

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/359879072>

Ο Πελοποννησιακός Πόλεμος: Προεπισκόπηση με N-grams & Παραγοντική Ανάλυση

Presentation · August 2019

DOI: 10.13140/RG.2.2.13438.25929

CITATIONS

0

READ

1

2 authors, including:



Nikos Koutsoupias

University of Macedonia

121 PUBLICATIONS 113 CITATIONS

[SEE PROFILE](#)


Some of the authors of this publication are also working on these related projects:



GeDiAn [View project](#)




Economic Crisis [View project](#)



Ο Πελοποννησιακός Πόλεμος: Προεπισκόπηση με N-grams & Παραγοντική Ανάλυση

Νικόλαος Κουτσουπιάς & Κυριάκος Μικέλης
Πανεπιστήμιο Μακεδονίας



Ο Πελοποννησιακός Πόλεμος:
Προεπισκόπηση με N-grams & Παραγοντική Ανάλυση

Περιεχόμενα

- Σώμα Κειμένου (το έργο “Πόλεμος” του Θουκυδίδη)
- Συχνότητες, Νέφη Λέξεων (wordclouds) & Λογισμικά
- Εξόρυξη Κειμένου & Παραγοντική Ανάλυση (ΠΑ)
- Ngrams & ΠΑ
- Συμπεράσματα & Επόμενα Βήματα

Ο Πελοποννησιακός Πόλεμος: Προεπισκόπηση με N-grams & Παραγοντική Ανάλυση

Corpus

- Project Gutenberg's The History of the Peloponnesian War, by Thucydides
- (Original: History -21yrs /of 27th)
- Translator: Richard Crawley
- Release Date: December, 2004 [EBook #7142]
- Language: English

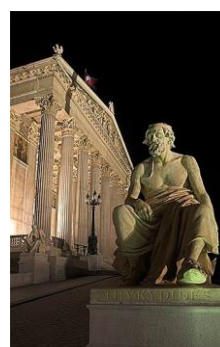


[Σώμα](#) | [Συχνότητες](#) | [Εξόρυξη](#) | [Ngrams](#) | [Συμπεράσματα](#)

Ο Πελοποννησιακός Πόλεμος: Προεπισκόπηση με N-grams & Παραγοντική Ανάλυση

Corpus

- Project Gutenberg's The History of the Peloponnesian War, by Thucydides
- 200k WORDS
- 26 CHAPTERS
- 8 BOOKS



[Σώμα](#) | [Συχνότητες](#) | [Εξόρυξη](#) | [Ngrams](#) | [Συμπεράσματα](#)

Ο Πελοποννησιακός Πόλεμος:
Προεπισκόπηση με N-grams & Παραγοντική Ανάλυση

- **Συχνότητες, Νέφη Λέξεων & Λογισμικά**

Software Used:

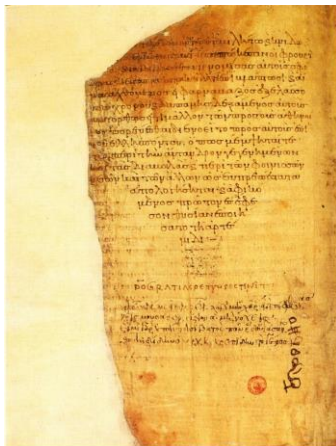
- - R
- - R-STUDIO

- R packages:
 - -FactoMineR
 - -Factoshiny
 - -factoextra
 - -tm
 - -ngram

Σώμα | Συχνότητες | Εξόρυξη | Ngrams | Συμπεράσματα

Ο Πελοποννησιακός Πόλεμος:
Προεπισκόπηση με N-grams & Παραγοντική Ανάλυση

- **Συχνότητες, Νέφη Λέξεων & Λογισμικά**



Book	Words (unigrams)
b1	32579
b2	25217
b3	24778
b4	29358
b5	21725
b6	25229
b7	21804
b8	24503

Σώμα | Συχνότητες | Εξόρυξη | Ngrams | Συμπεράσματα

Ο Πελοποννησιακός Πόλεμος:
Προεπισκόπηση με N-grams & Παραγοντική Ανάλυση

• Συχνότητες, Νέφη Λέξεων & Λογισμικά



WORDCLOUDS: Related Research

- ◆ IEEE Trans on visualization & computer graphics
- ◆ J of the Association for Inf Science & Technology
- ◆ Journal of Applied Statistics
- ◆ Online Learning
- ◆ Interactive Learning Environments
- ◆ The Gerontologist
- ◆ J of Environmental Planning & Management
- ◆ J of Small Business & Enterprise Development
- ◆ European J of Management & Business Economics

