

LING 570: Hw5
Due date: 11:45pm on Nov 4

All the examples are under `~/dropbox/09-10/570/hw5/examples/`.

Q1 (20 points): write a script, **ngram_count.sh**, that collects unigrams, bigrams, and trigrams.

⇒ See `ngram_count-fall09.pl`

Q2 (20 points): write a script, **build_lm.sh**, that builds an LM using ngram counts:

⇒ See `build_lm.pl`

Q3 (30 points): Write a script, **ppl.sh**, that calculates the perplexity of a test data given an LM.

⇒ See `ppl.pl`

Q4 (30 points) Use `examples/wsj sec0 19.word` as training data, and calculate the perplexity of `examples/wsj sec22.word` by running the codes created in Q1-Q3. Fill out the following table.

lambda_1	lambda_2	lambda_3	perplexity
0.05	0.15	0.8	373.02
0.1	0.1	0.8	349.29
0.2	0.3	0.5	228.89
0.2	0.5	0.3	210.03
0.2	0.7	0.1	211.29
0.2	0.8	0	236.52
1.0	0	0	998.67