Correlation: Lexical graduation in the word 'mito' in YouTube Users Comments

Rodrigo Esteves de Lima-Lopes
State University of Campinas | rll307@unicamp.br
Carolina Palma de Sousa Arruda
State University of Campinas | carolpalma203@gmail.com

Contents

1 Introduction

This script was developed for the analysis of Portuguese. We hope it helps colleagues in the LC area and popularize the use of R. It is part of our research project developed with CNPQ. Please drop us a line if you have any doubts or need any help.

1.1 Purpose of this repository

This repository brings the scripts used for lexical analysis of Brazilian Portuguese YouTube comments we wrote for our paper:

> Lima-Lopes, R. E. de and Arruda, C. P. S. (2020) Strategies for Gradation in YouTube comments. (under review)

This script is specifically about correlation analysis.

1.2 Packages

For the correlation we are going to use two packages:

1. `stats` is responsible for the basic statistic analysis. Usually it loads together with `R`
2. `data.table` is responsible for some table manipulation
3. `tidyverse` is responsible for data manipulation

2 Running the test

Load the matrix of strategies:

```
load("correlation.RData")
```

After loading the data, normalise the matrix by 1000:

```
occurences_relative <- matrix_occurences %>%
  mutate_at(
    vars(-matches("word|total")),
    ~ .x * 1000 / total
  )
```

Now we delete unnecessary data:

```
occurences_relative$word <- NULL
occurences_relative$total <- NULL
```

Creating the matrix:

```
occurences_relative <- as.matrix(occurences_relative)
rownames(occurences_relative)<-matrix_occurences$word
```

Running the test:

```
correlation<- cor(occurences_relative)
```