

## **Tema nr. 1**

**-PSDT-**

**Analiza unui set de date format din 7 poezii**

## Descrierea setului de date analizat

Setul de date poezii este format din 7 poezii în limba engleza în care fiecare linie oferă informații cu privire la numărul versului în cadrul poeziei, versul și titlul poeziei din care face parte versul. Structura setului de date neprocesat este prezentat în Fig. 1 a).

line	versuri	title
1	You may write me down in history	Still_I_Rise
2	With your bitter, twisted lies,	Still_I_Rise
3	You may trod me in the very dirt	Still_I_Rise
4	But still, like dust, I'll rise.	Still_I_Rise

a)

line	title	word
1	Still_I_Rise	you
1	Still_I_Rise	may
1	Still_I_Rise	write
1	Still_I_Rise	me
1	Still_I_Rise	down

b)

Fig. 1. Structura setului de date analizat a) înainte de procesare și b) după procesare.

Structura setului de date după aplicarea funcției `unnest_tokens()` este prezentat în Fig. 1 b). Observăm că setul inițial conține un număr de 375 de observații, în timp ce setul de date procesat este format din 2546 observații (Fig. 2 a) și b)). Setul de date conține atât cuvinte de legătură, cât și cuvinte ce pot exprima sentimente.

```
> dim(poezii)
[1] 375 3
```

a)

```
> dim(poezii_tokenized)
[1] 2546 3
```

b)

Fig. 2. Structura setului de date analizat a) înainte de procesare și b) după procesare.

În Fig. 3 a) și b) este prezentată sintaxa în R folosită și rezultatul obținut la nivelul întregului set de date.

```
temp <- poezii_tokenized %>%
  group_by(word) %>%
  count(word, sort = TRUE)

> dim(temp)
[1] 1077 2
```

a)

Nr. Crt.	word	n
1	the	171
2	and	115
3	of	64
4	a	61
5	in	39
6	i	36

b)

Fig. 3. Prezentăm a) sintaxa în R și b) rezultatul obținut.

Observăm că la nivelul setului de date avem un număr de 1077 de cuvinte unice (Fig. 3 a) și b)). Prin urmare, ne propunem să vedem câte cuvinte de legătură avem și care sunt cele mai folosite, respectiv câte cuvinte urmează să analizăm.

```

cuvinte_legatura <- temp %>%
  inner_join(stop_words) %>%
  select(-lexicon) %>%
  as.data.frame() %>%
  distinct(word, n)
-----
> dim(cuvinte_legatura)
[1] 210 2

```

a)

Nr.Crt.	word	n
1	the	171
2	and	115
3	of	64
4	a	61
5	in	39
6	i	36

b)

Fig. 4. Prezentam a) sintaxa în R și dimensiunea setului de date analizat si b) rezultatul obtinut.

Astfel, avem un număr de 210 de cuvinte de legatura la nivelul întregului set de date (Fig. 4 a)), iar cel mai folosit cuvânt este *the* cu 171 apariții (Fig. 4 b)), în timp ce cuvântul *thou* este cel mai utilizat cuvânt cu un numar de 18 aparitii (Fig. 5 b)).

```

temp <- poezii_tokenized %>%
  group_by(word) %>%
  filter(word != '') %>%
  count(word, sort = TRUE) %>%
  anti_join(stop_words) %>%
  ungroup() %>%
  arrange(desc(n)) %>%
  mutate(nr = row_number())
> dim(temp)
[1] 867 3

```

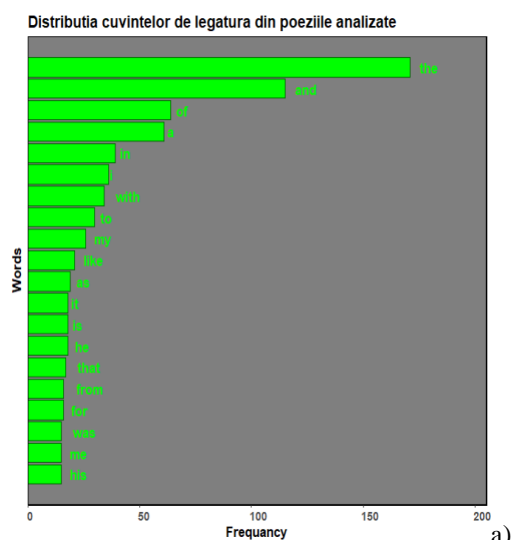
a)

Nr.Crt.	word	n	nr
1	thou	18	1
2	thy	13	2
3	hear	10	3
4	rise	10	4
5	thee	10	5
6	ice	8	6

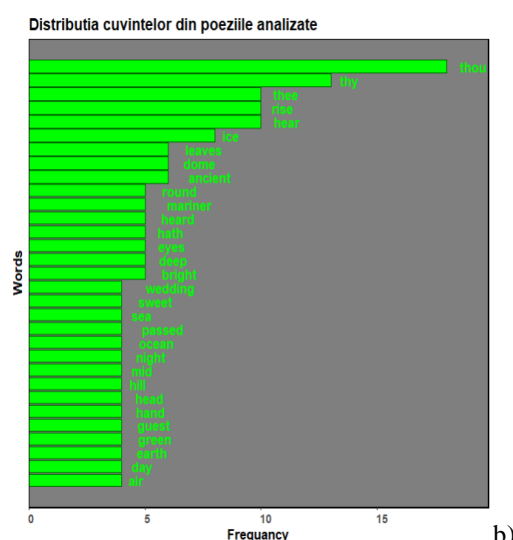
b)

Fig. 5. Prezentam a) sintaxa în R și dimensiunea setului de date analizat si b) rezultatul obtinut.

Și cum o imagine vorbește mai mult decăt o mie de cuvinte în Fig. 6 a) si b) prezentăm distribuția cuvintelor de legatura și a cuvintelor folosite de autori.



a)



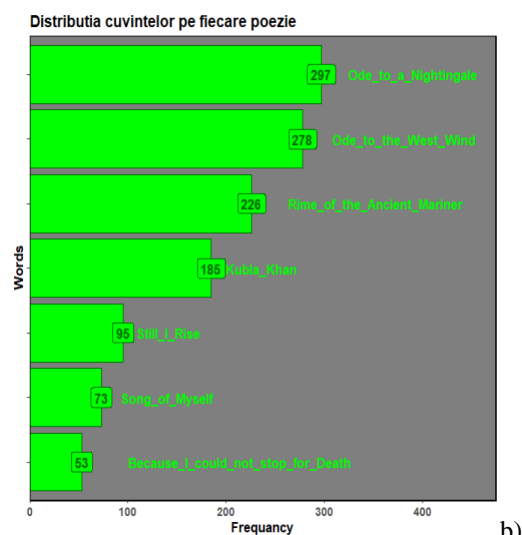
b)

Fig. 6. Distributia a) cuvintelor de legatura și b) cuvintelor.

## Analiza setului de date fără cuvinte de legatura

Nr.Crt.	title
1	Still_I_Rise
2	Kubla_Khan
3	Song_of_Myself
4	Because_I_could_not_stop_for_Death
5	Ode_to_a_Nightingale
6	Rime_of_the_Ancient_Mariner
7	Ode_to_the_West_Wind

a)



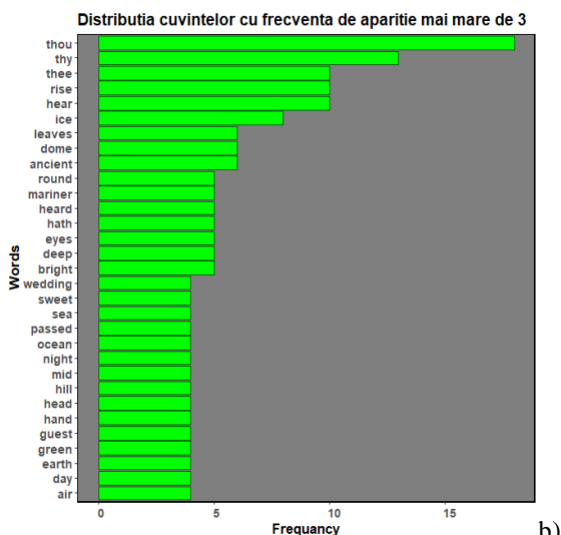
b)

Fig. 7. Prezntam a) denumirea poeziilor și b) distribuția cuvintelor pe fiecare poezie analizata.

