Dictionary Analysis

Riccardo Ruta

7/5/2022

Contents

RICORSO ALLA RETORICA POPULISTA.

A livello di partiti politici, quali fanno maggiormente ricorso alla retorica populista?

A livello di singoli politici, quali fanno maggiormente ricorso alla retorica populista?

Per rispondere a questa domanda, ricorrere ai seguenti dizionari 3(+1) dizionari:

- Rooduijn & Pauwels: Rooduijn, M., and T. Pauwels. 2011. "Measuring Populism: Comparing Two Methods of Content Analysis." West European Politics 34 (6): 1272–1283.
- Decadri & Boussalis: Decadri, S., & Boussalis, C. (2020). Populism, party membership, and language complexity in the Italian chamber of deputies. Journal of Elections, Public Opinion and Parties, 30(4), 484-503.
- Grundl: Gründl J. Populist ideas on social media: A dictionary-based measurement of populist communication. New Media & Society. December 2020.
- Decadri & Boussalis + Grundl: questo è semplicemente una versione più estesa del dizionario D&B, che contiene anche alcuni termini presi da Grundl.

1) First step, import the words and create the dictionary

```
# import dictionaries file
dict <-  read_excel("data/populism_dictionaries.xlsx")
variable.names(dict)
```

```
## [1] "Rooduijn_Pauwels_Italian"
## [2] "Grundl_Italian_adapted"
## [3] "Decadri_Boussalis"
## [4] "Decadri_Boussalis_Grundl_People"
## [5] "Decadri_Boussalis_Grundl_Common Will"
## [6] "Decadri_Boussalis_Grundl_Elite"
```

```
# create the dictionary
Rooduijn_Pauwels_Italian <-
  dictionary(list(populism =
                    (dict$Rooduijn_Pauwels_Italian
                     [!is.na(dict$Rooduijn_Pauwels_Italian)])))

Grundl_Italian_adapted <-
  dictionary(list(populism =
                    dict$Grundl_Italian_adapted
                  [!is.na(dict$Grundl_Italian_adapted)]))

Decadri_Boussalis <-
  dictionary(list(populism =
                    dict$Decadri_Boussalis
                  [!is.na(dict$Decadri_Boussalis)]))

Decadri_Boussalis_Grundl <-
  dictionary(list(people =
                    dict$Decadri_Boussalis_Grundl_People
                  [!is.na(dict$Decadri_Boussalis_Grundl_People)],
                  common_will =
                    dict$`Decadri_Boussalis_Grundl_Common Will`
                  [!is.na(dict$`Decadri_Boussalis_Grundl_Common Will`)],
                  elite =
                    dict$Decadri_Boussalis_Grundl_Elite
                  [!is.na(dict$Decadri_Boussalis_Grundl_Elite)]))
```

```
my_dictionary <- dictionary(list(populism = c(Rooduijn_Pauwels_Italian$populism,
                                              Grundl_Italian_adapted$populism,
                                              Decadri_Boussalis$populism,
                                              Decadri_Boussalis_Grundl$people,
                                              Decadri_Boussalis_Grundl$common_will,
                                              Decadri_Boussalis_Grundl$elite)))
head(my_dictionary$populism)
```

I also create one extra dictionary that include all the populist words

```
## [1] "antidemocratic*" "casta"           "consens*"        "corrot*"
## [5] "disonest*"       "elit*"
```

```
tail(my_dictionary$populism)
```

```
## [1] "raccomandati"
## [2] "bugie dei partiti, falsita dei partiti"
## [3] "mazzett?"
## [4] "prendere in giro, bullarsi di"
## [5] "banchier?"
## [6] "lobbist*"
```

2) Import the DFM prepared in previous steps and apply dictionaries

```
# Daily Dictionary analysis with Decadri_Boussalis_Grundl on the whole dataset
dfm_dict1  <- dfm_lookup(DFM_trimmed, dictionary = Decadri_Boussalis_Grundl)
# Group by date
dfm_by_date1 <- dfm_group(dfm_dict1, groups= date)
dfm_by_date1
```

