Subset	Words	Characters
Training	599,126	3,739,262
Development	5,000	31,416
Test	5,000	31,239