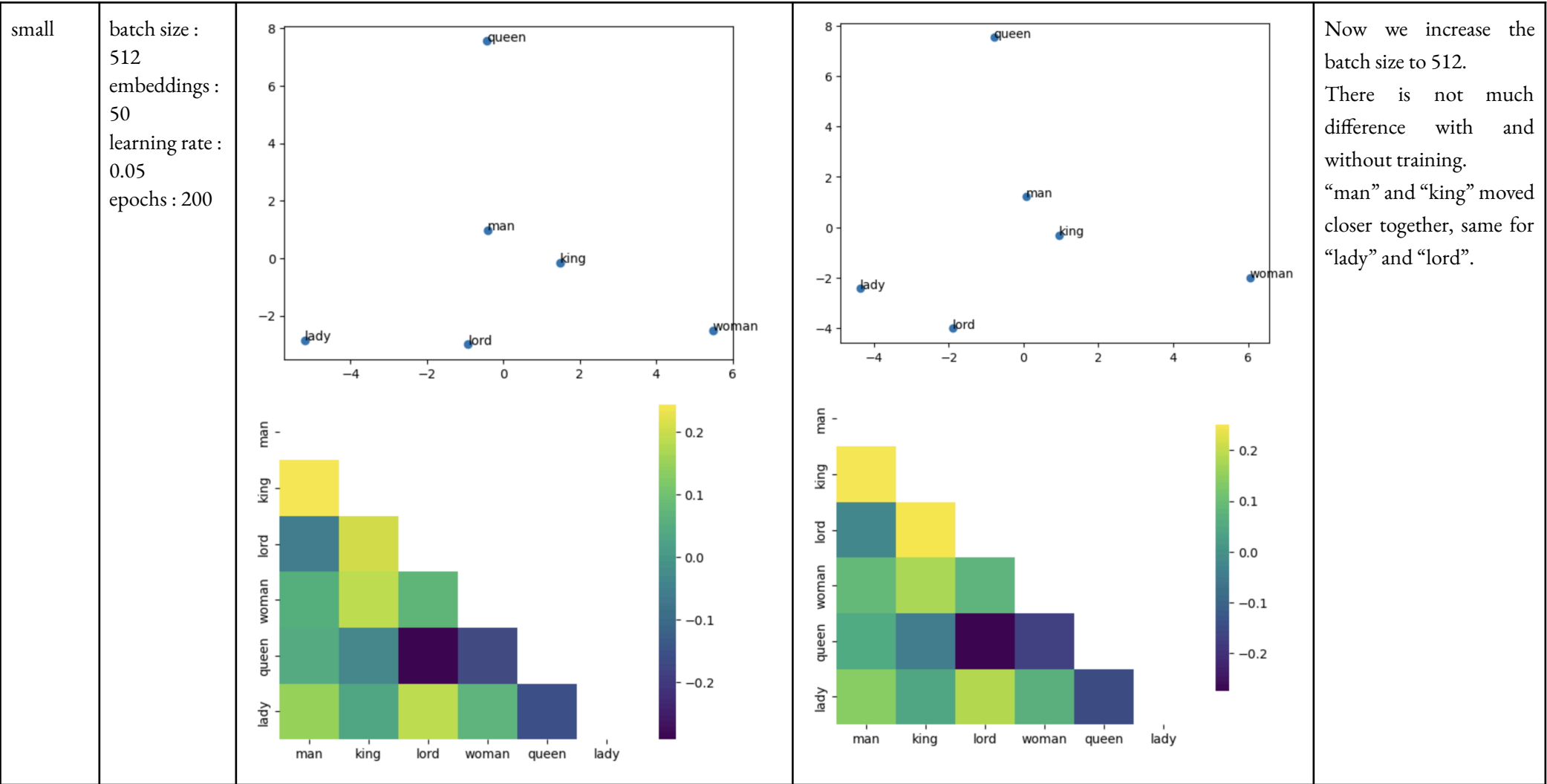
CBOW models recap chart¹

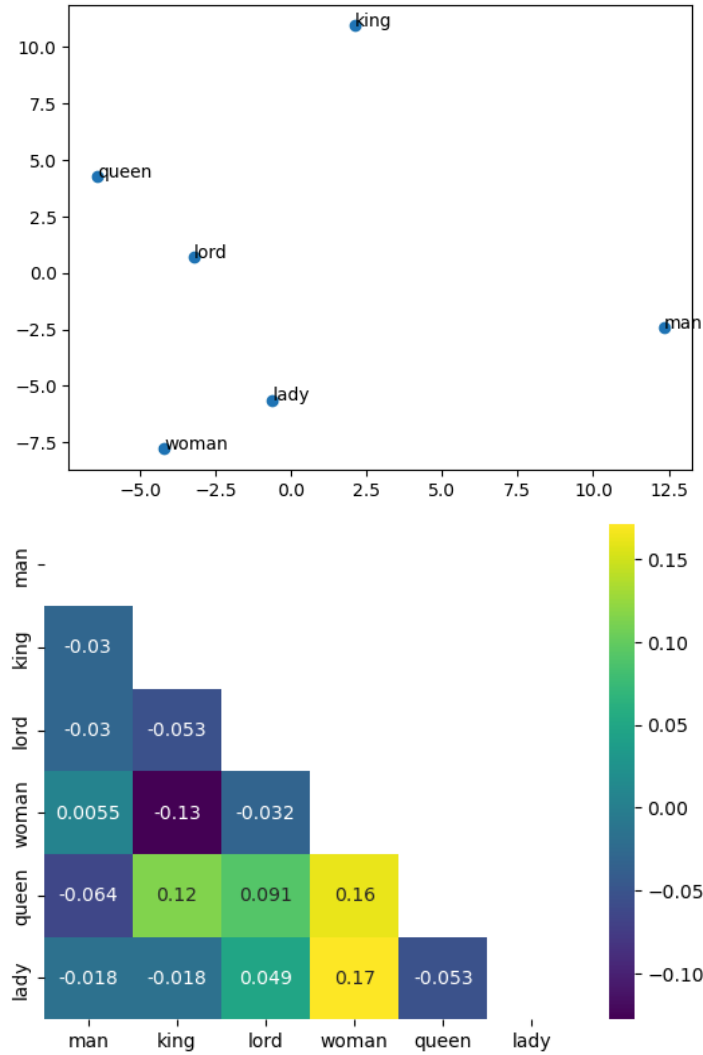
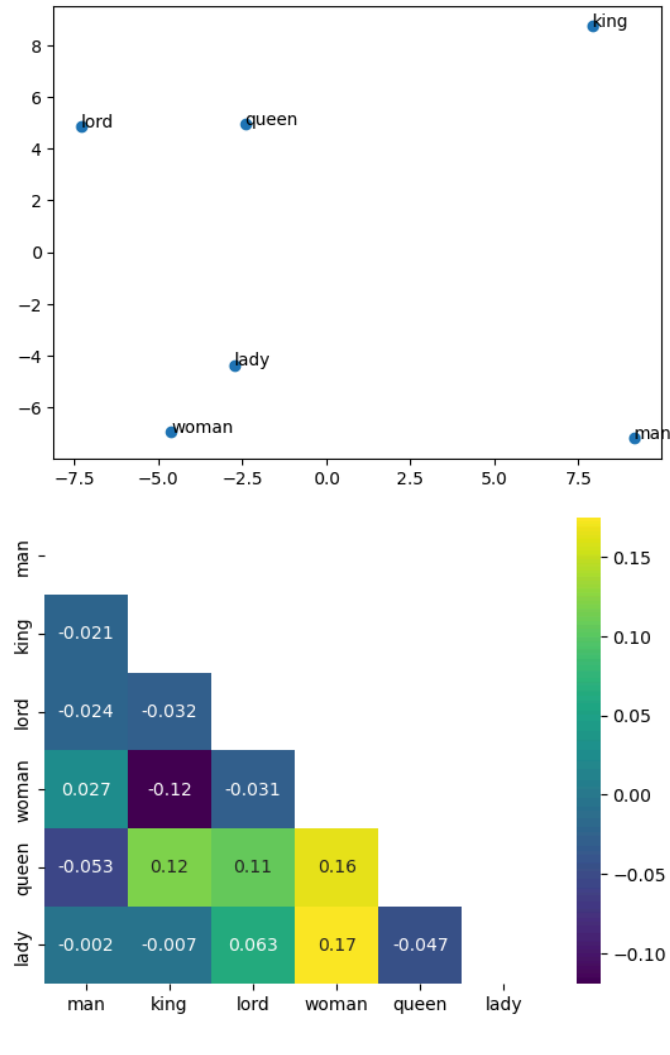
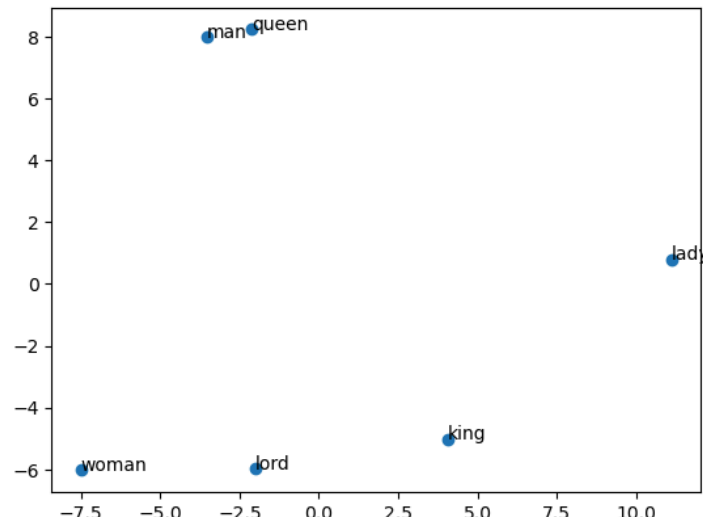
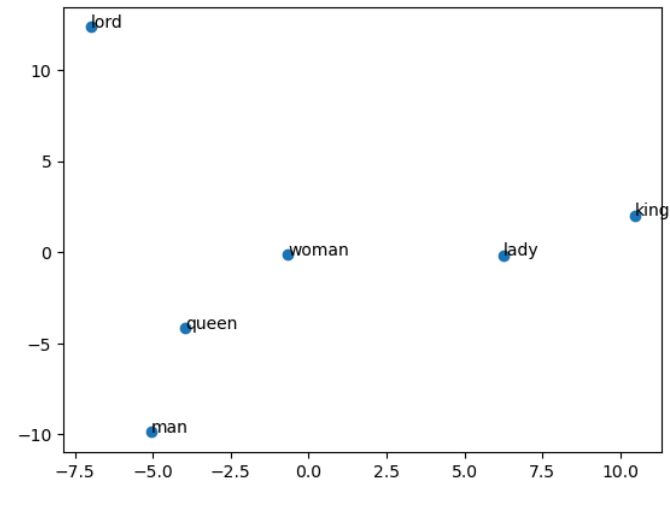
Corpus	Parameters	Results before training	Results after training	Comments																																																																																																		
small	batch size : 128 embeddings : 50 learning rate : 0.1 epochs : 200	 <p>Scatter plot showing word embeddings before training. The x-axis ranges from -4 to 6, and the y-axis ranges from -2 to 6. Points are labeled: king, queen, man, lord, woman, and lady. The heatmap shows pairwise distances between these words, with a color scale from -0.15 (dark purple) to 0.05 (yellow).</p> <table><tr><th></th><th>man</th><th>king</th><th>lord</th><th>woman</th><th>queen</th><th>lady</th></tr><tr><th>man</th><td></td><td></td><td></td><td></td><td></td><td></td></tr><tr><th>king</th><td>0.088</td><td></td><td></td><td></td><td></td><td></td></tr><tr><th>lord</th><td>-0.091</td><td>0.021</td><td></td><td></td><td></td><td></td></tr><tr><th>woman</th><td>0.0029</td><td>0.05</td><td>-0.0062</td><td></td><td></td><td></td></tr><tr><th>queen</th><td>-0.12</td><td>-0.0093</td><td>-0.16</td><td>-0.0058</td><td></td><td></td></tr><tr><th>lady</th><td>0.029</td><td>-0.14</td><td>0.043</td><td>0.084</td><td>-0.048</td><td></td></tr></table>		man	king	lord	woman	queen	lady	man							king	0.088						lord	-0.091	0.021					