

Greek Prime Ministers programmatic statements analysis

Ανάλυση περιεχομένου των προγραμματικών δηλώσεων των πρωθυπουργών Ελλάδας (2009 – 2019) χρησιμοποιώντας τη γλώσσα R.

Για την παρούσα ανάλυση αξιοποιήθηκαν τα κείμενα των προγραμματικών δηλώσεων των ελλήνων πρωθυπουργών (Κ.Μητσοτάκης, Α.Τσίπρας, Α.Σαμαράς, Γ.Παπανδρέου) και στατιστικά πακέτα και αλγόριθμοι της γλώσσας προγραμματισμού R.

Στο **πρώτο μέρος** πραγματοποιήθηκε στατιστική επεξεργασία των προγραμματικών δηλώσεων όπου εξάγονται βασικά χαρακτηριστικά του κειμένου και της γλώσσας όπως αριθμός μοναδικών λέξεων, μεγαλύτερες λέξεις, ταχύτητα ανάγνωσης κτλ.

Στο **δεύτερο μέρος**, μέσω ληματοποίησης (tokenization) του κειμένου, δημιουργήθηκαν λίστες συχνότητας λέξεων, ανάλυση κανόνων συσχέτισης (term correlation) των συχνότερα εμφανιζόμενων λέξεων, επισημείωση μερών του λόγου και λεξικογραφική ανάλυση (annotation analysis). Τα βασικά αποτελέσματα συνοδεύονται από τις αντίστοιχες γραφικές παραστάσεις.

Τέλος, για κάθε κείμενο προγραμματικών δηλώσεων δημιουργήθηκε και το αντίστοιχο word cloud.

Β.Λαμπρακάκης, 2019

Το σύνολο του κώδικα και η αντίστοιχη τεκμηρίωση είναι προσβάσιμα στο αποθετήριο GitHub, στην ηλεκτρονική διεύθυνση https://github.com/vlamprakakis/Primeministers_Statements_TextMining

Στατιστικά κειμένων προγραμματικών δηλώσεων (1)

Αριθμός λέξεων			
Μητσοτάκης, 20-07-2019	Τσίπρας, 8-2-2015	Σαμαράς, 6-7-2012	Παπανδρέου, 16-10-2009
> length(words[[1]])	> length(words[[1]])	> length(words[[1]])	> length(words[[1]])
[1] 6382	[1] 8119	[1] 4813	[1] 5374
Αριθμός μοναδικών λέξεων			
Μητσοτάκης, 20-07-2019	Τσίπρας, 8-2-2015	Σαμαράς, 6-7-2012	Παπανδρέου, 16-10-2009
> length(unique(words[[1]]))	> length(unique(words[[1]]))	> length(unique(words[[1]]))	> length(unique(words[[1]]))
[1] 2166	[1] 2563	[1] 1634	[1] 1912
Αριθμός προτάσεων			
Μητσοτάκης, 20-07-2019	Τσίπρας, 8-2-2015	Σαμαράς, 6-7-2012	Παπανδρέου, 16-10-2009
> nsentence(text)	> nsentence(text)	> nsentence(text)	> nsentence(text)
text1	text1	text1	text1
316	345	240	346
Αριθμός χαρακτήρων κειμένου			
Μητσοτάκης, 20-07-2019	Τσίπρας, 8-2-2015	Σαμαράς, 6-7-2012	Παπανδρέου, 16-10-2009
> count_characters(text)	> count_characters(text)	> count_characters(text)	> count_characters(text)
[1] 43125	[1] 55587	[1] 31985	[1] 36477

Στατιστικά κειμένων προγραμματικών δηλώσεων (2)

Μεγαλύτερες λέξεις (άνω των 15 χαρακτήρων)

Μητσοτάκης, 20-07-2019	Τσίπρας, 8-2-2015	Σαμαράς, 6-7-2012	Παπανδρέου, 16-10-2009
> filter(tab, nchar(word) >15)	> filter(tab, nchar(word) >15)	> filter(tab, nchar(word) >15)	> filter(tab, nchar(word) >15)
# A tibble: 15 x 2	# A tibble: 21 x 2	# A tibble: 17 x 2	# A tibble: 16 x 2
word count	word count	word count	word count
<chr> <dbl>	<chr> <dbl>	<chr> <dbl>	<chr> <dbl>
1 αποτελεσματικότητα 2	1 διαπραγματευόμαστε 4	1 αποκρατικοποιήσεις 5	1 ανταγωνιστικότητα 2
2 ανταποκρινόμαστε 1	2 διαπραγματευτική 3	2 αποκρατικοποίηση 3	2 αναχρηματοδότησης 1
3 αντιλαμβανόμαστε 1	3 ανταγωνιστικότητα 2	3 αποφασιστικότητα 3	3 ανταγωνιστικότητας 1
4 αντιμετωπίζοντας 1	4 ανταποδοτικότητα 2	4 ανταγωνιστικότητα 2	4 αποποινικοποίηση 1
5 απελευθερώνοντας 1	5 αντισυνταγματικά 2	5 μεταρρυθμιστικές 2	5 αποτελεσματικότητας 1
6 αποκομματικοποίηση 1	6 αποφασιστικότητά 2	6 αναδιαπραγμάτευση 1	6 αστυνομικοκεντρική 1
7 αποτελεσματικότητας 1	7 διαπραγματεύσεις 2	7 αναδιαπραγμάτευσης 1	7 δραστηριοποιούμε 1
8 δημιουργικότητας 1	8 διαπραγματευτούμε 2	8 αποφασιστικότητά 1	8 επιχειρηματικότητα 1
9 εκσυγχρονίζονται 1	9 ραδιοτηλεοπτικών 2	9 αυτοαπασχολούμενους 1	9 κινηματογραφικός 1
10 εντατικοποιούνται 1	10 χρηματοπιστωτικής 2	10 εγκληματικότητας 1	10 λαθρομετανάστευσης 1
11 επαναπροκηρυχθεί 1	11 αδιαπραγμάτευτες 1	11 κοινοβουλευτικός 1	11 ξαναλειτουργήσει 1
12 ιδιωτικοποιήσουμε 1	12 αμφιταλαντεύσεις 1	12 πολλαπλασιάζονται 1	12 περιθωριοποιείται 1
13 κοινοβουλευτικού 1	13 ανακεφαλαιοποιήθηκαν 1	13 προγραμματισμένη 1	13 προσανατολισμένη 1
14 χρηματοπιστωτικής 1	14 αναπροσανατολισμό 1	14 σταθεροποιητικού 1	14 συνδιαμορφώσουμε 1
15 χρονοδιαγραμμάτων 1	15 αντισυνταγματικές 1	15 συμπαρασταθήκατε 1	15 συνειδητοποιημένο 1

Στατιστικά κειμένων προγραμματικών δηλώσεων (3)

Μέσο μήκος λέξης (αριθμός χαρακτήρων)

Μητσοτάκης, 20-07-2019	Τσίπρας, 8-2-2015	Σαμαράς, 6-7-2012	Παπανδρέου, 16-10-2009
> lapply(word_lengths, mean)	> lapply(word_lengths, mean)	> lapply(word_lengths, mean)	> lapply(word_lengths, mean)
[[1]]	[[1]]	[[1]]	[[1]]
[1] 5.599499	[1] 5.708708	[1] 5.515271	[1] 5.602903

Εκτιμώμενος χρόνος ανάγνωσης (σε λεπτά)

Μητσοτάκης, 20-07-2019	Τσίπρας, 8-2-2015	Σαμαράς, 6-7-2012	Παπανδρέου, 16-10-2009
> read_time_minutes	> minutes	> minutes	> minutes
[1] 39.49	[1] 40.59	[1] 24.06	[1] 26.87

Πραγματικός χρόνος ανάγνωσης σε λεπτά (από YouTube)

Μητσοτάκης, 20-07-2019	Τσίπρας, 8-2-2015	Σαμαράς, 6-7-2012	Παπανδρέου, 16-10-2009
61.14	87.12	38.10	48.05

Εμφάνιση συχνότητας λέξεων (κάτω από τη λέξη εμφανίζεται ο αριθμός συχνότητας εμφάνισης στο κείμενο των προγραμματικών δηλώσεων)

Μητσοτάκης

> freq

έχω	ανέβει	βήμα	εκατοντάδες
5	1	3	1
ευχαριστώ	ιδιαίτερη	κρύβω	κυρίες
2	1	1	9
κύριοι	σήμερα	σας	στιγμή
9	14	6	2

Τσίπρας

> freq

ένωσης	ανάκτηση	αναγέννηση	ανθρωπιστικής	αντιμετώπιση
4	2	1	2	5
αξιοπρέπειας	αποκατάσταση	αποτελούν	βασικούς	βουλευτές
5	6	6	1	19
δικαιοσύνη	επιλογή	ευρωπαϊκής	θεσμικό	ιανουαρίου
4	5	7	3	1

Σαμαράς

> freq

ακόμη	αυτές	γενικολογίες	ευκολότερες	εύκολες
9	5	1	1	1
θα	καλύπτουν	κυρίες	κύριοι	λένε
94	1	12	12	1
μιλήσω	πάντα	περιπτώσεις	περιστροφές	σας
3	5	4	1	2

Παπανδρέου

> freq

ιδίας	απαράλλαχτης	γυρίζει	γυρίσουμε	δέσμευσε
2	1	1	1	1
διαχειριστές	είμαστε	ελλάδα	ελληνικός	εμπιστευτήκε
1	6	17	3	1
εξουσίας	ζήτησε	ιστορίας	κάνουμε	κρίσιμες
2	1	2	12	1

