Q1: Sentiment Use

a. The simple sentences

Simple 1: “john fell down harry fellas-well mary was fine down by the stream the sun shone before it went down.”

Simple 2: “bill fell down jeff fell too down by the river the sun shone until it sunk down belinda was ill.”

The K-L score as follows.

```
KL-divergence between d1 and d2: 3.704719577426546
KL-divergence between d2 and d1: 3.0973054936655595
```

We can see that d1 and d2 do not have high scores, and the scores of d1 and d2 are not equal to the scores of d2 and d1. Since there are many repeated words in the two sentences, the similarity is high, and the divergence is relatively small. However, the scores in both directions are not the same, implying an asymmetry in KL scores.

b. Third sentence

Third sentence: “We were raised to believe that we were scattered on this world as nations and tribes so we can learn from each other and find beauty in our differences.”

The K-L score as follows.

```
KL-divergence between d1 and d2: 3.704719577426546
KL-divergence between d2 and d1: 3.0973054936655595
KL-divergence between d1 and d3: 7.399158944437483
KL-divergence between d2 and d3: 7.559106567866163
KL-divergence between d3 and d1: 7.10524155559411
KL-divergence between d3 and d2: 7.10524155559411
```

The third sentence was chosen entirely at random from Twitter. The content is almost irrelevant to the first two sentences. We saw an increase in KL scores for d3, d1, and d2. This score makes sense because the first and second sentences have very little in common with the third sentence and are entirely different in content. So they are significantly divergent from each other, and the corresponding KL scores increase accordingly.

c. epsilon and gamma

Since two sentences, there may be cases where all the words differ. One sentence has words that are not in the other sentence. Therefore, this situation needs to be considered when calculating the KL score. Thus, a back-off smoothing model is introduced.

In this model, Epsilon then assigns weights to those words that do not appear in the model sentences. However, it is a minimum value that cannot be equal to 0, preventing the KL score from tending to infinity. On the other hand, Gamma is a normalization factor that makes the probability of a term in the range of 0 to 1 satisfy the probability property.

I am scaling down the epsilon by a factor of 1000 in the above calculation results in the following.

```
KL-divergence between d1 and d2: 8.584550747686238
KL-divergence between d2 and d1: 7.129631609778013
KL-divergence between d1 and d3: 16.614248203984346
KL-divergence between d2 and d3: 16.773633407017847
KL-divergence between d3 and d1: 16.32166516451297
KL-divergence between d3 and d2: 16.32166516451297
```

Since the vast majority of words in the third sentence did not overlap with the first two sentences, those words were given smaller values, resulting in an increased score for KL.