Manual unigrams, special token probability=0.6, entropy=2.3040