In Fig. 7 b) este prezentată distribuția cuvintelor pe fiecare poezie. Observam ca poezia *Ode\_to\_a\_Nightingale* si *Ode\_to\_the\_West\_Wind* contin cele mai multe cuvinte, respectiv 297 de cuvinte și 185 de cuvinte urmate de, iar la polul opus se afla *Song\_of\_Myself* si *Because\_I\_could\_not\_stop\_for\_Death* cu cate 73 de cuvinte, respective 53 de cuvinte.

Nr.Crt.	word	n
1	thou	18
2	thy	13
3	hear	10
4	rise	10
5	thee	10
6	ice	8
7	ancient	6
8	dome	6
9	leaves	6
10	bright	5

a)



b)

Fig. 8. Distribuția cuvintelor la nivelul poeziilor analizate.

Din Fig. 8 a) observam ca cel mai utilizat cuvântul este *thou* si apare in poezia *Ode\_to\_the\_West\_Wind*, dar și în alte poezii, urmat de cuvântul *rise*, *thy*, *ice* etc.

In Fig. 9 sunt prezentate distribuțiile cuvintelor la nivelul fiecărei poezii în parte. Observam ca și în acest caz cuvântul *thou* din poezia *Ode\_to\_the\_West\_Wind* se afla in top, fiind cel mai utilizat cuvinte din poezia *Ode\_to\_the\_West\_Wind*.

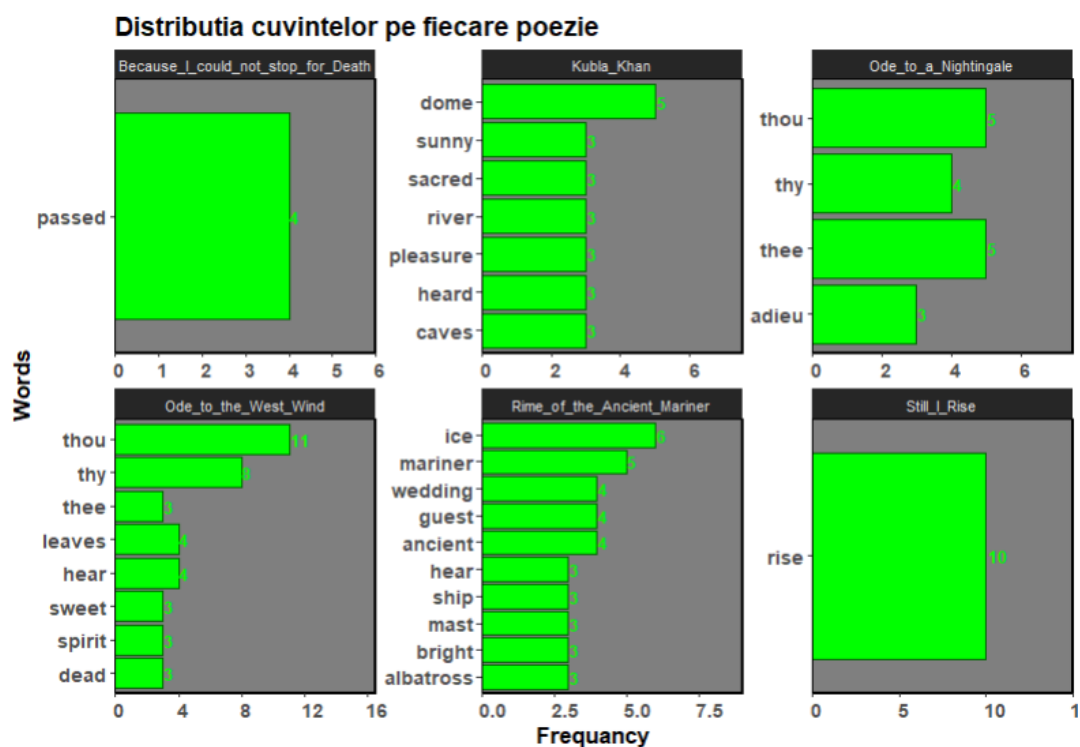


Fig. 9. Distribuția cuvintelor la nivelul fiecărei poezii analizate.

In Fig. 9 prezentăm distribuția cuvintelor care apar cu o frecvență mai mare de 3 în șase poezii diferite: *Because I could not stop for Death*, *Kubla Khan*, *Ode to a Nightingale*, *Still I Rise* etc.

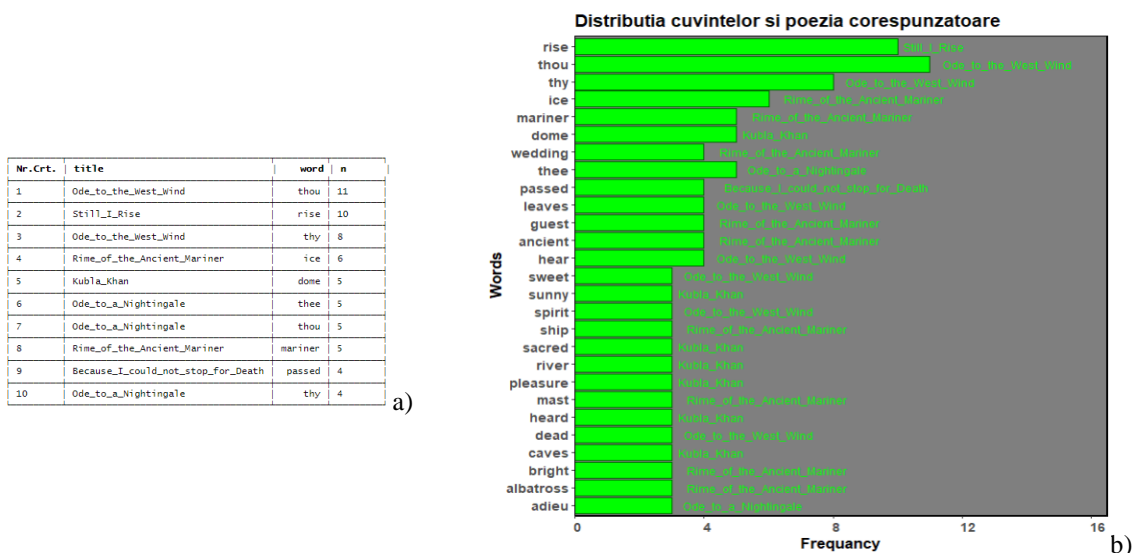


Fig. 10. Prezentam a) denumirea poeziilor și b) distribuția cuvintelor pe fiecare poezie analizata.

Fiecare grafic individual arată cuvintele cheie din fiecare poezie și frecvența lor. De exemplu, in poezia *Because I could not stop for Death*: cuvântul *passed* apare de 4 ori, in timp ce in poezia *Kubla Khan*: cuvintele *dome* și *sunny* apar de 5 ori, dar apar și cuvinte precum

*sacred, river, pleasure, heard* și *caves*, fiecare de 3 ori. In poezia *Ode to a Nightingale*: cuvintele *thou* și *thee* apar de 5 ori fiecare, iar *thy* de 4 ori, iar in poezia *Still I Rise*: cuvântul *rise* apare de 10 ori, fiind, cel mai probabil, tema poeziei (Fig. 10 a) și b)). De exemplu, cuvântul *thou* are 11 apariții în poezia *Ode to the West Wind*, în timp ce în poezia *Rime of the Ancient Mariner* și *Ode to a Nightingale* are 5, respective 2 apariții. Un alt cuvânt ce apare în 3 poezii diferite, *Ode to the West Wind*, *Rime of the Ancient Mariner* și *Ode to a Nightingale*, este *thy*.

