Kneser-Ney smoothing

Idea:

- The unigram distribution captures the word frequency
- We need to capture the diversity of contexts for the word

$$\hat{p}(w) \propto |x : c(x\,w) > 0|$$

malt

This is the …

Kong

- Probably, the most popular smoothing technique

https://web.stanford.edu/~jurafsky/slp3/4.pdf