woman	0.0029	0.05	-0.0062				queen	-0.12	-0.0093	-0.16	-0.0058			lady	0.029	-0.14	0.043	0.084	-0.048		 <p>Scatter plot showing word embeddings after training. The x-axis ranges from -4 to 6, and the y-axis ranges from -4 to 8. Points are labeled: lord, lady, queen, man, king, and woman. The heatmap shows pairwise distances between these words, with a color scale from -0.10 (dark purple) to 0.15 (yellow).</p> <table><tr><th></th><th>man</th><th>king</th><th>lord</th><th>woman</th><th>queen</th><th>lady</th></tr><tr><th>man</th><td></td><td></td><td></td><td></td><td></td><td></td></tr><tr><th>king</th><td>0.13</td><td></td><td></td><td></td><td></td><td></td></tr><tr><th>lord</th><td>0.054</td><td>0.19</td><td></td><td></td><td></td><td></td></tr><tr><th>woman</th><td>0.2</td><td>0.019</td><td>0.0023</td><td></td><td></td><td></td></tr><tr><th>queen</th><td>-0.12</td><td>0.039</td><td>0.081</td><td>-0.00089</td><td></td><td></td></tr><tr><th>lady</th><td>0.045</td><td>-0.062</td><td>0.17</td><td>0.12</td><td>0.0074</td><td></td></tr></table>		man	king	lord	woman	queen	lady	man							king	0.13						lord	0.054	0.19					woman	0.2	0.019	0.0023				queen	-0.12	0.039	0.081	-0.00089			lady	0.045	-0.062	0.17	0.12	0.0074		As we can see, “lord” and “lady” moved closer together (closer values on the X axis), and “lord” and “man” moved farther apart, maybe because here the social class tends to be a more important criteria than gender ?
	man	king	lord	woman	queen	lady																																																																																																
