1.2 N-gram Language Modeling Perplexity (28%)

(a)

(b)

(c)

(d)

(e)

1.3 Interpolation (17%)

(a)

(b)

(c)

(d)

2 Byte-Pair Encoding (BPE) (30%)

(a)

(b)

(c)

3 WordPiece (25%)

(a)

(b)

(c)

4 Open-ended Exploration for Tokenization (10% for 517, extra credit for 447)

(a)