

## R

### 共起スコア

## MI スコア

### 計算の仕方

1. 「strong tea」の共起頻度
2. 「strong」の頻度
3. 「tea」の頻度
4. コーパスの総語数

MI スコア =  $\log_2((\text{共起頻度} \times \text{コーパス総語数}) \div (\text{片方の頻度} \times \text{もう片方の頻度}))$

### 特徴

- ・低頻度の単語によるまれな共起表現ほどスコアが大きくなる
- ・分母が小さくなるので

### R のスクリプト

- ・使い方
  - ・このスクリプトを R で読み込んで、
  - ・後は必要な頻度数と総語数を入れる。

共起表現 xy 頻度、単語 x の頻度、単語 y の頻度 コーパスの総語数 N

MI(xy, x, y, N)