Nr.Crt.	title	word	n
1	Ode_to_the_West_Wind	thou	11
2	Ode_to_a_Nightingale	thou	5
3	Rime_of_the_Ancient_Mariner	thou	2

a)

Nr.Crt.	title	word	n
1	Ode_to_the_West_Wind	thy	8
2	Ode_to_a_Nightingale	thy	4
3	Rime_of_the_Ancient_Mariner	thy	1

b)

Nr.Crt.	title	word	n
1	Kubla_Khan	air	1
2	Ode_to_a_Nightingale	air	1
3	Ode_to_the_West_Wind	air	1
4	Still_I_Rise	air	1

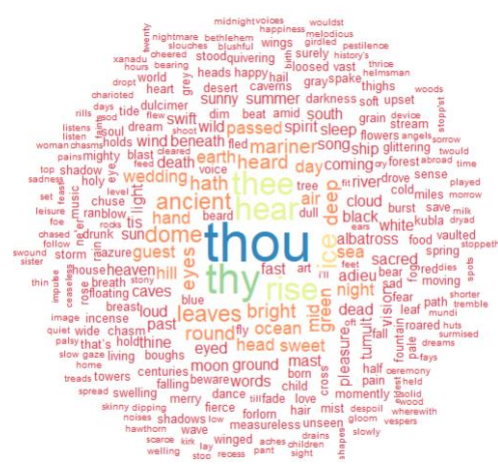
c)

Nr.Crt.	title	word	n
1	Rime_of_the_Ancient_Mariner	ice	6
2	Kubla_Khan	ice	2

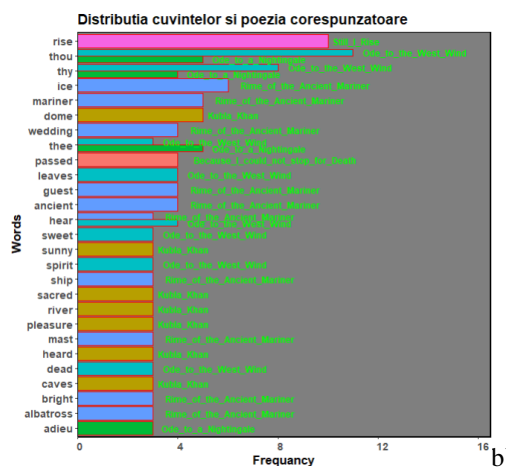
d)

Fig. 11. Frecvența de apariție a cuvintelor a) thou, b) thy, c) air și d) ice în poezii.

Cuvântul *air* apare o singura dată în 4 poezii diferite: *Ode to the West Wind*, *Still I Rise*, *Kubla Khan* și *Ode to a Nightingale*, în timp ce *ice* are 6 apariții în poezia *Rime of the Ancient Mariner* și 2 apariții în poezia *Kubla Khan*.



a)



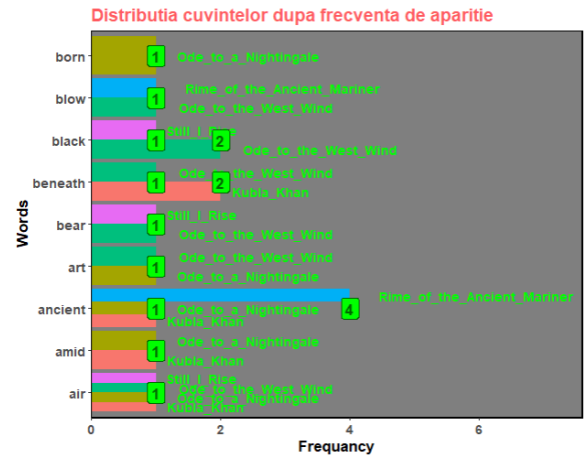
b)

Fig. 12. Prezentam a) norul de cuvinte și b) distribuția cuvintelor pe fiecare poezie analizată.

Pentru a ne face o imagine mai clară cu privire la cuvintele comune și frecvența lor de apariție în mai mult de o poezie în Fig. 13 sunt prezentate frecvența de apariție a cuvintelor din diferite poezii.

Nr.Crt.	word	title	n
1	air	Ode_to_a_Nightingale	1
2	air	Still_I_Rise	1
3	dance	Ode_to_a_Nightingale	1
4	dance	Still_I_Rise	1
5	dream	Ode_to_a_Nightingale	1
6	dream	Still_I_Rise	1
7	eyes	Ode_to_a_Nightingale	1
8	eyes	Still_I_Rise	2
9	pain	Ode_to_a_Nightingale	1
10	pain	Still_I_Rise	1
11	past	Ode_to_a_Nightingale	2
12	past	Still_I_Rise	1

a)



b)

Fig. 13. Prezentam a) frecvența de apariție cuvintelor din diferite poezii și b) distribuția cuvintelor comune din diferite poezii analizate.

## Analiza poeziei Still\_I\_Rise

Fiecare linie din setul de date oferă informații cu privire la numărul versului în cadrul poeziei, versul și titlul poeziei din care face parte versul. Structura setului de date neprocesat este prezentat în Fig. 14 a).

Structura dataframe-ului Still\_I\_Rise analizat

Setul de date: contine 1 poezii

line	versuri	title
1	You may write me down in history	Still_I_Rise
2	With your bitter, twisted lies,	Still_I_Rise
3	You may trod me in the very dirt	Still_I_Rise
4	But still, like dust, I'll rise.	Still_I_Rise
5	Does my sassiness upset you?	Still_I_Rise
6	Why are you beset with gloom?	Still_I_Rise
7	'Cause I walk like I've got oil wells	Still_I_Rise

a)

Structura dataframe-ului analizat

Setul de date: contine 1 poezii

line	title	word
1	Still_I_Rise	you
1	Still_I_Rise	may
1	Still_I_Rise	write
1	Still_I_Rise	me
1	Still_I_Rise	down
1	Still_I_Rise	in

b)

Fig. 14. Structura setului de date analizat a) înainte de procesare și b) după procesare.

În Fig. 14 b) este prezentat rezultatul obținut după aplicarea funcției `unnest_tokens()`.

Observăm că poezia inițială conține un număr de 43 de versuri/observații, în timp ce poezia procesată este formată din 240 de cuvinte/observații (Fig. 15 a) și b)). Setul de date conține atât cuvinte de legătură, cât și cuvinte ce pot exprima sentimente.

```
> dim(poezii %>%
+   filter(title == 'Still_I_Rise'))
[1] 43  3
```

a)

```
> dim(poezii_tokenized %>%
+   filter(title == 'Still_I_Rise'))
[1] 240 3
```

b)

Fig. 15. Structura setului de date analizat a) înainte de procesare și b) după procesare.

În Fig. 16 a) și b) este prezentată sintaxa în R folosită și rezultatul obținut la nivelul poeziei *Still I Rise*.

```
temp <- poezii_tokenized %>%
  filter(title == 'Still_I_Rise') %>%
  group_by(word) %>%
  count(word, sort = TRUE)

> dim(temp)
[1] 133 2
```

a)

Nr.Crt.	word	n
1	i	12
2	you	11
3	rise	10
4	like	9
5	the	9
6	my	8
7	and	6
8	in	6

b)

Fig. 16. Prezentăm a) sintaxa în R și b) rezultatul obținut.

Observăm că poezia *Still I Rise* conține un număr de 133 de cuvinte unice (Fig. 16 a) și b)). Prin urmare, ne propunem să vedem câte cuvinte de legatură avem și care sunt cele mai folosite, respectiv câte cuvinte urmează să analizăm.

```
cuvinte_legatura <- temp %>%
  inner_join(stop_words) %>%
  select(-lexicon) %>%
  as.data.frame() %>%
  distinct(word, n)

> dim(cuvinte_legatura)
[1] 50 2
```

a)

Nr.Crt.	word	n
1	i	12
2	you	11
3	like	9
4	the	9
5	my	8
6	and	6

b)

Fig. 17. Prezentăm a) sintaxa în R și dimensiunea setului de date analizat și b) rezultatul obținut.

Astfel, avem un număr de 50 de cuvinte de legatură la nivelul poeziei (Fig. 17 a)), iar cel mai folosit cuvânt este *I* cu 12 apariții (Fig. 17 b)), în timp ce cuvântul *rise* este cel mai utilizat cuvânt cu un număr de 10 apariții (Fig. 18 b)).

```
temp <- poezii_tokenized %>%
  group_by(word) %>%
  filter(word != '') %>%
  filter(title == 'Still_I_Rise') %>%
  count(word, sort = TRUE) %>%
  anti_join(stop_words) %>%
  ungroup() %>%
  arrange(desc(n)) %>%
  mutate(nr = row_number())

> dim(temp)
[1] 83 3
```

a)

Nr.Crt.	word	n	nr
1	rise	10	1
2	eyes	2	2
3	that's	2	3
4	upset	2	4
5	air	1	5
6	ancestors	1	6
7	awful	1	7

b)

Fig. 18. Prezentăm a) sintaxa în R și dimensiunea setului de date analizat și b) rezultatul obținut.

În Fig. 19 a) și b) prezentăm distribuția cuvintelor de legatură și a cuvintelor folosite de autor în poezia *Still I Rise* care sunt în număr de 95 de cuvinte.