Decadri_Boussalis_Grundl

```
## Document-feature matrix of: 839 documents, 3 features (12.08% sparse) and 3 docvars.
##             features
## docs         people common_will elite
##   2020-01-01      1           0     1
##   2020-01-02      7           1     6
##   2020-01-03      8           1     6
##   2020-01-04     20           0     4
##   2020-01-05     22           1     3
##   2020-01-06      8           0     4
## [ reached max_ndoc ... 833 more documents ]
```

```
# Group by week
dfm_by_week1 <- dfm_group(dfm_dict1, groups= week)
dfm_by_week1
```

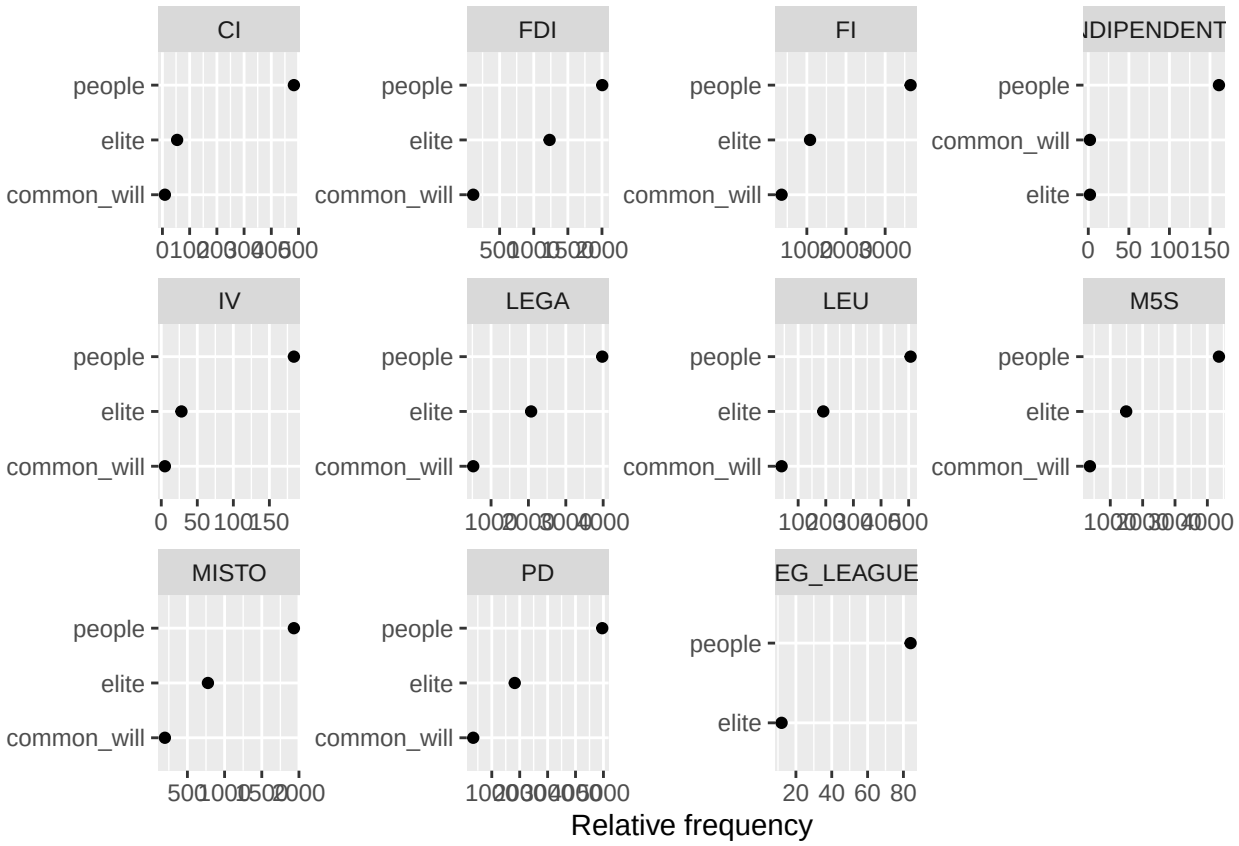
```
## Document-feature matrix of: 121 documents, 3 features (0.55% sparse) and 1 docvar.
##     features
## docs people common_will elite
##    1     58           3    20
##    2    154          26    54
##    3    232          25    91
##    4    248          11    99
##    5    249          17    73
##    6    166           7    83
## [ reached max_ndoc ... 115 more documents ]
```

```
# Group by month
dfm_by_month1 <- dfm_group(dfm_dict1, groups= month)
dfm_by_month1
```

```
## Document-feature matrix of: 28 documents, 3 features (0.00% sparse) and 1 docvar.
##     features
## docs people common_will elite
##    1    890          80   334
##    2    763          42   341
##    3    892          37   248
##    4    768          21   415
##    5    689          14   376
##    6    636          24   424
## [ reached max_ndoc ... 22 more documents ]
```

```
tstat1 <- textstat_frequency(dfm_dict1, groups = party_id)

ggplot(data = tstat1, aes(x = factor(nrow(tstat1):1), y = frequency)) +
    geom_point() +
    facet_wrap(~ group, scales = "free") +
    coord_flip() +
    scale_x_discrete(breaks = nrow(tstat1):1,
                     labels = tstat1$feature) +
    labs(x = NULL, y = "Relative frequency")
```




```
# Daily Dictionary analysis with Rooduijn_Pauwels_Italian on the whole dataset
dfm_dict2  <- dfm_lookup(DFM_trimmed, dictionary = Rooduijn_Pauwels_Italian)
# Group by date
dfm_by_date2 <- dfm_group(dfm_dict2, groups= date)
dfm_by_date2
```