Πόσες φορές εμφανίζεται η κάθε συχνότητα

(1η γραμμή -> συχνότητα, 2η γραμμή -> αριθμός εμφάνισης συχνότητας)

Μητσοτάκης

```
> head(table(freq), 20)
```

freq

1	2	3	4	5	6	7	8	9	10	11	12	13	14	20	25	81	129
1562	284	124	42	19	18	12	14	12	3	4	4	3	1	2	1	1	1

Τσίπρας

```
> head(table(freq), 20)
```

freq

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	21
1704	389	149	80	51	31	28	21	15	3	6	5	4	2	1	1	2	4	2	2

Σαμαράς

```
> head(table(freq), 20)
```

freq

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	17	18	19	35	54
1094	257	87	48	32	18	12	10	4	2	1	8	2	3	1	1	1	1	1	1

Παπανδρέου

```
> head(table(freq), 20)
```

freq

1	2	3	4	5	6	7	8	9	10	11	12	14	16	17	81	95
1328	275	104	51	32	17	16	9	8	6	1	5	1	1	1	1	1

Οι 20 πιο συχνά εμφανιζόμενες λέξεις και η συχνότητα εμφάνισης τους

Μητσοτάκης

```
> head(freq, 20)
```

θα	μας	ελλάδα	μόνο	όμως	σήμερα	κυβέρνηση	νέα	πολιτική	επιτέλους
129	81	25	20	20	14	13	13	13	12
μπορεί	κάθε	μπορούμε	γιατί	ανάπτυξη	όλα	μέσα	θέλω	ας	υπουργείο
12	12	12	11	11	11	11	10	10	10

Τσίπρας

```
> head(freq, 20)
```

θα	μας	κυβέρνηση	όμως	μόνο	λαού	νέα	βουλευτές	πλαίσιο
105	102	26	23	22	21	21	19	19
κυρίες	κύριοι	χώρας	χωρίς	κάθε	ελληνική	ελλάδα	ελληνικού	γιατί
18	18	18	18	17	17	16	15	14
έχει	κοινωνικής							
14	13							

Σαμαράς

```
> head(freq, 20)
```

θα	μας	πρέπει	μπορεί	ώστε	έχει	γίνει	ελλάδα	όλα
94	54	35	19	18	17	15	14	14
κάθε	αν	σήμερα	κυρίες	κύριοι	συνάδελφοι	μόνο	όχι	χώρα
14	13	13	12	12	12	12	12	12
κάνουμε	ήδη							
12	12							

Παπανδρέου

```
> head(freq, 20)
```

μας	θα	ελλάδα	απαξίωση	κάθε	κάνουμε	μαζί	έχει	χώρας
95	81	17	16	14	12	12	12	12
όλων	όμως	κυρίες	κύριοι	συνάδελφοι	χώρα	έχουμε	δημόσιο	πολίτη
12	11	10	10	10	10	10	10	9
κατάσταση	σήμερα							
9	9							

Λέξεις με συχνότητα εμφάνισης τουλάχιστον 10 φορές

Μητσοτάκης

```
> findFreqTerms(dtm, lowfreq=10)
```

```
[1] "σήμερα" "γιατί" "ελλάδα" "επιτέλους" "κυβέρνηση" "μας" "μόνο" "νέα"  
[9] "πολιτική" "όμως" "θα" "ανάπτυξη" "μπορεί" "κάθε" "όλα" "θέλω"  
[17] "μπορούμε" "ας" "μέσα" "υπουργείο"
```

Τσίπρας

```
> findFreqTerms(dtm, lowfreq=10)
```

```
[1] "βουλευτές" "κοινωνικής" "κυβέρνησης" "κυρίες" "κύριοι" "λαού" "μας"  
[8] "πλαίσιο" "χώρας" "χρόνια" "χώρα" "όμως" "ελλάδα" "θα"  
[15] "έχουμε" "γιατί" "άμεσα" "χωρίς" "κάθε" "ελληνικού" "κυβέρνηση"  
[22] "ευρώπη" "όχι" "ελληνική" "δημόσιο" "μόνο" "νέα" "ώστε"  
[29] "είμαστε" "πρέπει" "έχει" "τέλος" "εμείς" "μάχη" "συστήματος"  
[36] "ρόλο" "εργασίας"
```

Σαμαράς

```
> findFreqTerms(dtm, lowfreq=10)
```

```
[1] "θα" "κυρίες" "κύριοι" "συνάδελφοι" "έχει" "αν" "μπορεί"  
[8] "μόνο" "σήμερα" "όχι" "ελλάδα" "χώρα" "πρέπει" "γιατί"  
[15] "μας" "όλα" "ώστε" "κάθε" "πιο" "γίνει" "κάνουμε"  
[22] "ήδη" "μέτρα"
```

Παπανδρέου

```
> findFreqTerms(dtm, lowfreq=10)
```

```
[1] "ελλάδα" "κάνουμε" "κυρίες" "κύριοι" "μας" "συνάδελφοι" "χώρα"  
[8] "κάθε" "έχουμε" "θα" "μαζί" "έχει" "χώρας" "όμως"  
[15] "απαξίωση" "όλων" "δημόσιο"
```


ΣΥΣΧΕΤΙΣΕΙΣ ΜΕΤΑΞΥ ΟΡΩΝ

Υπολογίζεται ο βαθμός συσχέτισης της λέξης αναζήτησης με κάθε άλλη λέξη του κειμένου.

Οι τιμές του συντελεστή κυμαίνονται από 0 έως 1. Η τιμή 1 σημαίνει ότι οι δύο λέξεις πάντα εμφανίζονται μαζί στο κείμενο ενώ οι τιμές που προσεγγίζουν το 0 σημαίνουν ότι οι λέξεις έχουν χαμηλό αριθμό εμφάνισης συγκριτικά με τη λέξη αναζήτησης.

Θεωρητικά, τιμές του συντελεστή από 0.3 - 0.5 καταδεικνύουν μέτρια συσχέτιση και από τιμές 0.5 -1 ισχυρή συσχέτιση.

Μητσοτάκης

Συσχέτιση όρων με τον όρο "θα" και συντελεστή συσχέτισης ≥ 0.45

```
> # find correlation of term "θα" with other terms
```

```
> findAssocs(tdm, c("θα"), corlimit = 0.45)
```

\$θα

κατάσταση	χρειαστεί	ψηφιακή	ασφαλιστική	αυτόματα	βαθμίδα	γέννηση
0.52	0.52	0.50	0.48	0.48	0.48	0.48
γνησιότητα	δηλώνεται	εάν	εγγράφεται	εισάγεται	ενημερώνεται	εξετάζεται
0.48	0.48	0.48	0.48	0.48	0.48	0.48
εξυπηρετείται	επικοινωνίας	επικοινωνεί	εργαλείο	ηλεκτρονικά	νοσοκομείο	παιδί
0.48	0.48	0.48	0.48	0.48	0.48	0.48
παιδιών	προχωρά	συνολικά	τίτλων	υπόσταση		
0.48	0.48	0.48	0.48	0.48		

Μητσοτάκης

Συσχέτιση όρων με τον όρο "ελλάδα" και συντελεστή συσχέτισης ≥ 0.7

```
> # find correlation of term "ελλάδα" with other terms
```

```
> findAssocs(tdm, c("ελλάδα"), corlimit = 0.7)
```

\$ελλάδα

2004	έκανε	αγώνες	βρίσκει	θεωρώ	ιστορίας	μοναδική	ολυμπιακούς
0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76
ορόσημο	ραγδαία	σύγχρονης	υπερήφανους	χρονιά			
0.76	0.76	0.76	0.76	0.76			

ΣΥΣΧΕΤΙΣΕΙΣ ΜΕΤΑΞΥ ΟΡΩΝ

Μητσοτάκης

Συσχέτιση όρων με τον όρο "κυβέρνηση" και συντελεστή συσχέτισης ≥ 0.5

```
> # find correlation of term "κυβέρνηση" with other terms
```

```
> findAssocs(tdm, c("κυβέρνηση"), corlimit = 0.5)
```

```
$κυβέρνηση
```

έγινε	αναστολές	γεφυρώνει	λύσεις	ξέρει	παρωπίδες	πείρα	σύνθεσή	τεχνοκράτη
0.52	0.52	0.52	0.52	0.52	0.52	0.52	0.52	0.52

Μητσοτάκης

Συσχέτιση όρων με τον όρο "ανάπτυξη" και συντελεστή συσχέτισης ≥ 0.5

```
> # find correlation of term "ανάπτυξη" with other terms
```

```
> findAssocs(tdm, c("ανάπτυξη"), corlimit = 0.5)
```

```
$ανάπτυξη
```

πρέπει	επενδύσεις	υγεία	μεταρρυθμιστική	τομή
0.57	0.52	0.51	0.51	0.51

Μητσοτάκης

Συσχέτιση όρων με τον όρο "χώρα" και συντελεστή συσχέτισης ≥ 0.4

```
> # find correlation of term "χώρα" with other terms
```

```
> findAssocs(tdm, c("χώρα"), corlimit = 0.4)
```

```
$χώρα
```

εθνικές	συνδυάζει	αγροτική	γίνει	εκσυγχρονισμό
0.49	0.46	0.46	0.45	0.44

ΣΥΣΧΕΤΙΣΕΙΣ ΜΕΤΑΞΥ ΟΡΩΝ

Τσίπρας

Συσχέτιση όρων με τον όρο "θα" και συντελεστή συσχέτισης ≥ 0.33

```
> # find correlation of term "θα" with other terms
> findAssocs(tdm, c("θα"), corlimit = 0.33)
$θα
```

