Vocabulary Growth Curves 1000 1200 1400 V (types) observed interpolation N (k tokens)