Rooduijn_Pauwels_Italian

```
## Document-feature matrix of: 839 documents, 1 feature (0.60% sparse) and 3 docvars.
##             features
## docs         populism
##   2020-01-01        1
```

```
##   2020-01-02        5
##   2020-01-03        6
##   2020-01-04        4
##   2020-01-05        3
##   2020-01-06        4
## [ reached max_ndoc ... 833 more documents ]
```

```
# Group by week
dfm_by_week2 <- dfm_group(dfm_dict2, groups= week)
dfm_by_week2
```

```
## Document-feature matrix of: 121 documents, 1 feature (0.00% sparse) and 1 docvar.
##     features
## docs populism
##    1       19
##    2       50
##    3       84
##    4       92
##    5       67
##    6       77
## [ reached max_ndoc ... 115 more documents ]
```

```
# Group by month
dfm_by_month2 <- dfm_group(dfm_dict2, groups= month)
dfm_by_month2
```

```
## Document-feature matrix of: 28 documents, 1 feature (0.00% sparse) and 1 docvar.
##     features
## docs populism
##    1      309
##    2      311
##    3      233
##    4      399
##    5      336
##    6      400
## [ reached max_ndoc ... 22 more documents ]
```

```
# Daily Dictionary analysis with Grundl_Italian_adapted on the whole dataset
dfm_dict3  <- dfm_lookup(DFM_trimmed, dictionary = Grundl_Italian_adapted)
# Group by date
dfm_by_date3<- dfm_group(dfm_dict3, groups= date)
dfm_by_date3
```