συμφωνία	πρόγραμμα	επιδιώκουμε	κίνητρα	εταίροι
0.38	0.37	0.37	0.37	0.33

Τσίπρας

Συσχέτιση όρων με τον όρο "κυβέρνηση" και συντελεστή συσχέτισης ≥ 0.5

```
> # find correlation of term "κυβέρνηση" with other terms
> findAssocs(tdm, c("κυβέρνηση"), corlimit = 0.5)
$κυβέρνηση
```

νέα	απερχόμενη	αποδέχθηκε	βαθύ	βράχια	γέφυρες	γκρεμίζει
0.78	0.52	0.52	0.52	0.52	0.52	0.52
δημιουργήσει	δυστύχημα	εξαναγκάσει	κάψει	μείγματος	ξαναμπί	οδηγούσε
0.52	0.52	0.52	0.52	0.52	0.52	0.52
περιπέτειες	προηγούμενη	σπαρτά	φεύγοντας	φρόντισε		
0.52	0.52	0.52	0.52	0.52		

Τσίπρας

Συσχέτιση όρων με τον όρο "λαού" και συντελεστή συσχέτισης ≥ 0.39

```
> # find correlation of term "λαού" with other terms
> findAssocs(tdm, c("λαού"), corlimit = 0.39)
$λαού
```

κυρίαρχη	ελληνικού	ανάκτηση	τέλους
0.54	0.44	0.40	0.40

ΣΥΣΧΕΤΙΣΕΙΣ ΜΕΤΑΞΥ ΟΡΩΝ

Τσίπρας

Συσχέτιση όρων με τον όρο "ελλάδα" και συντελεστή συσχέτισης ≥ 0.4

```
> # find correlation of term "ελλάδα" with other terms
```

```
> findAssocs(tdm, c("ελλάδα"), corlimit = 0.4)
```

\$ελλάδα

αλλάζει	ατζέντα	διαφορετικό	δύσκολο	ενδιαφέροντος	επίκεντρο	θετικό
0.45	0.45	0.45	0.45	0.45	0.45	0.45
κομπάρσος	πρωταγωνιστής	τοπίο	εξελιξεν	επίμονη	επωφελή	θετικός
0.45	0.45	0.45	0.45	0.45	0.45	0.45
καταλήξουμε	καταλύτης					
0.45	0.45					

Τσίπρας

Συσχέτιση όρων με τον όρο "μάχη" και συντελεστή συσχέτισης ≥ 0.5

```
> # find correlation of term "μάχη" with other terms
```

```
> findAssocs(tdm, c("μάχη"), corlimit = 0.5)
```

\$μάχη

πραγματικά	ενάντια	έτοιμη	έφτασε	αγώνας	αδυσώπητη
0.88	0.86	0.72	0.72	0.72	0.72
απαξίωσε	γκρεμού	δηλητηρίασε	θεσμούς	καταδυναστεύει	κόστος
0.72	0.72	0.72	0.72	0.72	0.72
μεγάλος	φοροαποφυγή	φοροδιαφυγή	χείλος	σύστημα	αλλάξουμε
0.72	0.72	0.72	0.72	0.57	0.54
διοικήσαμε	μόνοι	ξεκάθαρη	οικοδομήσαμε	οικοδόμησαν	πελατελικών
0.54	0.54	0.54	0.54	0.54	0.54
περιουτολογώ	σχέσεων	σύγχρονη	αποτέλεσε	διαφθορά	αιτία
0.54	0.54	0.54	0.50	0.50	0.50
δώσει	ζωή	πραγματική	υψηλή		
0.50	0.50	0.50	0.50		

ΣΥΣΧΕΤΙΣΕΙΣ ΜΕΤΑΞΥ ΟΡΩΝ

Σαμαράς

Συσχέτιση όρων με τον όρο "θα" και συντελεστή συσχέτισης ≥ 0.50

```
> # find correlation of term "θα" with other terms
```

```
> findAssocs(tdm, c("θα"), corlimit = 0.50)
```

\$θα

έσοδα	ξεκινήσει	φυσικά	έλεγα
0.53	0.53	0.53	0.52
αισθητικής	ακίνητης	αναβαθμιστεί	αναξιοποίητα
0.52	0.52	0.52	0.52
ανενεργά	ανεξάρτητα	αντίστοιχη	αξία
0.52	0.52	0.52	0.52
αποκτήσουν	αποτελέσει	αποφύγουμε	γειτονικές
0.52	0.52	0.52	0.52
δημιουργήσουν	δυνατόν	δόμησης	γης
0.52	0.52	0.52	0.52
επενδυτικό	επικράτεια	καθαρίσουν	καταπατητή
0.52	0.52	0.52	0.52
κυριότητας	κόβονται	κόσμου	ληλατούνται
0.52	0.52	0.52	0.52
μένουν	μέρη	μεγάλα	μεγάλης
0.52	0.52	0.52	0.52
νομοθετικές	οποίου	πέραν	παραμένει
0.52	0.52	0.52	0.52
περιμένουμε	περιμένουν	περιουσία	περιουσιακά
0.52	0.52	0.52	0.52
πιλότο	προσδιοριστούν	προχωράει	προχωρήσει
0.52	0.52	0.52	0.52
σούνιο	στοιχεία	συντελεστές	σχέδιο
0.52	0.52	0.52	0.52
τεράστια	υπεραξίες	υπερεκμετάλλευσης	υποβάθμισης
0.52	0.52	0.52	0.52
χρήσεις	όχημα	ώθηση	υπουργοί
0.52	0.52	0.52	0.51

ΣΥΣΧΕΤΙΣΕΙΣ ΜΕΤΑΞΥ ΟΡΩΝ

Σαμαράς

Συσχέτιση όρων με τον όρο "πρέπει" και συντελεστή συσχέτισης ≥ 0.50

```
> # find correlation of term "πρέπει" with other terms
```

```
> findAssocs(tdm, c("πρέπει"), corlimit = 0.5)
```

\$πρέπει

αξιοπιστία	βγει	διάστημα	δούμε	ηγεσία	κάποιος	κερδίσουμε	κόσμο	ξανά
0.60	0.60	0.60	0.60	0.60	0.60	0.60	0.60	0.60
πέρασε	πούμε	προφανές	στήριγμά	ταλαιπωρία	τελευταίο	τολήσουμε	χαράξουμε	ακόμη
0.60	0.60	0.60	0.60	0.60	0.60	0.60	0.60	0.56
πέρα								
0.52								

Σαμαράς

Συσχέτιση όρων με τον όρο "ελλάδα" και συντελεστή συσχέτισης ≥ 0.40

```
> # find correlation of term "ελλάδα" with other terms
```

```
> findAssocs(tdm, c("ελλάδα"), corlimit = 0.4)
```

\$ελλάδα

έτοιμοι	αποφασισμένη	αυτοπεποίθηση	βγάλουν	δείξει	δυο
0.47	0.47	0.47	0.47	0.47	0.47
εμπνεύσει	λόγια	μάχες	πετύχει	αδύναμος	ανακτήσουμε
0.47	0.47	0.47	0.47	0.47	0.47
ανατολή	ευρώπης	θυμίσουμε	κοντεύει	κρίκος	μέση
0.47	0.47	0.47	0.47	0.47	0.47
ξανακάνουμε	προπύργιο	ρόλο	σταθεροποιητικό	σύννορά	σύννορα
0.47	0.47	0.47	0.47	0.47	0.47

ΣΥΣΧΕΤΙΣΕΙΣ ΜΕΤΑΞΥ ΟΡΩΝ

Σαμαράς

Συσχέτιση όρων με τον όρο "μέτρα" και συντελεστή συσχέτισης ≥ 0.45

```
> # find correlation of term "μέτρα" with other terms
```

```
> findAssocs(tdm, c("μέτρα"), corlimit = 0.45)
```

\$μέτρα

πέσει	ξεπεράσουμε	παίρνουμε	πρόσθετα	αν
0.56	0.51	0.51	0.51	0.50
2013	έβδομο	αίμα	αποκατασταθεί	γενικότερη
0.49	0.49	0.49	0.49	0.49
διευκολύνουν	ελληνικές	εντάσσεται	επιχείρηση	ηλεκτρονικού
0.49	0.49	0.49	0.49	0.49
θέμα	κατοίκους	κενό	κυρίαρχο	μικρή
0.49	0.49	0.49	0.49	0.49
οποιαδήποτε	περιουσιολογίου	προχωράμε	ρευστότητας	στήριξη
0.49	0.49	0.49	0.49	0.49
τράπεζες	φορολογούμενους	άκρο	αλλιώτικα	αναπτυξιακή
0.49	0.49	0.49	0.49	0.49
ας	γραφειοκράτες	δημόσιοι	εμπόδιο	εξοχήν
0.49	0.49	0.49	0.49	0.49
εχθρό	κακοί		καταδικασμένη	κατανικάς
0.49	0.49		0.49	0.49
καταφέρεις	νοσηρή	ξεριζώσεις	ολοκληρωμένη	παρακάμψεις
0.49	0.49	0.49	0.49	0.49
προσωπικά	στρατηγική	υπάλληλοι	χρειάζονται	προσέξτε
0.49	0.49	0.49	0.49	0.49
αδικίες	ακραίες	αναβληθούν	αφορμή	δείχνουμε
0.49	0.49	0.49	0.49	0.49
δημοσιονομικά	δημοσιονομικό	εκείνο	εννοούμε	επανορθωθούν
0.49	0.49	0.49	0.49	0.49
ζητούμε	ισοδύναμα	λόγω	παραπάνω	προσαρμογή
0.49	0.49	0.49	0.49	0.49
προτείνουμε	συνέβαινε	τονίσει	υφειακά	χαμηλοσυνταξιούχους
0.49	0.49	0.49	0.49	0.49
γραφειοκρατία				
0.48				