man																																																																																																						
king	0.088																																																																																																					
lord	-0.091	0.021																																																																																																				
woman	0.0029	0.05	-0.0062																																																																																																			
queen	-0.12	-0.0093	-0.16	-0.0058																																																																																																		
lady	0.029	-0.14	0.043	0.084	-0.048																																																																																																	
	man	king	lord	woman	queen	lady																																																																																																
man																																																																																																						
king	0.13																																																																																																					
lord	0.054	0.19																																																																																																				
woman	0.2	0.019	0.0023																																																																																																			
queen	-0.12	0.039	0.081	-0.00089																																																																																																		
lady	0.045	-0.062	0.17	0.12	0.0074																																																																																																	

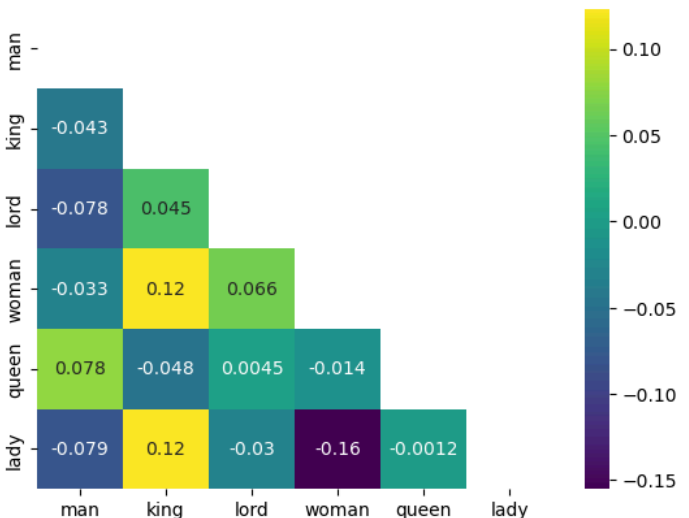
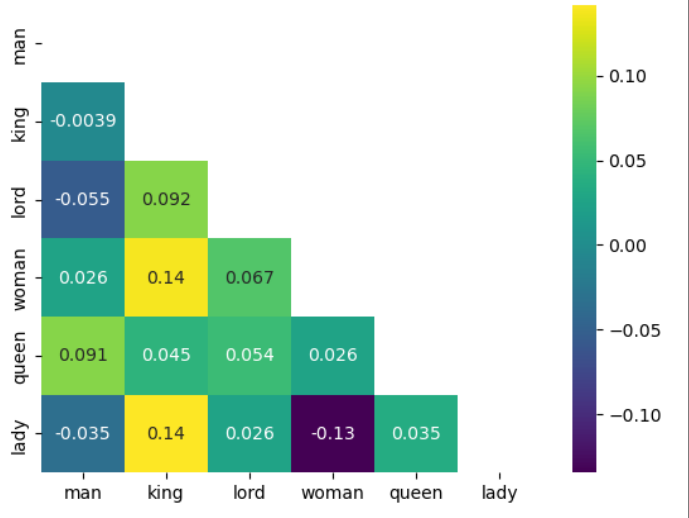
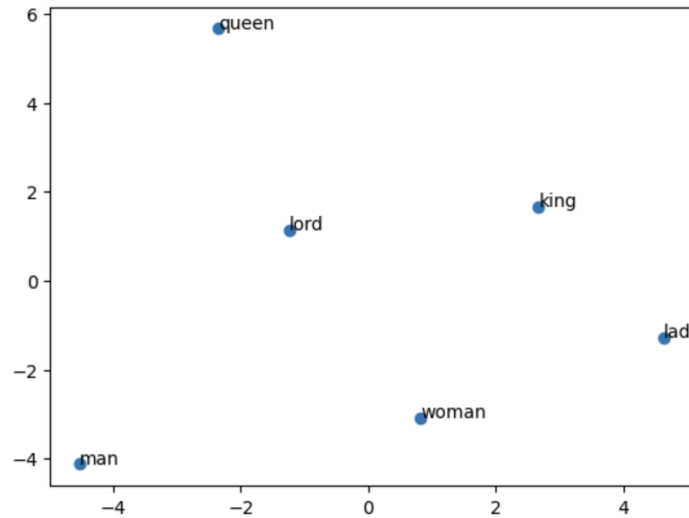
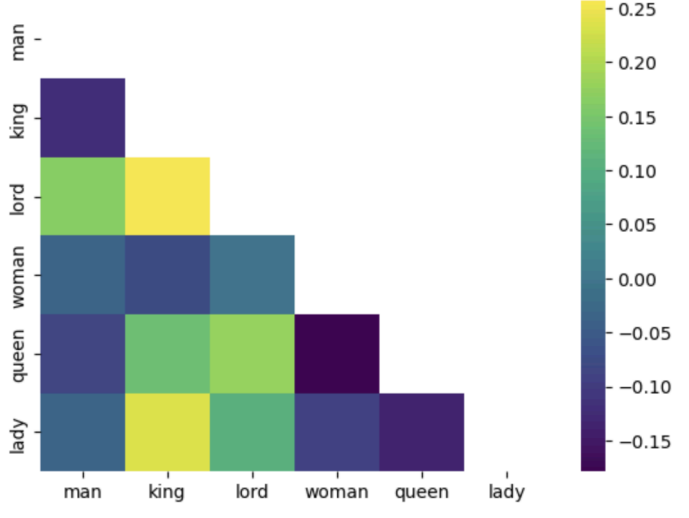
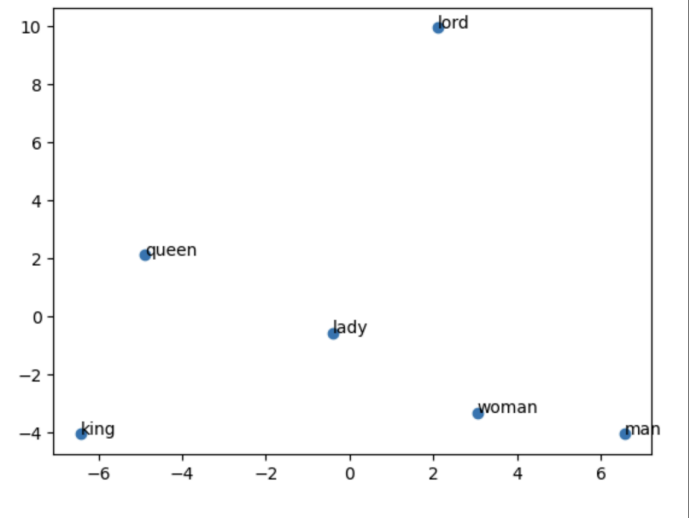
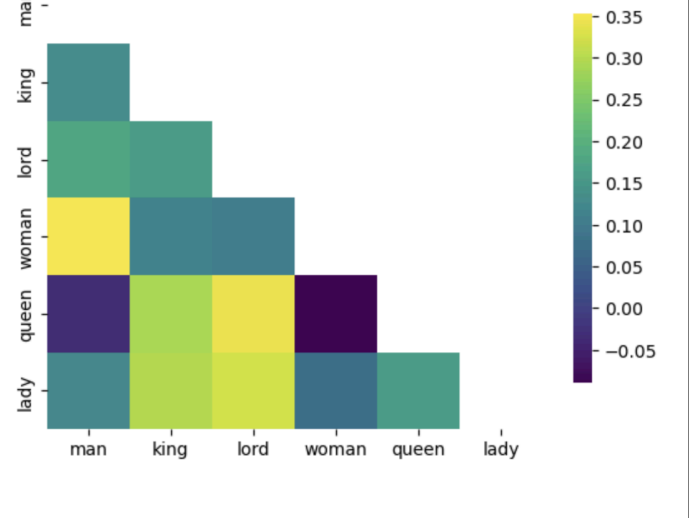
¹ We didn't manage to get the annotations on certain heatmaps because we ran the script on a different computer for some of the attempts because it was faster, but we couldn't solve this problem.







small	batch size : 256 embeddings : 200 learning rate : 0.05 epochs : 200	 <p>Scatter plot (left) and heatmap (right) for the small corpus. The scatter plot shows word embeddings for king, queen, lord, lady, man, and woman. The heatmap shows cosine similarity scores between these words.</p> <table><tr><th></th><th>man</th><th>king</th><th>lord</th><th>woman</th><th>queen</th><th>lady</th></tr><tr><th>man</th><td>-</td><td></td><td></td><td></td><td></td><td></td></tr><tr><th>king</th><td>-0.03</td><td>-</td><td></td><td></td><td></td><td></td></tr><tr><th>lord</th><td>-0.03</td><td>-0.053</td><td>-</td><td></td><td></td><td></td></tr><tr><th>woman</th><td>0.0055</td><td>-0.13</td><td>-0.032</td><td>-</td><td></td><td></td></tr><tr><th>queen</th><td>-0.064</td><td>0.12</td><td>0.091</td><td>0.16</td><td>-</td><td></td></tr><tr><th>lady</th><td>-0.018</td><td>-0.018</td><td>0.049</td><td>0.17</td><td>-0.053</td><td>-</td></tr></table>		man	king	