Grundl_Italian_adapted

```
## Document-feature matrix of: 839 documents, 1 feature (0.24% sparse) and 3 docvars.
##             features
## docs         populism
##   2020-01-01        0
```

```
##   2020-01-02        3
##   2020-01-03        3
##   2020-01-04        1
##   2020-01-05       20
##   2020-01-06       23
## [ reached max_ndoc ... 833 more documents ]
```

```
# Group by week
dfm_by_week3 <- dfm_group(dfm_dict3, groups= week)
dfm_by_week3
```

```
## Document-feature matrix of: 121 documents, 1 feature (0.00% sparse) and 1 docvar.
##     features
## docs populism
##    1       27
##    2       98
##    3       97
##    4       73
##    5       91
##    6       69
## [ reached max_ndoc ... 115 more documents ]
```

```
# Group by month
dfm_by_month3 <- dfm_group(dfm_dict3, groups= month)
dfm_by_month3
```

```
## Document-feature matrix of: 28 documents, 1 feature (0.00% sparse) and 1 docvar.
##     features
## docs populism
##    1      376
##    2      331
##    3      293
##    4      413
##    5      445
##    6      278
## [ reached max_ndoc ... 22 more documents ]
```

```
# Daily Dictionary analysis with Decadri_Boussalis on the whole dataset
dfm_dict4  <- dfm_lookup(DFM_trimmed, dictionary = Decadri_Boussalis)
# Group by date
dfm_by_date4<- dfm_group(dfm_dict4, groups= date)
dfm_by_date4
```

Decadri_Boussalis

```
## Document-feature matrix of: 839 documents, 1 feature (0.00% sparse) and 3 docvars.
##             features
## docs         populism
##   2020-01-01        2
```

```
##   2020-01-02       13
##   2020-01-03       16
##   2020-01-04       24
##   2020-01-05       26
##   2020-01-06       13
## [ reached max_ndoc ... 833 more documents ]
```

```
# Group by week
dfm_by_week4 <- dfm_group(dfm_dict4, groups= week)
dfm_by_week4
```

```
## Document-feature matrix of: 121 documents, 1 feature (0.00% sparse) and 1 docvar.
##     features
## docs populism
##    1       81
##    2      215
##    3      324
##    4      346
##    5      329
##    6      252
## [ reached max_ndoc ... 115 more documents ]
```

```
# Group by month
dfm_by_month4 <- dfm_group(dfm_dict4, groups= month)
dfm_by_month4
```

```
## Document-feature matrix of: 28 documents, 1 feature (0.00% sparse) and 1 docvar.
##     features
## docs populism
##    1     1237
##    2     1106
##    3     1132
##    4     1183
##    5     1045
##    6     1059
## [ reached max_ndoc ... 22 more documents ]
```

```
# Daily Dictionary analysis with my_dictionary on the whole dataset
dfm_dict5  <- dfm_lookup(DFM_trimmed, dictionary = my_dictionary)
# Group by date
dfm_by_date5<- dfm_group(dfm_dict5, groups= date)
dfm_by_date5
```

My dictionary

```
## Document-feature matrix of: 839 documents, 1 feature (0.00% sparse) and 3 docvars.
##             features
## docs         populism
##   2020-01-01        2
```

```
##   2020-01-02       16
##   2020-01-03       18
##   2020-01-04       25
##   2020-01-05       45
##   2020-01-06       36
## [ reached max_ndoc ... 833 more documents ]
```

```
# Group by week
dfm_by_week5 <- dfm_group(dfm_dict5, groups= week)
dfm_by_week5
```