ΣΥΣΧΕΤΙΣΕΙΣ ΜΕΤΑΞΥ ΟΡΩΝ

Παπανδρέου

Συσχέτιση όρων με τον όρο "θα" και συντελεστή συσχέτισης ≥ 0.25

```
> # find correlation of term "θα" with other terms
```

```
> findAssocs(tdm, c("θα"), corlimit = 0.25)
```

\$θα

φορολογούνται	επενδύσεις	ίσων	αδιέξοδο	βγάλουμε	δίλημμα
0.29	0.27	0.25	0.25	0.25	0.25
δημιουργήσουμε	εάν	ευκαιριών	θετικά	ισότητας	πραγματικό
0.25	0.25	0.25	0.25	0.25	0.25
σημερινό	συγκριτικό	συνεχίσουμε	χτίσουμε	αχανείς	γερό
0.25	0.25	0.25	0.25	0.25	0.25
εκλογικές	εκλογικού	επιφέροντας	μαύρο	μειώνει	πολιτικό
0.25	0.25	0.25	0.25	0.25	0.25
προχωρήσουμε	σπάει	χρήμα	χτύπημα	διανέμονται	ευνοούν
0.25	0.25	0.25	0.25	0.25	0.25
κέρδη	καταργηθούν	μεγαλομετόχους	συντελεστή	χαμηλό	δεσμεύομαι
0.25	0.25	0.25	0.25	0.25	0.25
δώσω	εαυτό	καλύτερό	τιμήσω	ψυχή	όλη
0.25	0.25	0.25	0.25	0.25	0.25

Παπανδρέου

Συσχέτιση όρων με τον όρο "ελλάδα" και συντελεστή συσχέτισης ≥ 0.40

```
> # find correlation of term "κυβέρνηση" with other terms
```

```
> findAssocs(tdm, c("ελλάδα"), corlimit = 0.4)
```

\$ελλάδα

αξιών	κερδίζει	ανθρωπίνων	δημοκρατικής	διαλόγου	ειρήνης	εμπέδωσης
0.48	0.46	0.43	0.43	0.43	0.43	0.43
ευνομούμενης	ευρώπης	οικουμένης	παρουσίας	προάσπισης	προώθησης	πρωτοβουλίας
0.43	0.43	0.43	0.43	0.43	0.43	0.43
αντιπροσώπων	ελλήνων	ενώπιον	κατακτήσουμε	ξεκινήσαμε	ομιλία	πλευρά
0.43	0.43	0.43	0.43	0.43	0.43	0.43

ΣΥΣΧΕΤΙΣΕΙΣ ΜΕΤΑΞΥ ΟΡΩΝ

Παπανδρέου

Συσχέτιση όρων με τον όρο "απαξίωση" και συντελεστή συσχέτισης ≥ 0.50

```
> # find correlation of term "πολιτική" with other terms
```

```
> findAssocs(tdm, c("απαξίωση"), corlimit = 0.5)
```

\$απαξίωση

έννοιας	αδύναμο	αναξιοκρατία	βυθίζεται	εισοδήματος	ηθικής	θεσμών
0.84	0.84	0.84	0.84	0.84	0.84	0.84
κερδοσκοπία	οικογένειας	περιβάλλοντος	πολιτεία	πολυδιάστατη	προστασίας	συνταξιούχο
0.84	0.84	0.84	0.84	0.84	0.84	0.84
σωσίβιο	φτωχό	πάντων	άτομο	έφερε	διακυβέρνησης	κοινωνίας
0.84	0.84	0.70	0.58	0.58	0.58	0.58
παιδείας	έλλειψη					
0.58	0.57					

Παπανδρέου

Συσχέτιση όρων με τον όρο "δημόσιο" και συντελεστή συσχέτισης ≥ 0.35

```
> # find correlation of term "πολιτική" with other terms
```

```
> findAssocs(tdm, c("δημόσιο"), corlimit = 0.35)
```

\$δημόσιο

αγρότης	επιχειρηματίας	νέος	μην	χρέος	πρέπει
0.44	0.44	0.44	0.44	0.37	0.35

Παπανδρέου

Συσχέτιση όρων με τον όρο "ανάπτυξη" και συντελεστή συσχέτισης ≥ 0.40

```
> # find correlation of term "ανάπτυξη" with other terms
```

```
> findAssocs(tdm, c("ανάπτυξη"), corlimit = 0.4)
```

\$ανάπτυξη

πράσινη	γεωργίας	κατεύθυνση	υπαίθρου	αποτελέσει
0.86	0.60	0.60	0.42	0.42

ΛΕΞΙΚΟΓΡΑΦΙΚΗ ΑΝΑΛΥΣΗ (ANNOTATION ANALYSIS)

Περιλαμβάνει την ανάλυση στα μέρη του λόγου του κειμένου των προγραμματικών δηλώσεων και τη συχνότητα των αντίστοιχων όρων ανά κατηγορία.

ΣΥΜΒΟΛΟ	ΕΡΜΗΝΕΙΑ	ΕΡΜΗΝΕΙΑ (ελληνικά)
ADJ	adjective	επίθετο
ADP	adposition	προσθήκη
ADV	adverb	επίρρημα
AUX	auxiliary	βοηθητικό
CCONJ	coordinating conjunction	συνδετικό
DET	determiner	προσδιοριστικό
INTJ	interjection	επιφώνημα
NOUN	noun	ουσιαστικό
NUM	numeral	αριθμητικό
PART	particle	σύνθετη λέξη
PRON	pronoun	αντωνυμία
PROPN	proper noun	κύριο όνομα
PUNCT	punctuation	σημείο στίξης
SCONJ	subordinating conjunction	συνδετικό
SYM	symbol	σύμβολο
VERB	verb	ρήμα
X	other	άλλο

ΛΕΞΙΚΟΓΡΑΦΙΚΗ ΑΝΑΛΥΣΗ (ANNOTATION ANALYSIS)

Μητσοτάκης

```
> table(x$upos)
```

ADJ	ADP	ADV	AUX	CCONJ	DET	NOUN	NUM	PART	PRON	PROPN	PUNCT	SCONJ	VERB	X
619	506	354	365	350	1101	1431	96	49	388	90	792	57	664	10

Τσίπρας

```
> table(x$upos)
```

ADJ	ADP	ADV	AUX	CCONJ	DET	NOUN	NUM	PART	PRON	PROPN	PUNCT	SCONJ	VERB	X
925	666	392	412	470	1623	2185	73	96	532	77	896	77	760	17

Σαμαράς

```
> table(x$upos)
```

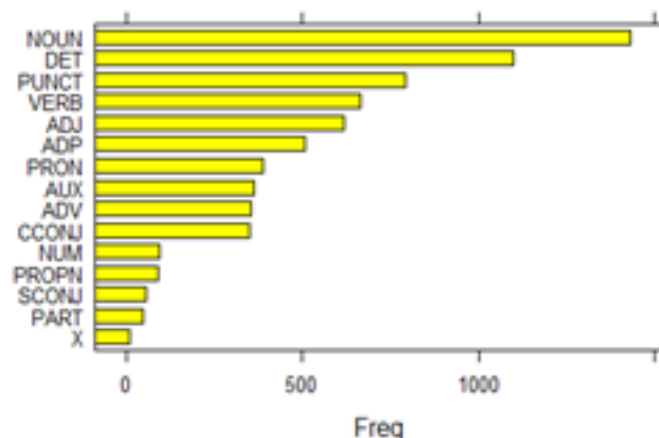
ADJ	ADP	ADV	AUX	CCONJ	DET	NOUN	NUM	PART	PRON	PROPN	PUNCT	SCONJ	VERB	X
432	377	347	405	244	726	1103	37	62	397	41	544	80	603	6

Παπανδρέου

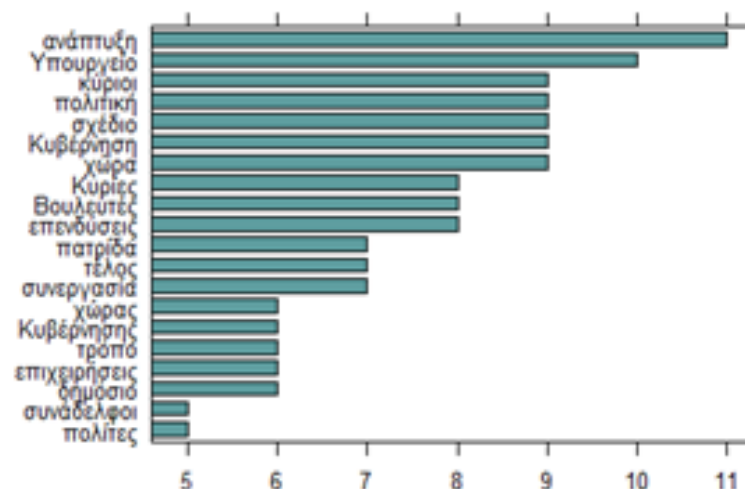
```
> table(x$upos)
```

ADJ	ADP	ADV	AUX	CCONJ	DET	NOUN	NUM	PART	PRON	PROPN	PUNCT	SCONJ	VERB	X
592	524	240	271	301	1094	1504	56	34	343	60	798	44	466	11

