

corpus coverage vs vocab size