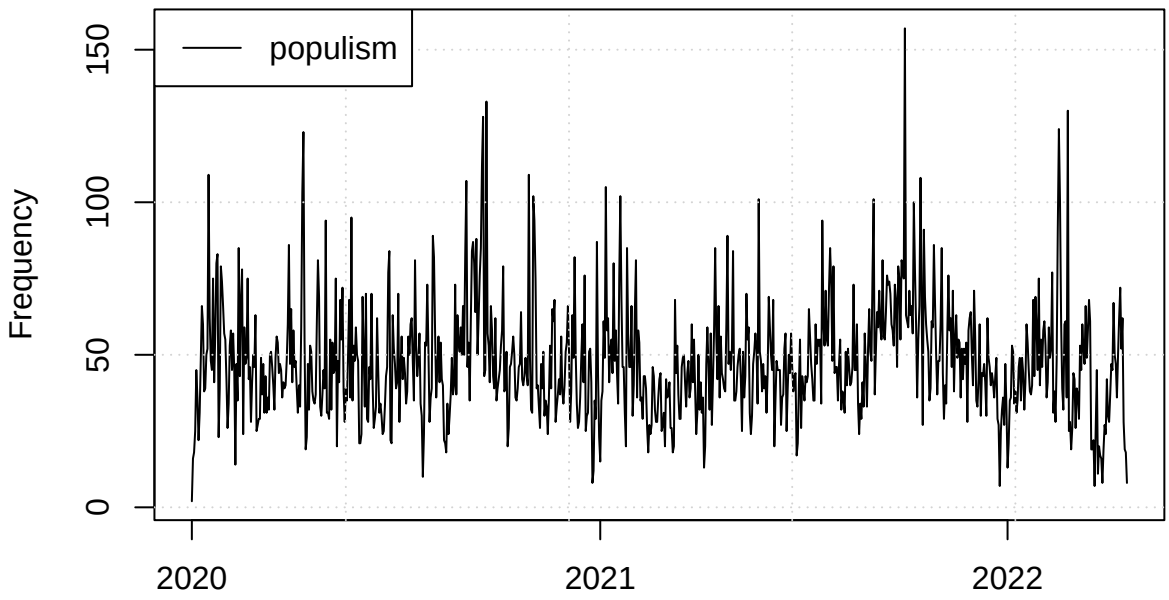
```
## Document-feature matrix of: 121 documents, 1 feature (0.00% sparse) and 1 docvar.
##     features
## docs populism
##    1      106
##    2      301
##    3      405
##    4      398
##    5      399
##    6      306
## [ reached max_ndoc ... 115 more documents ]
```

```
# Group by month
dfm_by_month5 <- dfm_group(dfm_dict5, groups= month)
dfm_by_month5
```

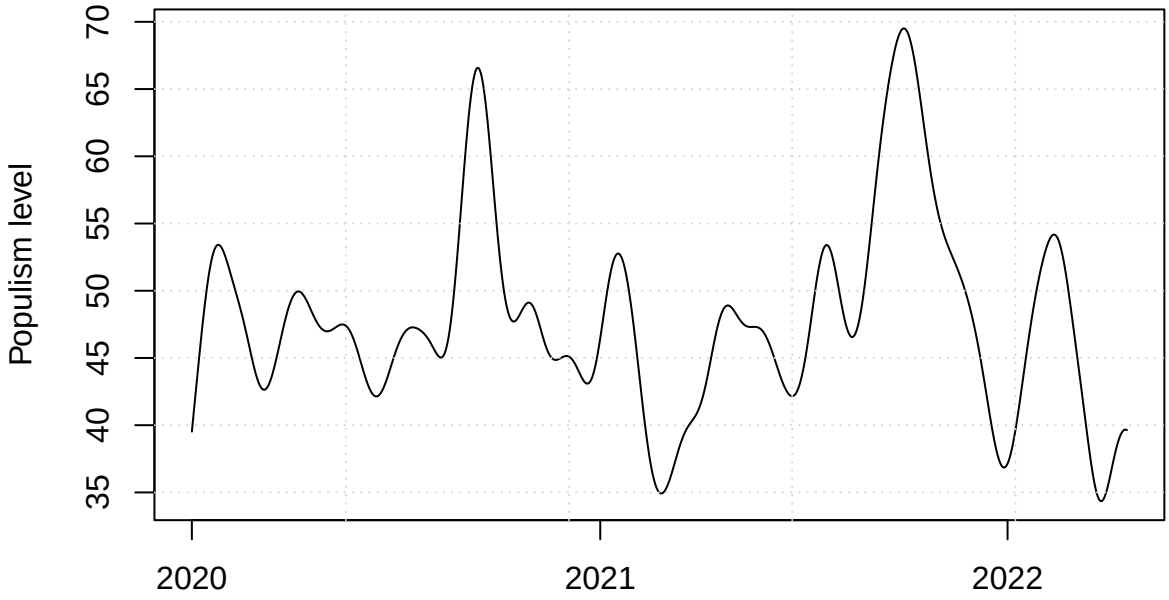
```
## Document-feature matrix of: 28 documents, 1 feature (0.00% sparse) and 1 docvar.
##     features
## docs populism
##    1     1541
##    2     1352
##    3     1381
##    4     1492
##    5     1424
##    6     1282
## [ reached max_ndoc ... 22 more documents ]
```