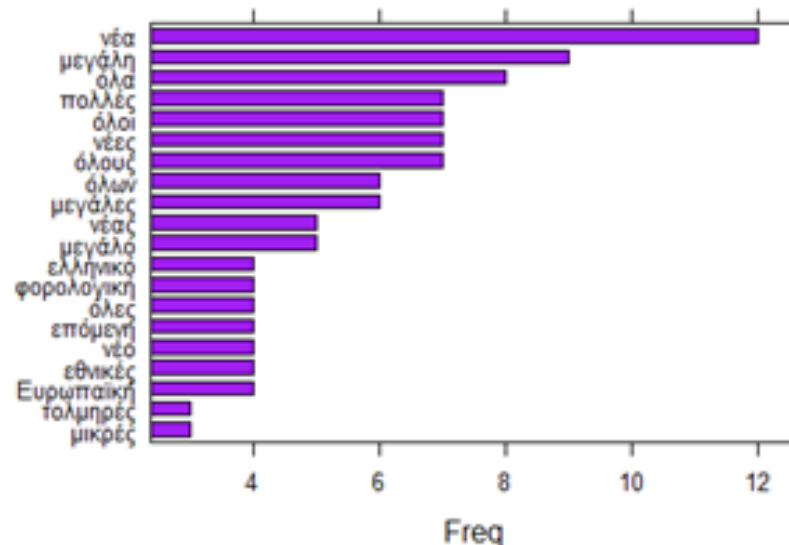
UPOS (Universal Parts of Speech)
frequency of occurrence



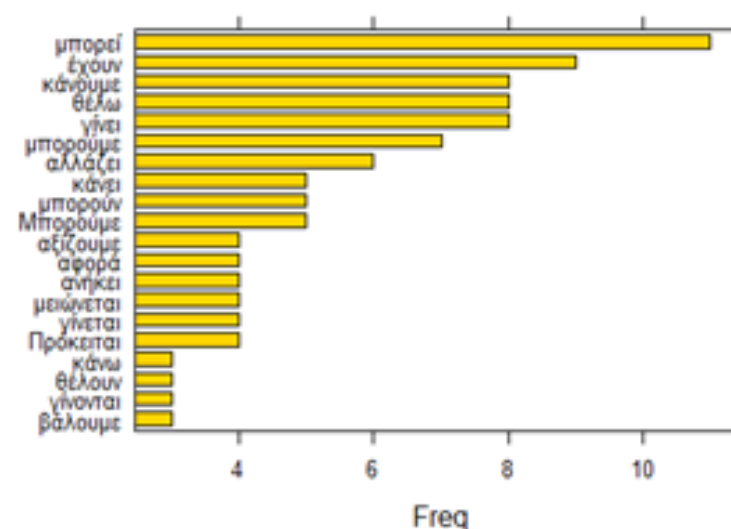
Most occurring nouns



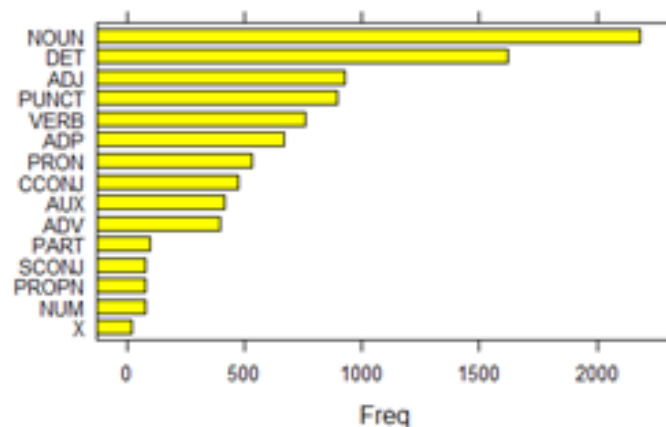
Most occurring adjectives



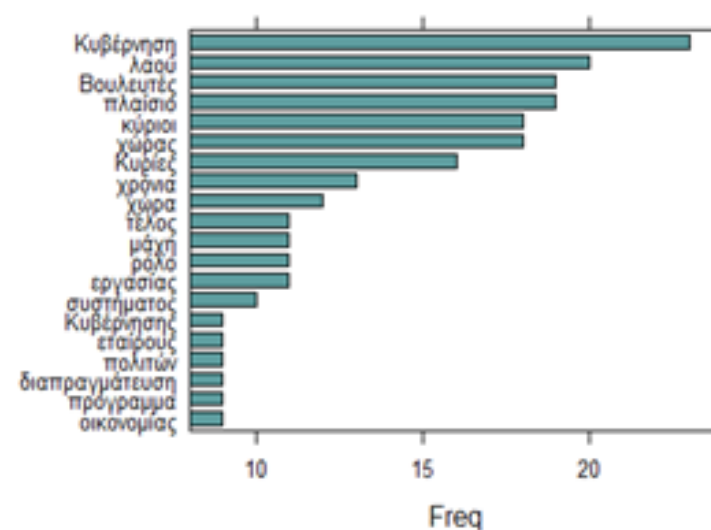
Most occurring Verbs



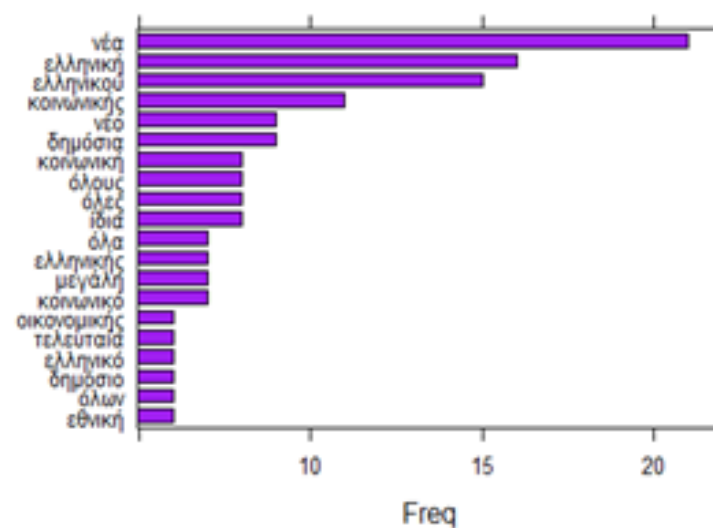
UPOS (Universal Parts of Speech)
frequency of occurrence



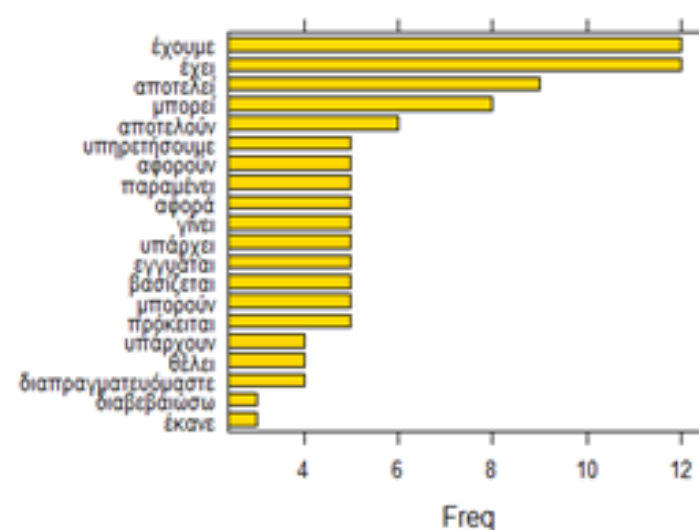
Most occurring nouns



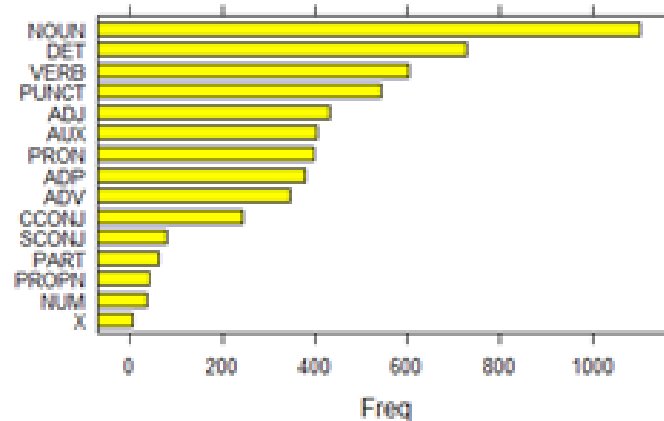
Most occurring adjectives



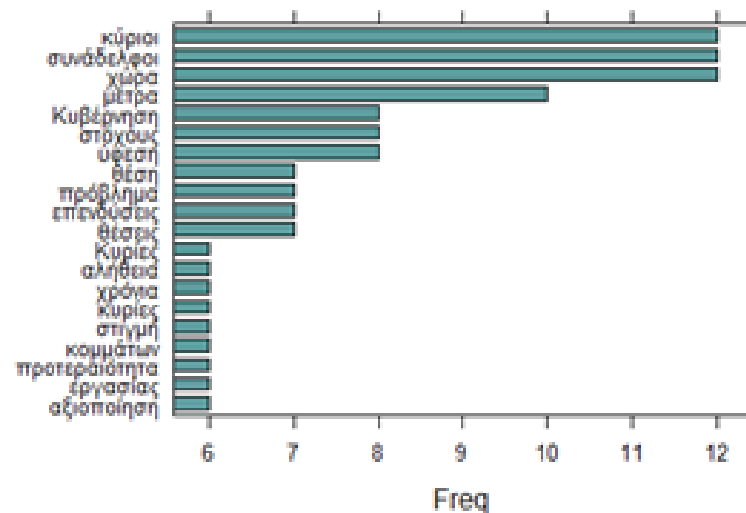
Most occurring Verbs



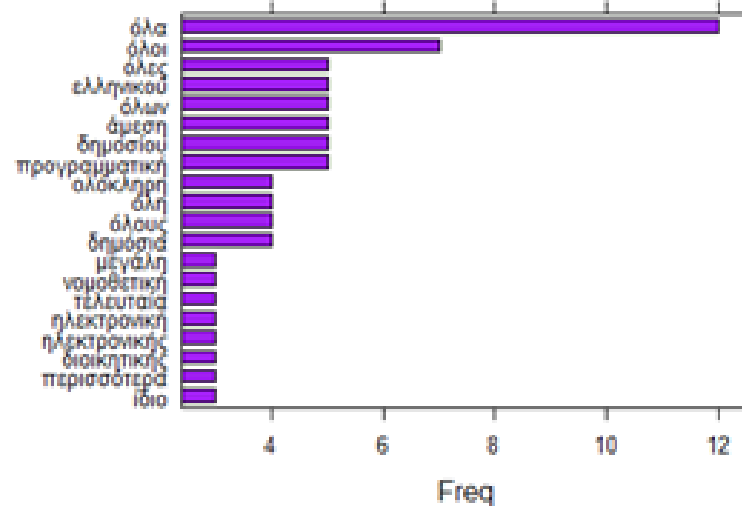
UPOS (Universal Parts of Speech)
frequency of occurrence