lord	woman	queen	lady	man	-						king	-0.03	-					lord	-0.03	-0.053	-				woman	0.0055	-0.13	-0.032	-			queen	-0.064	0.12	0.091	0.16	-		lady	-0.018	-0.018	0.049	0.17	-0.053	-	 <p>Scatter plot (left) and heatmap (right) for the small corpus. The scatter plot shows word embeddings for king, queen, lord, lady, man, and woman. The heatmap shows cosine similarity scores between these words.</p> <table><tr><th></th><th>man</th><th>king</th><th>lord</th><th>woman</th><th>queen</th><th>lady</th></tr><tr><th>man</th><td>-</td><td></td><td></td><td></td><td></td><td></td></tr><tr><th>king</th><td>-0.021</td><td>-</td><td></td><td></td><td></td><td></td></tr><tr><th>lord</th><td>-0.024</td><td>-0.032</td><td>-</td><td></td><td></td><td></td></tr><tr><th>woman</th><td>0.027</td><td>-0.12</td><td>-0.031</td><td>-</td><td></td><td></td></tr><tr><th>queen</th><td>-0.053</td><td>0.12</td><td>0.11</td><td>0.16</td><td>-</td><td></td></tr><tr><th>lady</th><td>-0.002</td><td>-0.007</td><td>0.063</td><td>0.17</td><td>-0.047</td><td>-</td></tr></table>		man	king	lord	woman	queen	lady	man	-						king	-0.021	-					lord	-0.024	-0.032	-				woman	0.027	-0.12	-0.031	-			queen	-0.053	0.12	0.11	0.16	-		lady	-0.002	-0.007	0.063	0.17	-0.047	-	<p>Here, we can notice that the cosine similarity score hasn't changed too much : we have the same dynamic.</p> <p>For the plot, we can see that “man” and “king” moved closer together (close values on the X axis), same for “lady” and “queen”.</p>