Nr.Crt.	title	n
1	Still_I_Rise	95



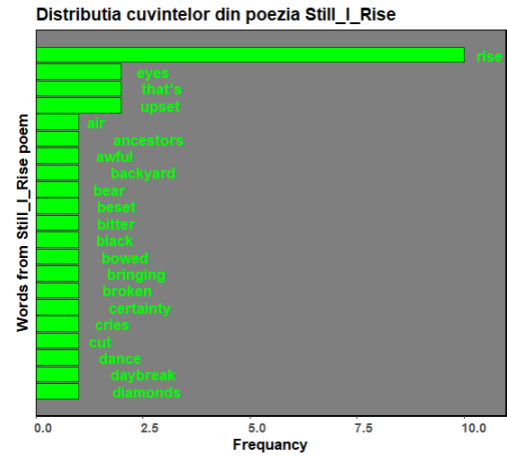
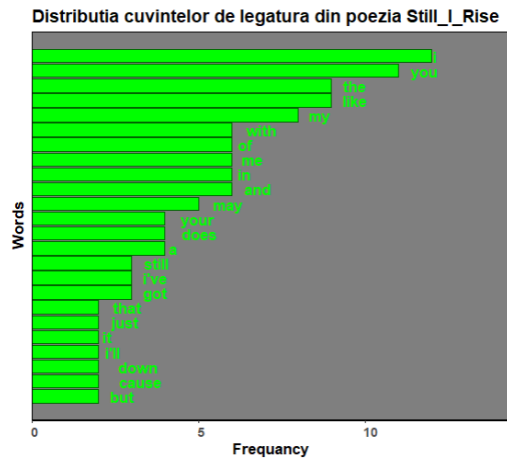


Fig. 19. Distribuția a) cuvintelor de legătură și b) cuvintelor folosite de autor.

## Analiza cuvintelor după frecvența de apariție a acestora in diferite poezii

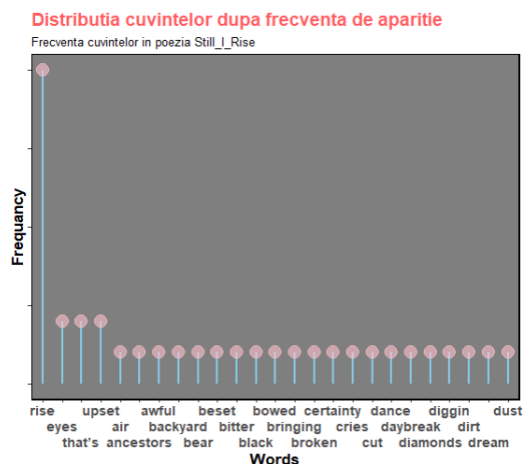


Fig. 20. Prezentăm a) distribuția cuvintelor și b) norul cuvintelor pe poezia *Still\_I\_Rise*.

In Fig. 20 a) si b) – Fig. 23 a) si b) sunt prezentate distribuțiile cuvintelor în poeziile analizate. Se poate observa ca cel mai folosit cuvânt in poezia *Still I Rise* este *rise*.

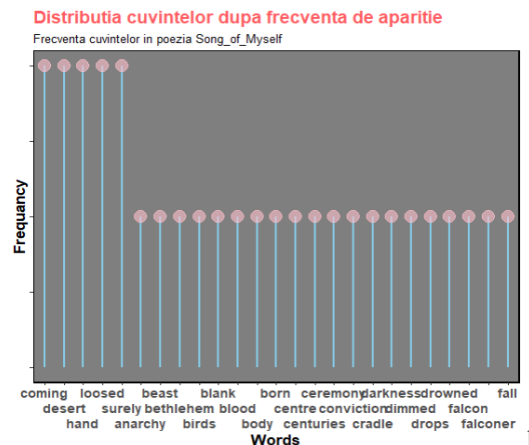
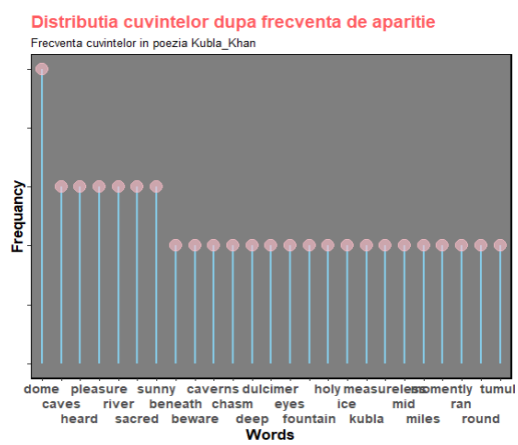


Fig. 21. Distribuția cuvintelor în poezia a) *Kubla\_Khan* și b) *Song\_of\_Myself*.

Observăm că în poezia *Kubla\_Khan*, *Song\_of\_Myself*, *Because\_I\_could\_not\_stop\_for\_Death*, respectiv *Ode\_to\_a\_Nightingale* cele mai utilizate cuvinte sunt *dome*, *coming*, *passed*, respectiv *thee/thou*.

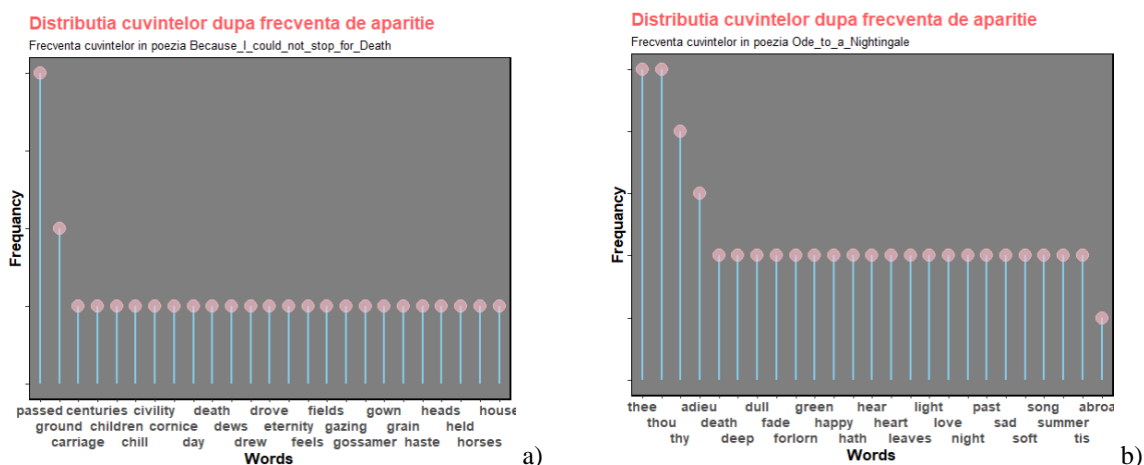


Fig. 22. Distribuția cuvintelor în poezia a) *Because\_I\_could\_not\_stop\_for\_Death* și b) *Ode\_to\_a\_Nightingale*.

În poezia *Rime\_of\_the\_Ancient\_Mariner* și *Ode\_to\_the\_West\_Wind* cele mai utilizate cuvinte sunt *ice*, respectiv *thou* după cum se poate observa și din Fig. a) și b).

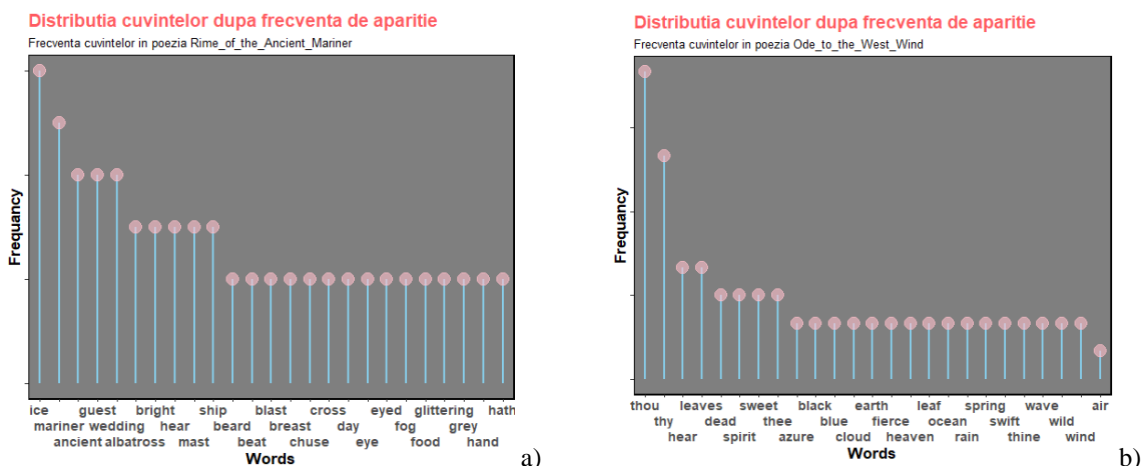


Fig. 23. Distribuția cuvintelor în poezia a) *Rime\_of\_the\_Ancient\_Mariner* și b) *Ode\_to\_the\_West\_Wind*.