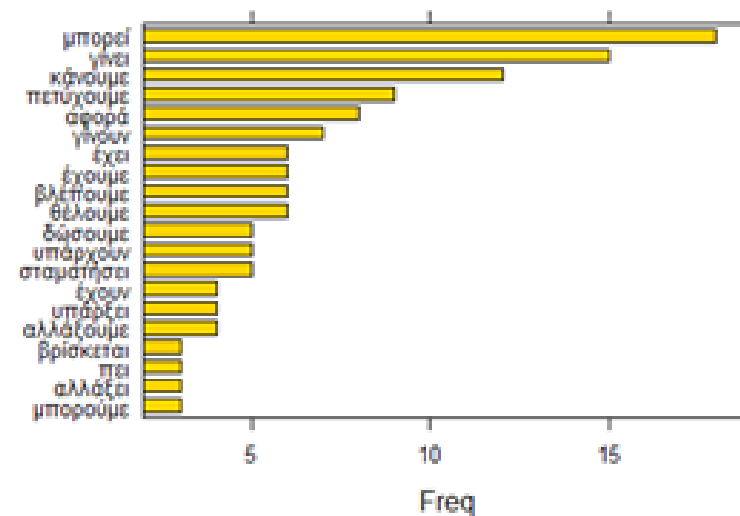
Most occurring nouns



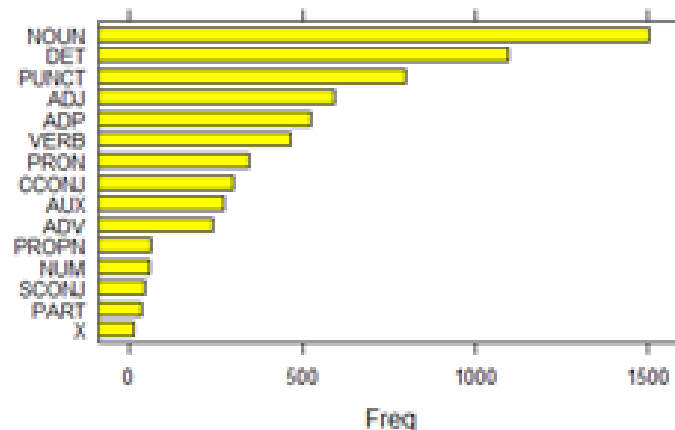
Most occurring adjectives



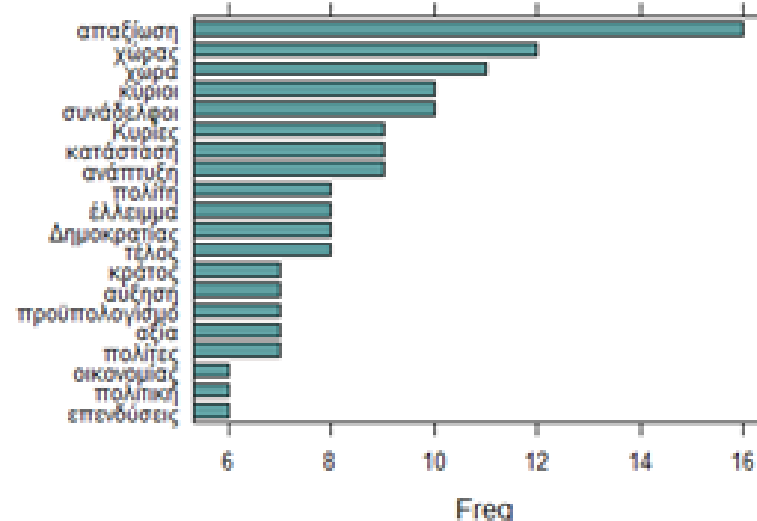
Most occurring Verbs



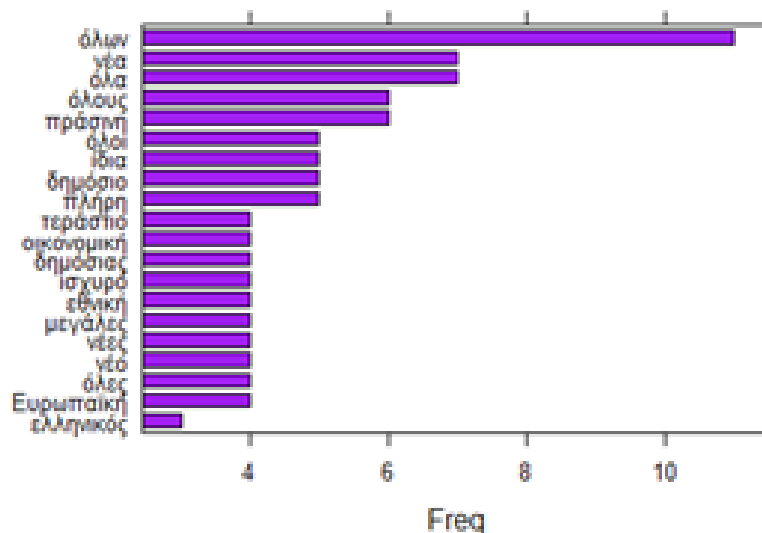
UPOS (Universal Parts of Speech)
frequency of occurrence



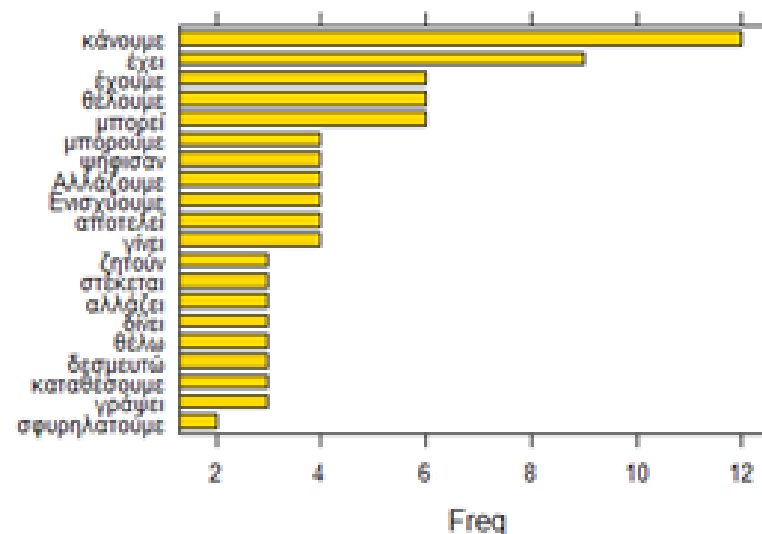
Most occurring nouns



Most occurring adjectives



Most occurring Verbs



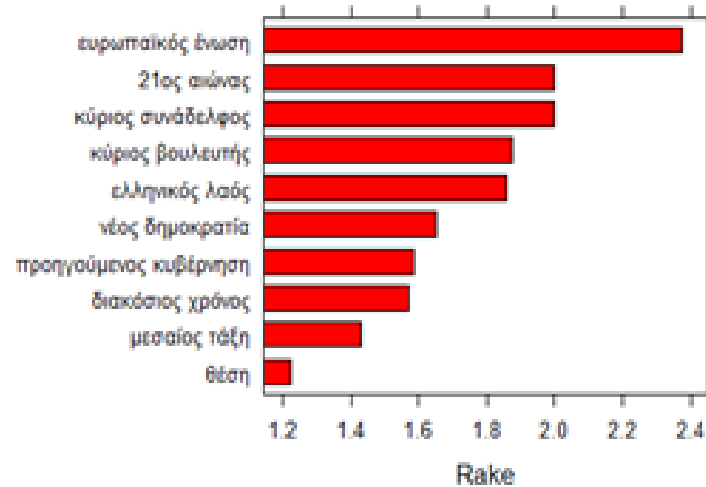
Εφαρμογή αλγόριθμου RAKE (Rapid Automatic Keyword Extraction)

Ο αλγόριθμος μηχανικής μάθησης RAKE "επιχειρεί" να προσδιορίσει φράσεις και λέξεις κλειδιά που περιέχονται σε ένα κείμενο.