	man	king	lord	woman	queen	lady																																																																																																
man	-																																																																																																					
king	-0.03	-																																																																																																				
lord	-0.03	-0.053	-																																																																																																			
woman	0.0055	-0.13	-0.032	-																																																																																																		
queen	-0.064	0.12	0.091	0.16	-																																																																																																	
lady	-0.018	-0.018	0.049	0.17	-0.053	-																																																																																																
	man	king	lord	woman	queen	lady																																																																																																
man	-																																																																																																					
king	-0.021	-																																																																																																				
lord	-0.024	-0.032	-																																																																																																			
woman	0.027	-0.12	-0.031	-																																																																																																		
queen	-0.053	0.12	0.11	0.16	-																																																																																																	
lady	-0.002	-0.007	0.063	0.17	-0.047	-																																																																																																
mid	batch size : 256 embeddings : 200 learning rate : 0.05 epochs : 200	 <p>Scatter plot for the mid corpus. The scatter plot shows word embeddings for man, queen, lady, woman, lord, and king.</p>	 <p>Scatter plot for the mid corpus. The scatter plot shows word embeddings for lord, woman, king, queen, man, and lady.</p>	<p>Experimenting on larger corpus takes a lot more time so we couldn't try many variations, but we can see the differences when we only change the corpus size if we compare this result with what we obtained previously.</p> <p>We can notice that with different sizes of corpus, the plots and heatmap are really different.</p> <p>For the comparison before and after training, we can see “lord” isolated, “queen”,</p>																																																																																																		

				<p>“woman” and “lady” a lot closer (with close values on the Y axis).</p> <p>“king”, “man” and “lord” stay away from each other.</p>
full	batch size : 128 learning rate : 0.05 embeddings : 50 epochs : 150	 	 	<p>We can hardly read the heatmaps, but we can see that the cosine similarity score of “king” and “man” has increased and became positive (we can see it with the color change). Same for “lady” and “queen”, “lady” and “woman”.</p> <p>“lord” and “king” moved farther apart and their cosine similarity score decreased.</p>