**Analiza comparativa a frecvenței de apariție a cuvintelor din poezia *Still I Rise* și alte poezii (*Rime of the Ancient Mariner*, *Ode\_to\_the\_West\_Wind* și *Ode to a Nightingale*)**

Analiza comparativă a frecvenței de apariție a cuvintelor din poezia *Still I Rise* și celelalte poezii (*Rime of the Ancient Mariner*, *Ode\_to\_the\_West\_Wind* și *Ode to a Nightingale*), respectiv rezultatul testului de corelație sunt prezentate în Fig. 24 a) și b).

Observăm că între poezia *Still I Rise* și poezia *Ode to a Nightingale* există o corelație de intensitate foarte mică, negativă, dar semnificativă ( $p\text{-value} < 0.05$ ). Intervalul de încredere

este foarte larg și include valoarea zero fiind în concordanță cu rezultatul testului de corelație (Fig. b)). O asociere negativă între textele celor două poezii indică faptul că atunci când proporția unui cuvânt din poezia *Still I Rise* crește, proporția aceluiași cuvânt în poezia *Ode to a Nightingale* scade (cuvintele care sunt cele mai frecvent folosite în poezia *Still I Rise* sunt rare în poezia *Ode to a Nightingale* (Fig. 25 a) – c))).

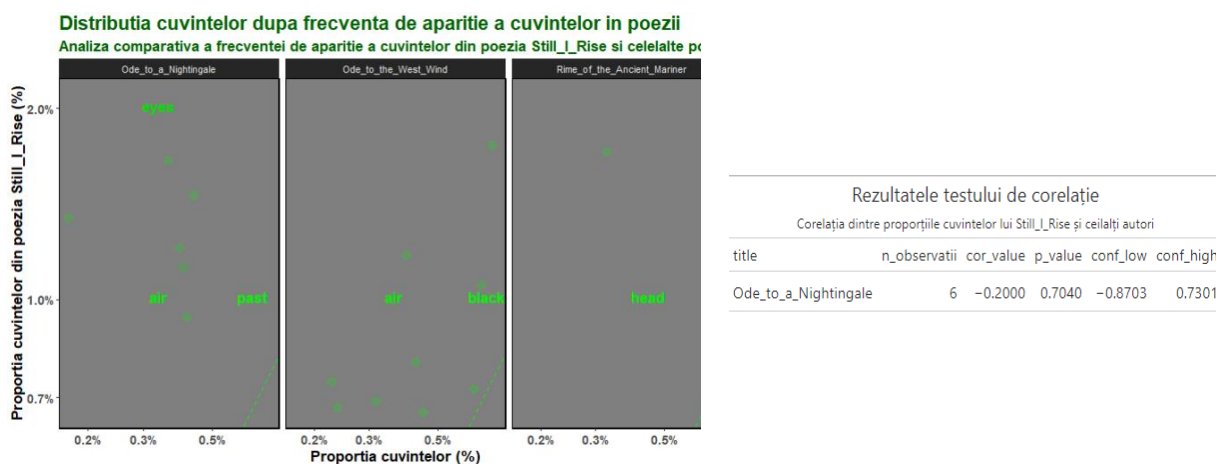


Fig. 24. Analiza comparativa a a) frecvenței de apariție a cuvintelor din poezia *Still I Rise* și alte poezii si b) rezultatul testului de corelație.

De exemplu, poezia *Still I Rise* folosește anumite cuvinte asociate cu subiectul poeziei, în timp ce poezia *Ode to a Nightingale* folosește un alt vocabular sau cuvintele comune, dar într-o proporție mică. Din punct de vedere al limbajului folosit cele două poezii sunt foarte diferite. Valoarea coeficientului de corelație de -0.200 dintre poezia *Still I Rise* și poezia *Ode to a Nightingale* sugerează o corelație negativă foarte slabă. Ceea ce indică faptul că pe măsură ce proporțiile cuvintelor din poezia *Ode to a Nightingale* cresc, cele din poezia *Still I Rise* pot să scadă (Fig. 25 c)). Același lucru îl constatăm și în cazul analizei legăturii dintre poezia *Still I Rise* și poeziile *Rime of the Ancient Mariner* și *Ode to the West Wind*, pentru care valoarea coeficientului de corelație este zero și care sugerează că nu există o legătură semnificativă între proporțiile cuvintelor din *Still I Rise* și celelalte poezii. Aceste poezii tratează teme/subiecte diferite, iar cuvintele comune care apar sunt pur întâmplătoare neexistând un pattern în utilizarea cuvintelor (Fig. a) – c)).

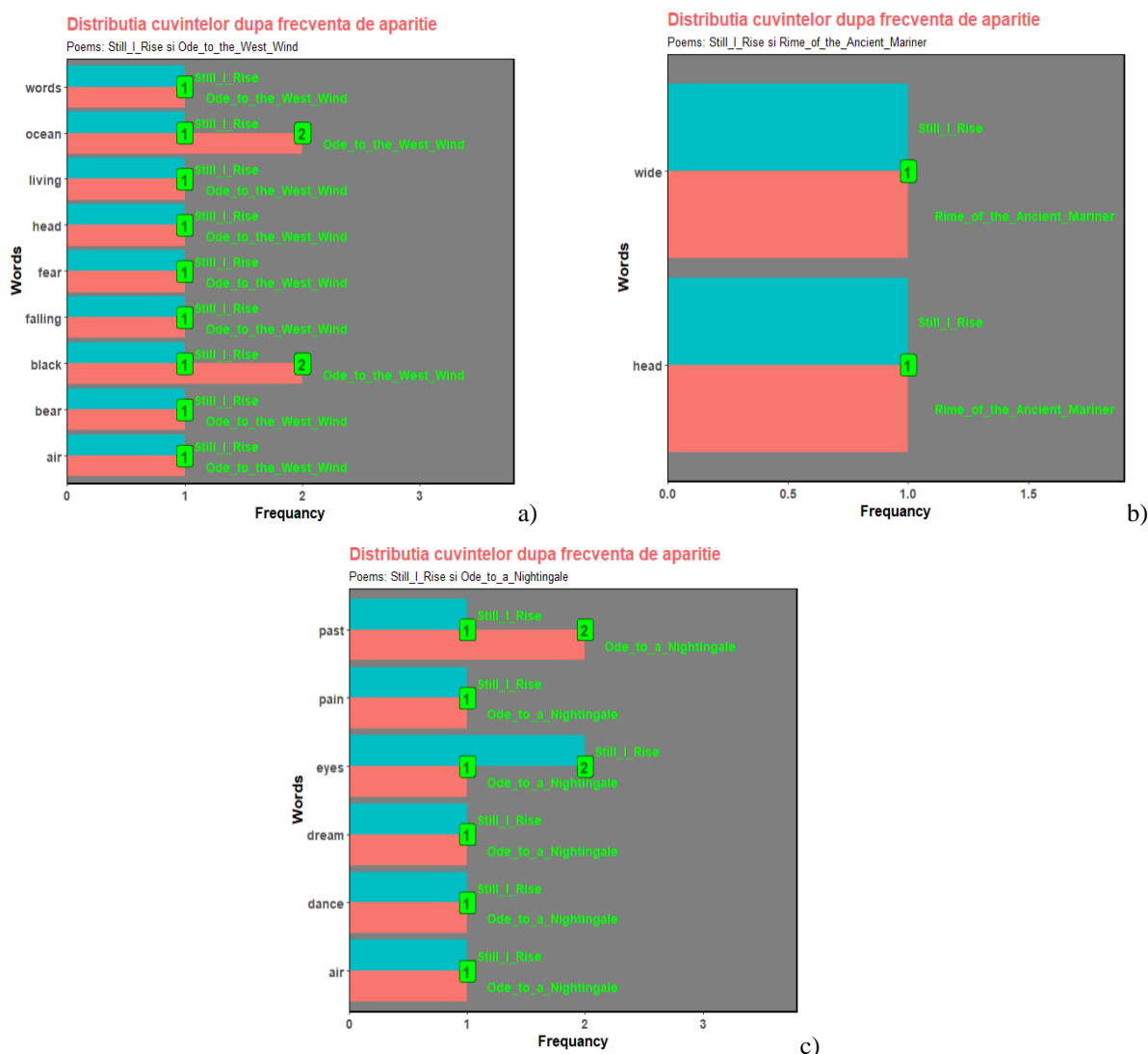


Fig. 25. Analiza comparativa a frecvenței de apariție a cuvintelor din poezia *Still I Rise* și a) poezia *Ode\_to\_the\_West\_Wind*, b) poezia *Rime of the Ancient Mariner* și c) poezia *Ode to a Nightingale*.