General level of populism in time

```
matplot(dfm_by_date5$date, dfm_by_date5, type = "l", lty = 1, col = 1:2,
        ylab = "Frequency", xlab = "")
grid()
legend("topleft", col = 1:2, legend = colnames(dfm_by_date5), lty = 1, bg = "white")
```




```
dat_smooth <- ksmooth(x = dfm_by_date5$date,
                      y = dfm_by_date5[,"populism"] - !dfm_by_date5[,"populism"],
                      kernel = "normal", bandwidth = 30)
plot(dat_smooth$x, dat_smooth$y, type = "l", ylab = "Populism level", xlab = "")
grid()
abline(h = 0, lty = 2)
```



Most populist party

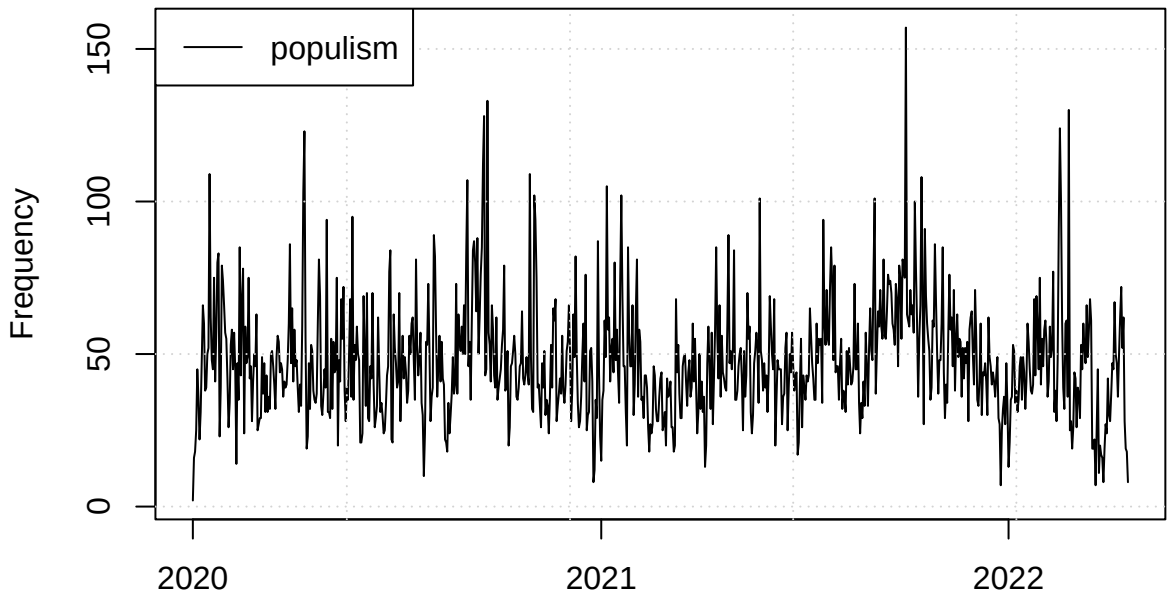
```
dfm_dict5_by_party <- dfm_group(dfm_dict5, groups = party_id)
dfm_dict5_by_party
```

```
## Document-feature matrix of: 11 documents, 1 feature (0.00% sparse) and 1 docvar.
##               features
## docs           populism
##   CI                700
##   FDI              4289
##   FI               6267
##   INDIPENDENTE      197
##   IV                251
##   LEGA             8024
## [ reached max_ndoc ... 5 more documents ]
```