Το αποτέλεσμα αυτό το επιτυγχάνει αφαιρώντας τα σημεία στίξης και τις "ουδέτερες" λέξεις από το κείμενο και υπολογίζοντας τις αλληλουχίες συνεχόμενων λέξεων ,το "βάρος" της κάθε λέξης και τη συχνότητα εμφάνισης της.

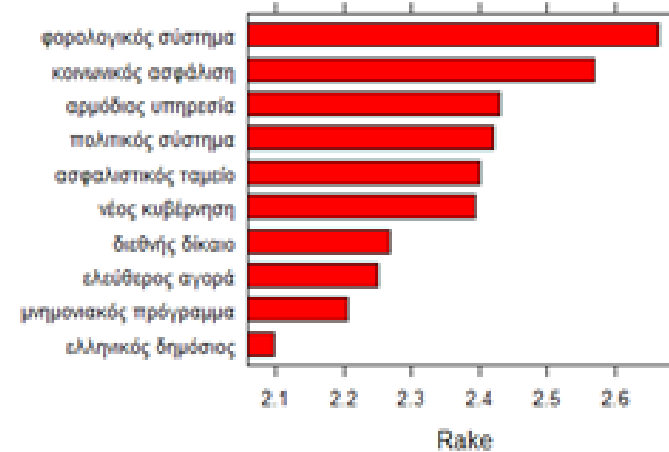
Μητσοτάκης

Keywords identified by RAKE



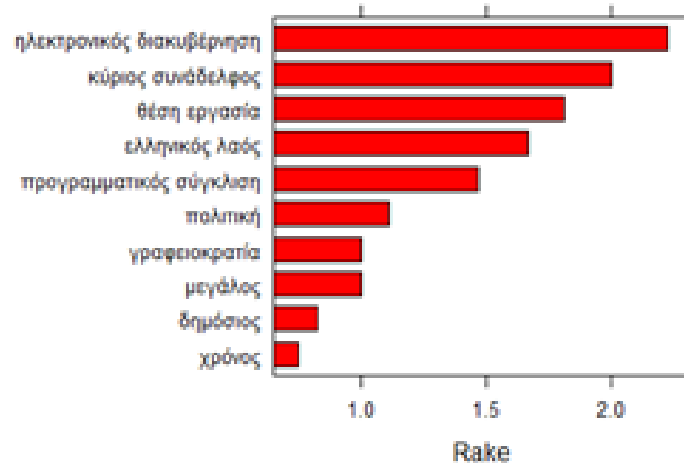
Τσίπρας

Keywords identified by RAKE



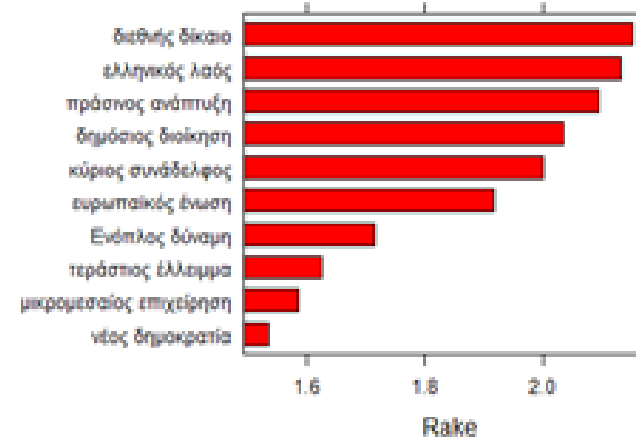
