Statistical NLP Assignment 4 Report

Jiyuan Lu   jl11046

Question 2:

Performance of the model:

	precision	recall	F1-Score
Total:	0.221961	0.525544	0.312106
PER:	0.435451	0.231230	0.302061
ORG:	0.475936	0.399103	0.434146
LOC:	0.147750	0.870229	0.252612
MISC:	0.491689	0.610206	0.544574

Observations:

LOC has the lowest precision but the highest recall and it has the lowest F1-score.

MISC has the highest precision, and it has the highest F1-score.

PER has the lowest recall.

Question 4:

Performance of the model:

	precision	recall	F1-Score
Total:	0.775298	0.614905	0.685849
PER:	0.762535	0.595756	0.668907
ORG:	0.611855	0.478326	0.536913
LOC:	0.876458	0.696292	0.776056
MISC:	0.830065	0.689468	0.753262

Observations:

For PER, ORG and MISC, their precision, recall and F1-score all improved.

For LOC, its precision greatly improved but its recall slightly degraded.

ORG now has the lowest precision, recall, and F1-score.

LOC now has the highest precision, recall, and F1-score.

By using an HMM tagger instead of a simple emission tagger, the improvement is the least for ORG, and the most for LOC.