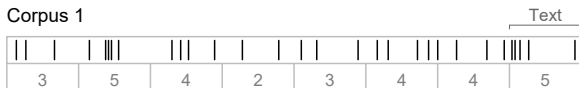


(a) Corpus-linguistic sense: Distribution of word tokens in the corpus

Corpus 1



Higher dispersion  
(more spread out)

Corpus 2



Lower dispersion  
(more concentrated)

Text-level  
token count

(b) Statistical sense: Distribution of text-level occurrence rates

Corpus 1



Lower dispersion  
(more concentrated)

Each dot  
denotes a text

Corpus 2



Higher dispersion  
(more spread out)

Number of occurrences