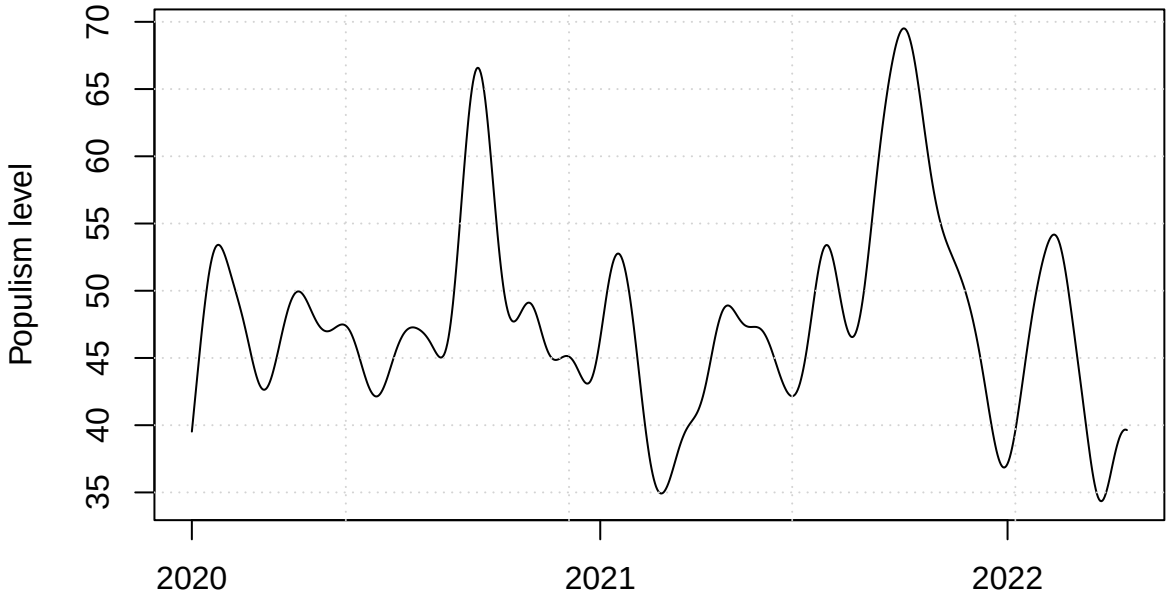
```
# Most populist party
dict_5_tstat_freq <- textstat_frequency(dfm_dict5, n = 5, groups = party_id)
kable(dict_5_tstat_freq)
```

feature	frequency	rank	docfreq	group
populism	700	1	645	CI
populism	4289	1	3833	FDI
populism	6267	1	5703	FI
populism	197	1	181	INDIPENDENTE
populism	251	1	233	IV
populism	8024	1	7353	LEGA
populism	929	1	820	LEU
populism	7224	1	6386	M5S
populism	3579	1	3219	MISTO
populism	8534	1	7744	PD
populism	120	1	109	REG_LEAGUES

```
matplot(dfm_by_date5$date, dfm_by_date5, type = "l", lty = 1, col = 1:2,
        ylab = "Frequency", xlab = "")
grid()
legend("topleft", col = 1:2, legend = colnames(dfm_by_date5), lty = 1, bg = "white")
```




```
dat_smooth <- ksmooth(x = dfm_by_date5$date,
                      y = dfm_by_date5[,"populism"] - !dfm_by_date5[,"populism"],
                      kernel = "normal", bandwidth = 30)
plot(dat_smooth$x, dat_smooth$y, type = "l", ylab = "Populism level", xlab = "")
grid()
abline(h = 0, lty = 2)
```


Populism level
35
40
45
50
55
60
65
70
2020
2021
2022