

Lunghezza del corpus ($V_i = \text{types}, C = \text{lung.corpus}$)	$ C = V_1 + V_2 + V_3 + \dots + V_n $
Frequenza Relativa di una parola: ($V_i = \text{types}, C = \text{lung.corpus}$)	$f(V_i) = \frac{ V_i }{ C }$ perc: $f(\%) = \frac{ V_i }{ C } \cdot 100$
Classe di Frequenza: ($V_i = \text{types.difreq}(i), V_C = \text{voc.corpus}$)	$ V_C = V_1 + V_2 + V_3 + \dots + V_{f_{\max}} $
Rapporto tipo-unità (TTR): ($C = \text{lung.corpus}, V_C = \text{voc.corpus}$)	$0 \leq \frac{ V_C }{ C } \leq 1$ Square TTR: $\frac{ V_C }{\sqrt{ C }}$
Distribuzione degli hapax: ($C = \text{lung.corpus}, V_1 = \text{types.difreq}(1)$)	$\frac{ V_1 }{ C }$
Frequenze cumulate di parole unità: ($f_i^c = \text{freq. cum. di val. } i, V_i = \text{types.difreq}(i)$)	$f_i^c = V_1 * 1 + V_2 * 2 + \dots + V_i * i$
Frequenze cumulate di parole tipo: ($V_i^c = \text{freq. cum. di val. } i, V_i = \text{types.difreq}(i)$)	$V_i^c = V_1 + V_2 + V_3 + \dots + V_i $
Frequenze relative cumulate di parole unità:	$\frac{f_i^c}{ C }$
Frequenze relative cumulate di parole tipo:	$\frac{V_i^c}{ V_C }$
Media Aritmetica: ($x_i = \text{valore}, n = \text{num.valori}$)	$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$
Deviazione Standard (scarto quadratico medio): ($\bar{x} = \text{media}, x_i = \text{valore}, n = \text{num.valori}$)	$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$ (varianza = σ^2)
Legge di Zipf: ($C = \text{cost. corrispondente alla parola di rango } 1, z = \text{rango}, a = \text{indice inverso della ricchezza lessicale.}$)	$f(z) = \frac{C}{z^a}$ Per Zipf: $a \approx 1$
Legge di Zipf in doppia scala logaritmica: ($C = \text{cost. corrispondente alla parola di rango } 1, z = \text{rango}, a = \text{indice inverso della ricchezza lessicale}$)	$\log f(z) = \log C - a \cdot \log z$
Frequenza media delle parole in C: ($V_C = \text{voc.corpus}, C = \text{lung.corpus}$)	$\bar{f}(C) = \frac{ C }{ V_C } = \frac{f_{V_1} + f_{V_2} + \dots + f_{V_n}}{n}$
Probabilità di trovare una parola V nel Corpus: ($C = \text{lung.corpus}, f_v = \text{freq. parola}$)	$P(v) \approx \frac{f_v}{ C }$
Probabilità dell'unione di eventi:	$P(A_1 \cup A_2 \cup \dots \cup A_n) = \sum_{i=1}^n P(A_i)$ ($A_i = \text{evento}$)
Probabilità congiunta (eventi indipendenti): ($A/B = \text{evento}$)	$P(A \cap B) = P(A) \cdot P(B)$
Probabilità congiunta (eventi dipendenti): ($A/B = \text{evento}$)	$P(A \cap B) = P(B) \cdot P(A B)$
Probabilità condizionata: ($A/B = \text{evento}$)	$P(A B) = \frac{P(A \cap B)}{P(B)}$
Mutua Informazione: ($V_i = \text{parola}, C = \text{lung.corpus}, f = \text{freq.rel.}$)	$MI(V_1, V_2) = \log_2 \frac{P(V_1, V_2)}{P(V_1) \cdot P(V_2)} = \log_2 \frac{f(V_1, V_2) \cdot C }{f(V_1) \cdot f(V_2)}$

n.b. $\rightarrow \log_2 x = \log_{10} x - \log_{10} 2$