

LING 570: Hw6

Q1 (15 points): Write a script, **create_2gram_hmm.sh**, that takes the annotated training data as input and creates an HMM for a bigram POS tagger with NO smoothing.

⇒ See `create_2gram_hmm.pl`

See slide #7 on 11_04_POS_p2.pdf, each state in the `output_hmm` corresponds to a tag (a.k.a. a POS tag or BOS or EOS).

Q2 (30 points): Write a script, **create_3gram_hmm.sh**, that takes the annotated training data as input and creates an HMM for a trigram POS tagger WITH smoothing.

⇒ See `create_3gram_hmm.pl`

See slide #9 on 11_04_POS_p2.pdf, each state in the `output_hmm` corresponds to a tag **pair** (a tag is a POS tag or BOS or EOS).

Q3 (40 points): Write a script, **check_hmm.sh**, that reads in a state-emission HMM file, check its format, and output a warning file.

⇒ See `check_hmm.pl` and `hmm.pm`

Q4 (15 points): Run the following commands and turn in the files generated by the commands:

⇒ See `2g_hmm*` and `3g_hmm*`