## Analiza comparativa a frecvenței de apariție a cuvintelor din poezia *Ode to a Nightingale* și alte poezii

Analiza comparativa a frecvenței de apariție a cuvintelor din poezia *Ode to a Nightingale* și celelalte poezii (*Rime of the Ancient Mariner*, *Ode\_to\_the\_West\_Wind* și *Still I Rise*), respectiv rezultatul testului de corelație sunt prezentate în Fig. 26 a) și b).

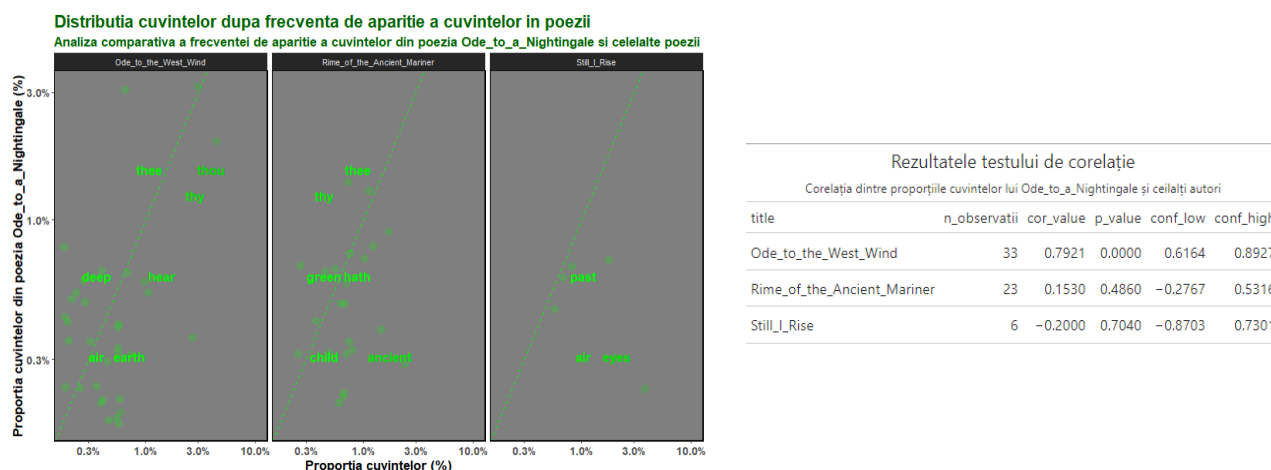
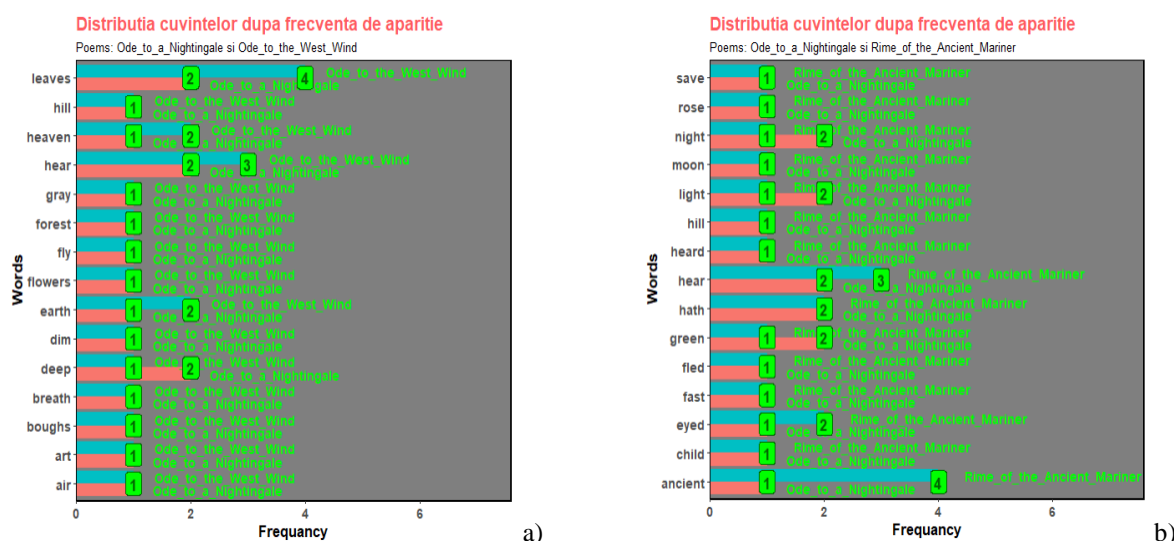


Fig. 26. Analiza comparativa a a) frecvenței de apariție a cuvintelor din poezia *Ode to a Nightingale* și alte poezii și b) rezultatul testului de corelație.

După cum se poate observa poezia *Ode to a Nightingale* pare sa aiba asociere cu toate poeziile analizate. Astfel, observam ca cea mai mare asociere exista intre poezia *Ode to a Nightingale* si poezia *Ode\_to\_the\_West\_Wind*, lucru sustinut atat de rezultatul testului de corelație ( $p - \text{value} < 0.05 \Rightarrow$  se respinge ipoteza nula, cu un risc asumat de 5%), cat și de valoarea coeficientului de corelație care înregistrează cea mai ridicată valoare (Fig. 26 b)), 0.79, dintre toate valorile obtinute pentru coeficientul de corelatie (cu alte cuvinte cuvintele care sunt frecvente în poezia *Ode to a Nightingale* au tendinta sa fie frecvente si in poezia *Ode\_to\_the\_West\_Wind* (Fig. 27 a) – c))).



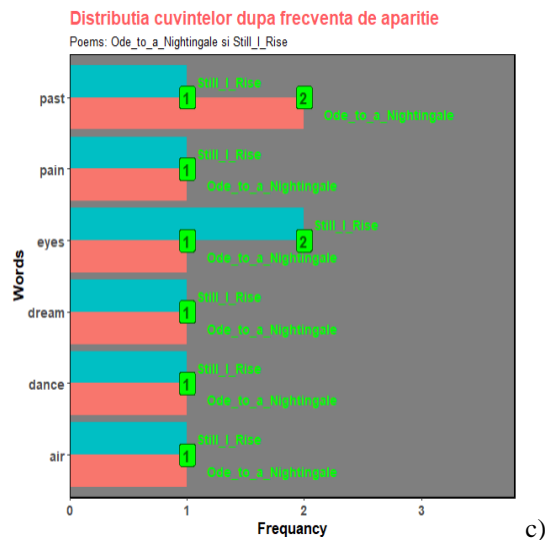


Fig. 27. Analiza comparativa a frecvenței de apariție a cuvintelor din poezia *Ode to a Nightingale* si a) poezia

*Ode\_to\_the\_West\_Wind*, b) poezia *Rime of the Ancient Mariner* si c) poezia *Still I Rise*.

Observam ca intre poezia *Ode to a Nightingale* si poeziile *Rime of the Ancient Mariner*, *Ode\_to\_the\_West\_Wind* si *Still I Rise* exista o asociere pozitiva, respectiv negativa (în cazul analizei coeficientului de corelatie dintre poezia *Ode to a Nightingale* si poezia *Still I Rise*). De asemenea, coeficientul de corelatie asociat legaturii dintre poezia *Ode to a Nightingale* si poeziile *Ode\_to\_the\_West\_Wind* si *Still I Rise* nu sunt semnificativ statistic ( $p - \text{value} > 0.05$ ).