Σοφαράς

Keywords identified by RAKE



Παπανδρέου

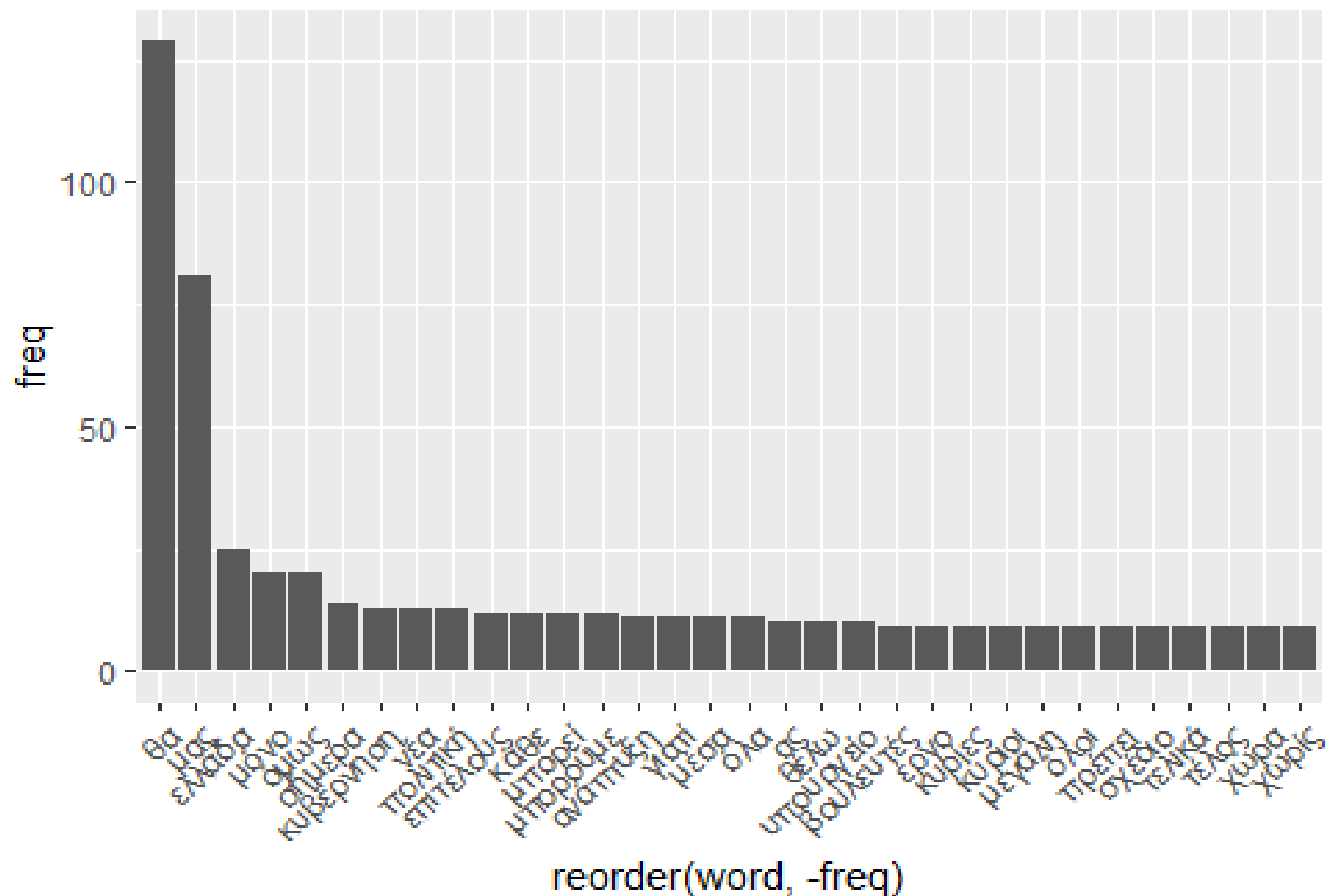
Keywords identified by RAKE



ΣΥΧΝΟΤΗΤΕΣ ΛΕΞΕΩΝ – ΓΡΑΦΗΜΑΤΑ

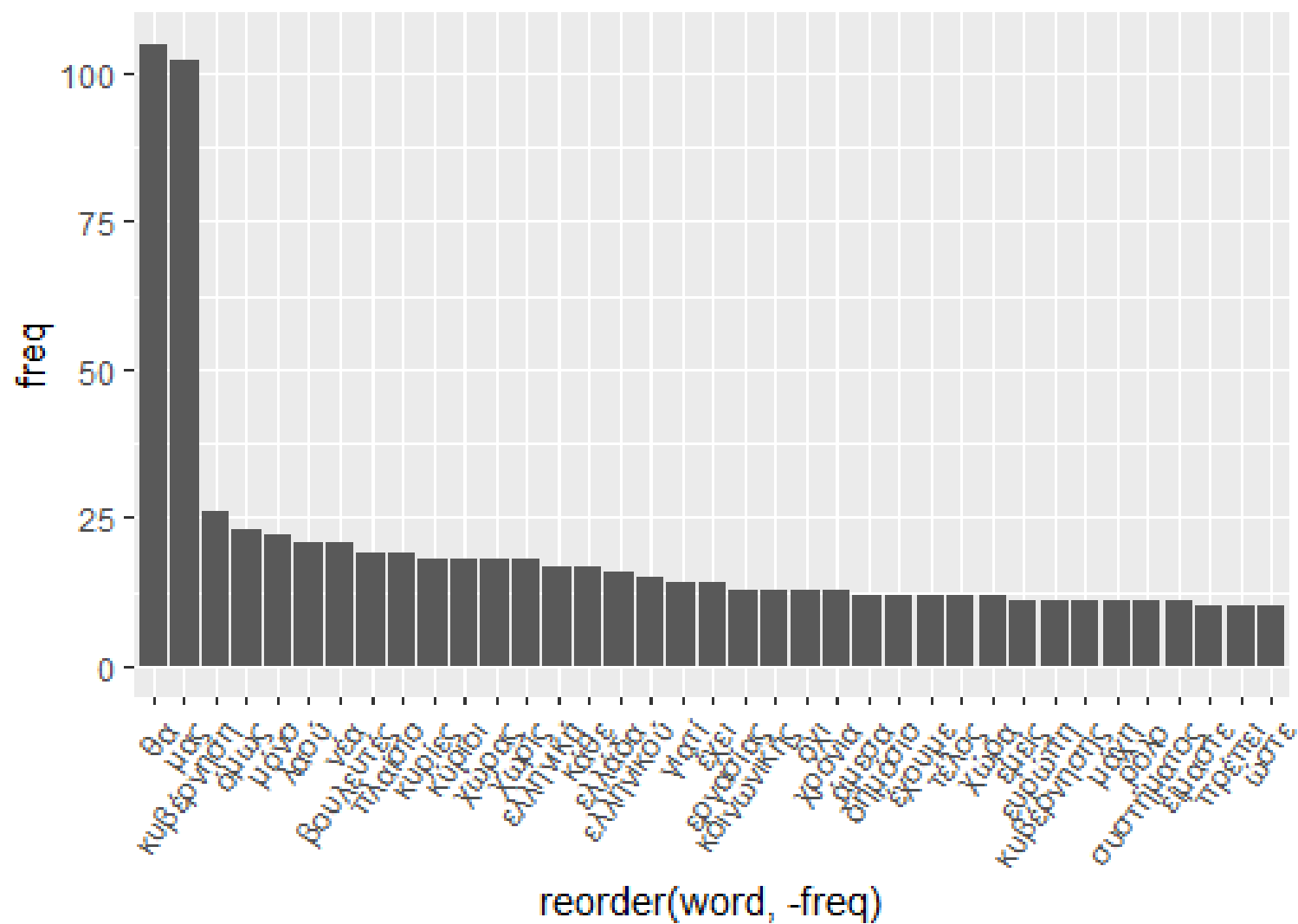
Μητσοτάκης

Λίστα συχνοτήτων λέξεων

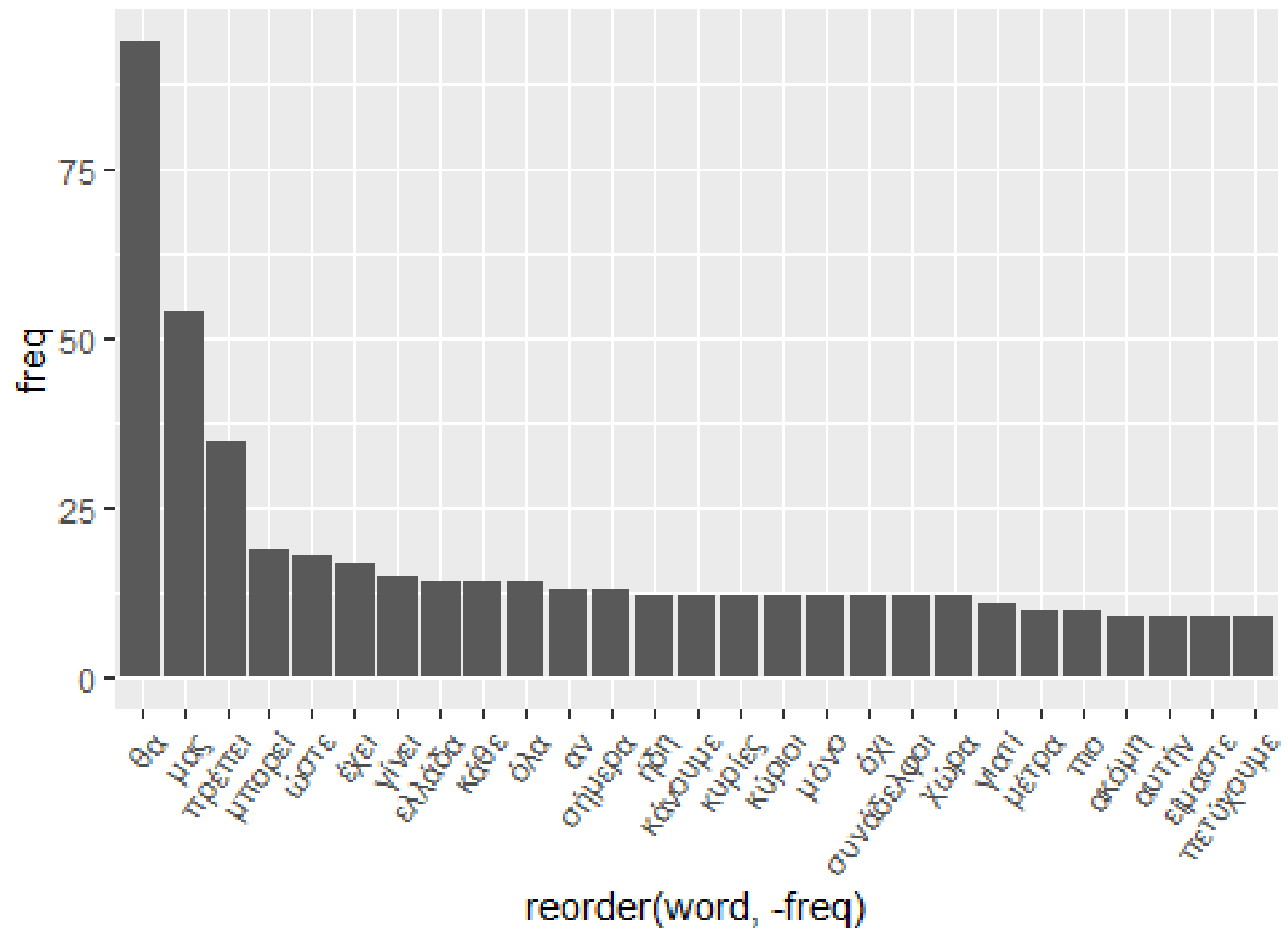


Τσίπρας

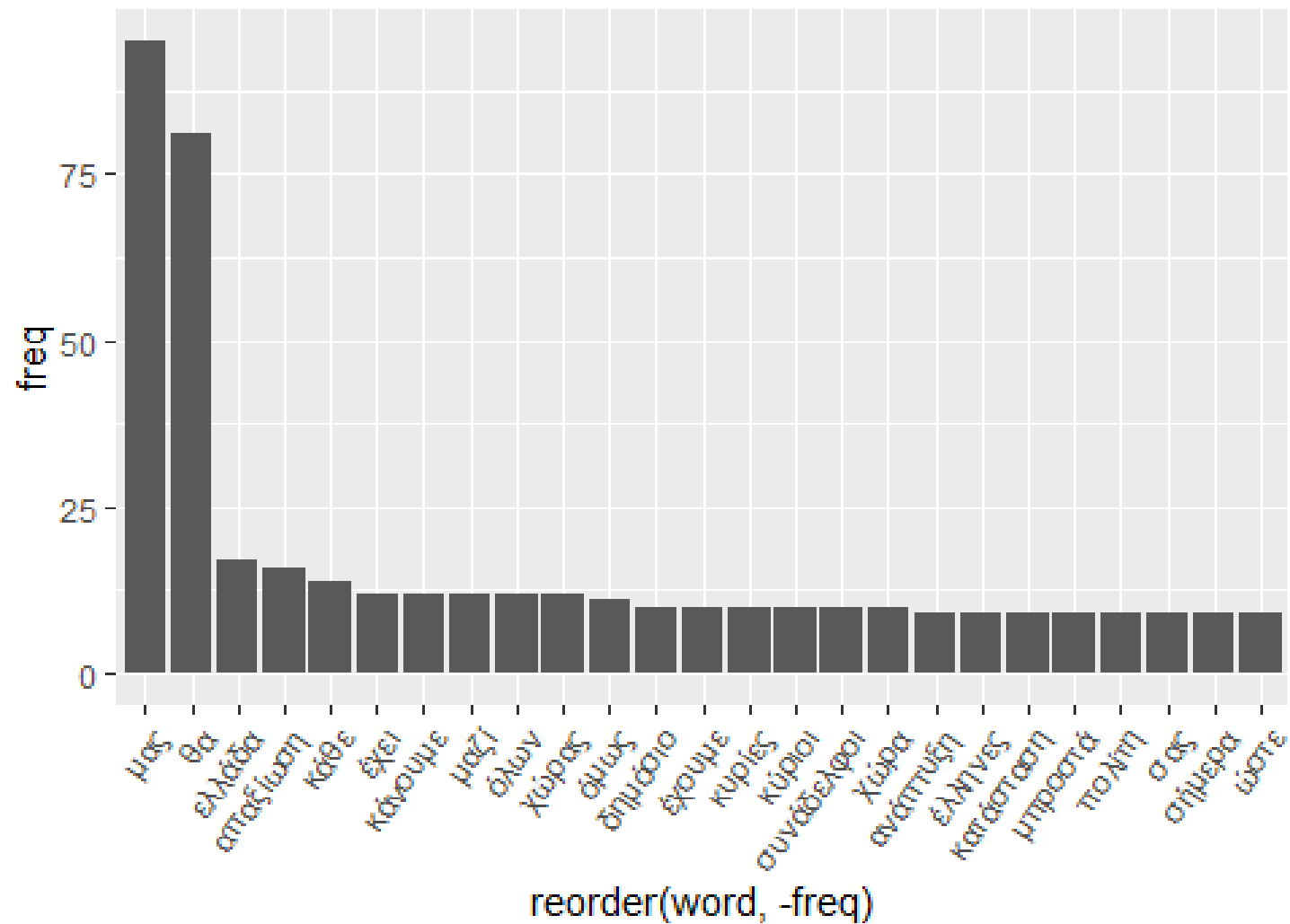
Λίστα συχνοτήτων λέξεων



Λίστα συχνοτήτων λέξεων



Λίστα συχνοτήτων λέξεων





Σαμαράς

WORD CLOUD