### Analiza comparativa a frecvenței de apariție a cuvintelor din poezia *Ode to the West Wind* si alte poezii

In Fig. 28 a) si b) este prezentată analiza comparative a frecvenței de apariție a cuvintelor din poezia *Ode\_to\_the\_West\_Wind* și celelalte poezii (*Rime of the Ancient Mariner* si *Ode to a Nightingale*), respectiv rezultatul testului de corelație.

Dupa cum se poate observa intre poezia *Ode\_to\_the\_West\_Wind* si *Ode to a Nightingale* exista o legatura puternica, lucru sustinut atat de valoarea coeficientului de corelație, cat și de rezultatul testului ( $p - \text{value} < 0.05 \Rightarrow$  se respinge ipoteza nula, cu un risc asumat de 5%). Semnul coeficientului de corelație, atât în cazul legăturii dintre poezie *Ode\_to\_the\_West\_Wind* si poezia *Rime of the Ancient Mariner*, cat si in cazul legaturii dintre *Ode\_to\_the\_West\_Wind* si poezia *Ode to a Nightingale*, indica faptul ca corelatia este pozitiva sau directa, lucru susținut și de reprezentarea grafica (Fig. 28 a)). Adica, pe masura ce proporția unui cuvânt din poezia *Ode\_to\_the\_West\_Wind* crește, proporția aceluiași cuvânt în celelalte poezii crește. De exemplu, cuvintele *leaves*, *heaven*, *hear* sau *earth* etc. sunt folosite atat in poezia *Ode\_to\_the\_West\_Wind*, cat si in poezia *Ode to a Nightingale*, motiv pentru care exista

o asociere între textele celor doua poezii, acestea abordand, cel mai probabil, o tema/subiect similară sau autorii folosesc un vocabular similar (Fig. 29 a)).

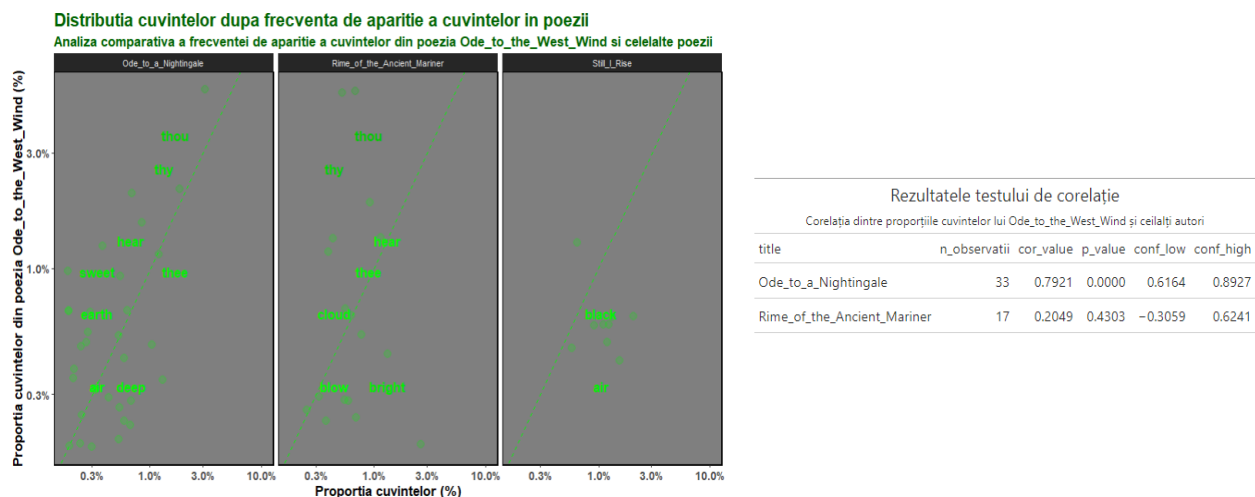
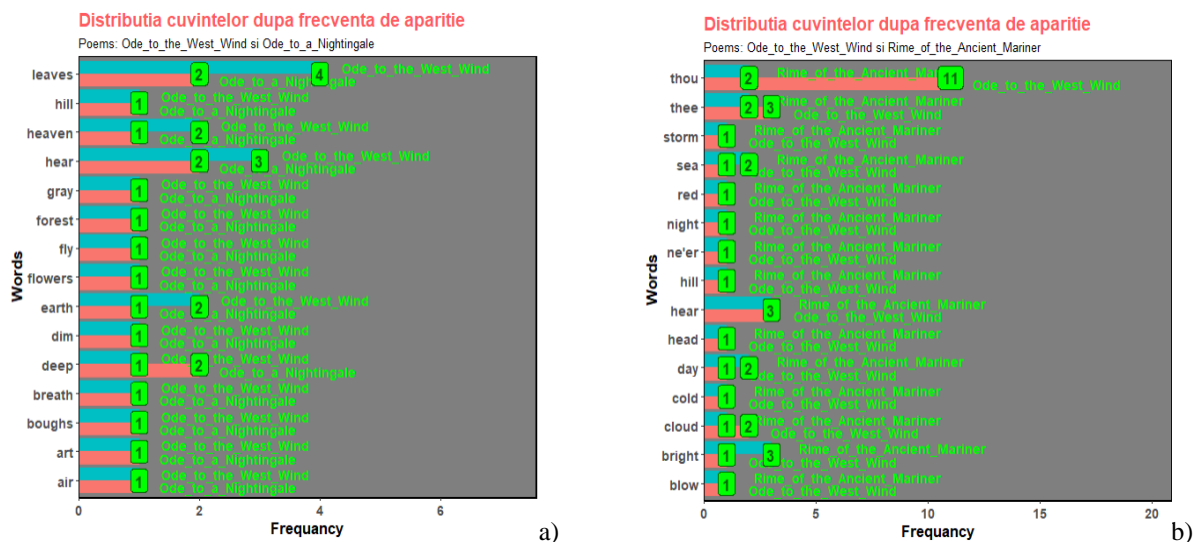


Fig. 28. Analiza comparativa a a) frecvenței de apariție a cuvintelor din poezia *Ode\_to\_the\_West\_Wind* si alte poezii si b) rezultatul testului de corelație.

De asemenea, se poate observa ca intre poezia *Ode\_to\_the\_West\_Wind* si poezia *Rime of the Ancient Mariner* exista o corelație de intensitate slabă, pozitivă, dar semnificativ statistic ( $p - value > 0.05$ ). Acest lucru sugerează ca, deși autorii folosesc un vocabular similar, exista o tendinta mica ca frecvențele cuvintelor sa varieze impreuna (Fig. 28 a) si b)). De exemplu, daca in poezia *Ode\_to\_the\_West\_Wind* cuvantul *thou* este folosit mai des (de 11 ori), in poezia *Rime of the Ancient Mariner* acest cuvânt este folosit, dar nu la fel (de 2 ori). Acest lucru poate sugera și faptul ca temele abordate sunt diferite (Fig. 29 b)).





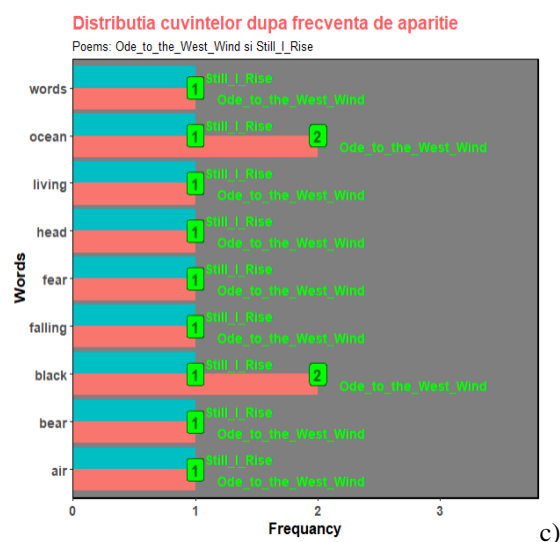


Fig. 29. Analiza comparativa a frecvenței de apariție a cuvintelor din poezia *Ode\_to\_the\_West\_Wind* si a) poezia *Ode to a Nightingale*, b) poezia *Rime of the Ancient Mariner* si c) poezia *Still I Rise*.

În ceea ce privește legătura între poezia *Ode\_to\_the\_West\_Wind* și poezia *Still I Rise* observăm că nu există nici o legătură între modul cum sunt folosite cuvintele de autori sau aceștia abordează teme diferite. Deși observăm că unele cuvinte apar în ambele poezii, apariția lor este pur întâmplătoare și nu urmează un pattern (Fig. 29 c)).

## Analiza comparativa a frecvenței de apariție a cuvintelor din poezia *Rime\_of\_the\_Ancient\_Mariner* si alte poezii

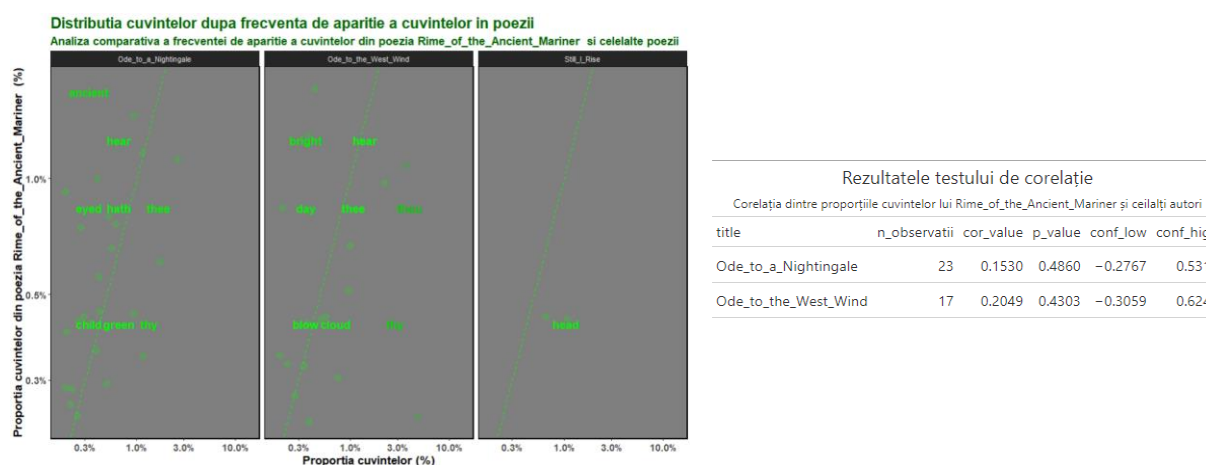


Fig. 30. Analiza comparativa a a) frecvenței de apariție a cuvintelor din poezia *Rime\_of\_the\_Ancient\_Mariner* si alte poezii si b) rezultatul testului de corelație.



Rezultatul testului de corelație obținut între proporțiile cuvintelor din poezia *Rime of the Ancient Mariner* și celelalte poezii (*Ode to a Nightingale* și *Ode to the West Wind*) arată că valoarea coeficientului de corelație este pozitiv și foarte mic (0.1530, respectiv 0.205), ceea ce indică o legatură directă, slabă, semnificativă ( $p - \text{value} > 0.05$ ), dar pozitivă. Interval de încredere pentru coeficientul de corelație este cuprins între  $[-0.2767, 0.5316]$ , respectiv  $[-0.3059, 0.6241]$  ceea ce arată că nu există o asociere între textele acestor poezii (nu există o legatură între proporțiile cuvintelor din *Rime of the Ancient Mariner* și *Ode to a Nightingale*, aceasta fiind nesemnificativă statistic (Fig. 30 a) și b))).

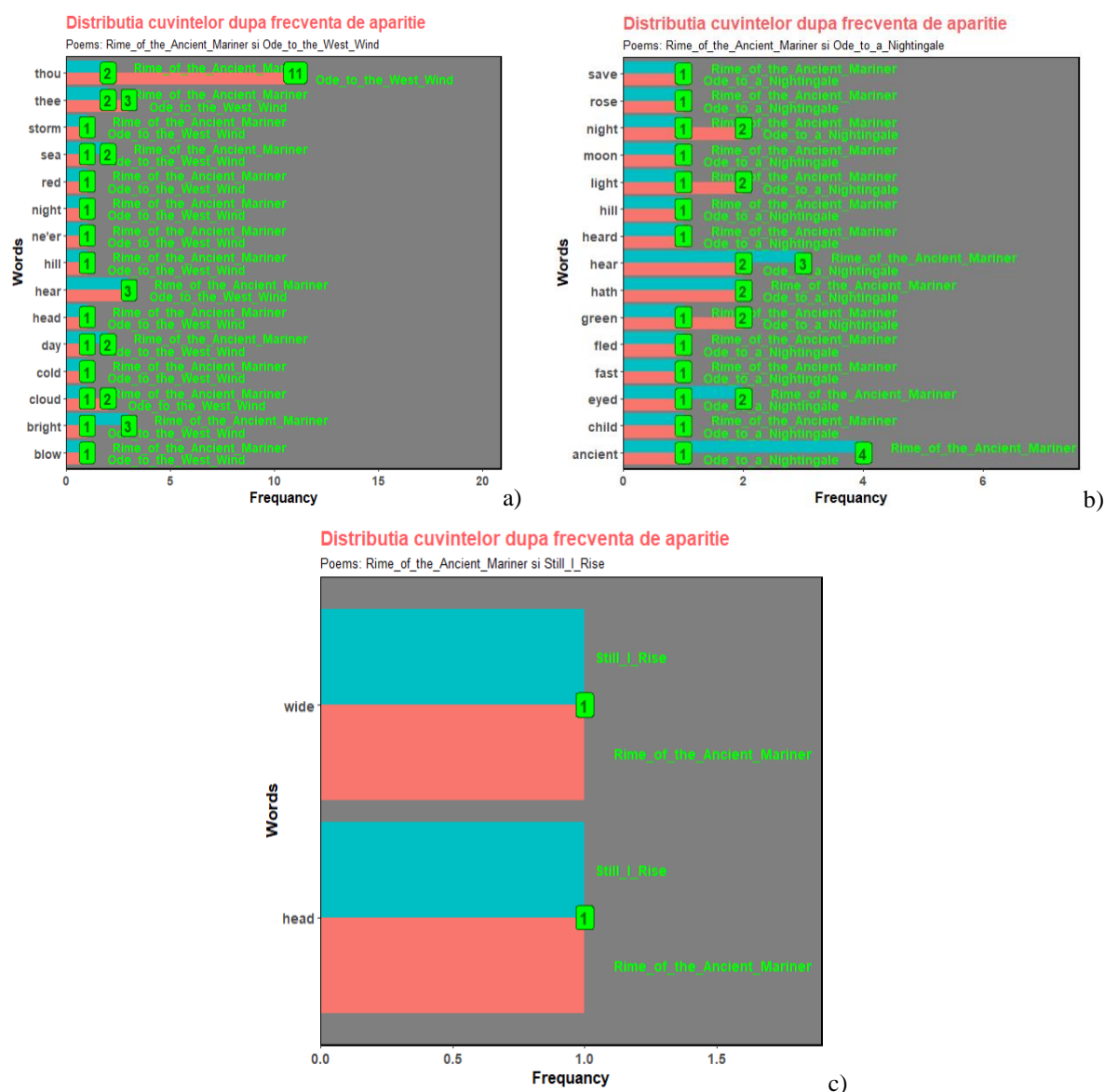


Fig. 31. Analiza comparativa a frecvenței de apariție a cuvintelor din poezia *Ode to the West Wind* și a) poezia *Ode to a Nightingale*, b) poezia *Rime of the Ancient Mariner* și c) poezia *Still I Rise*.

Prin urmare, nu exista nicio legatura intre modul cum sunt folosite cuvintele de autori sau aceștia abordează teme foarte diferite. Deși observam ca unele cuvinte apar în ambele poezii, apariția lor este intamplatoare si nu urmează un pattern, textele și temele celor doua poezii sunt foarte diferite (Fig. 31 b)).

In ceea ce priveste legatura dintre poezia *Rime of the Ancient Mariner* si poezia *Still I Rise* observam din Fig. 30 a) si Fig. 31 c) ca nu exista nici o legatura intre modul cum sunt folosite cuvintele de autori sau aceștia abordează teme foarte diferite sau folosesc un vocabular foarte diferit. Deși observam ca unele cuvinte apar în ambele poezii, apariția lor este pur intamplatoare si nu urmează un pattern (Fig. 31 c)). Prin urmare, din Fig. 30 a) si Fig. 31 c) observam ca vocabularul folosit de cei doi autori este foarte diferit (doar doua cuvinte comune) sau temele abordate sunt diametral opuse.