Σύμα | Συχνότητες | Εξόρυξη | Ngrams | Συμπεράσματα

Ο Πελοποννησιακός Πόλεμος:
Προεπισκόπηση με N-grams & Παραγοντική Ανάλυση

• Συχνότητες, Νέφη Λέξεων & Λογισμικά



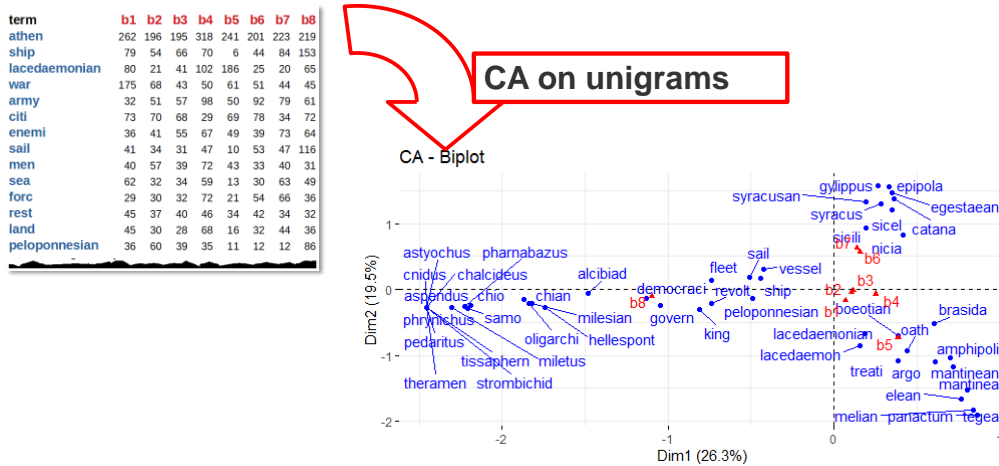
TERM	b1	b2	b3	b4	b5	b6	b7	b8
athenian	187	32	140	276	178	169	203	179
war	175	68	43	50	61	51	44	45
all	104	34	54	55	123	46	51	49
power	86	22	19	11	28	30	5	18
lacedaemonian	80	21	41	102	186	25	20	65
ship	79	54	66	70	6	44	84	153
athen	75	64	55	40	63	32	20	40
cit	73	70	68	29	69	78	34	72
corinthian	72	6	6	15	29	9	35	10
time	70	28	48	53	37	35	37	46
one	68	61	60	84	41	75	74	73
now	65	45	67	74	60	88	101	99
sea	62	32	34	59	13	30	63	49
first	58	47	52	58	41	59	52	46
also	57	77	52	50	46	57	54	71
place	57	45	42	65	44	32	29	34
made	55	46	48	48	45	26	28	45
corcyraean	53	2	17	9	0	0	2	0

1st Book Wordcloud

Σύμα | Συχνότητες | Εξόρυξη | Ngrams | Συμπεράσματα

Ο Πελοποννησιακός Πόλεμος:
 Προεπισκόπηση με N-grams & Παραγοντική Ανάλυση

- Συχνότητες, Νέφη Λέξεων & Λογισμικά



Σώμα | Συχνότητες | Εξόρυξη | Ngrams | Συμπεράσματα

Ο Πελοποννησιακός Πόλεμος:
 Προεπισκόπηση με N-grams & Παραγοντική Ανάλυση

- Εξόρυξη Κείμενου & Παραγοντική Ανάλυση (ΠΑ)

Corpus
 The History of the Peloponnesian War, by
 Thucydides
 200k Words / 26 Chapters / 8 Books

Stopword Removal & WordStemming

N-gram (bigrams) Extraction

Book	bigrams
b1	13581
b2	10934
b3	10603
b4	12408
b5	8889
b6	10517
b7	9116
b8	10094

Σώμα | Συχνότητες | Εξόρυξη | Ngrams | Συμπεράσματα

Ο Πελοποννησιακός Πόλεμος:
 Προεπισκόπηση με N-grams & Παραγοντική Ανάλυση

- Εξόρυξη Κείμενου & Παραγοντική Ανάλυση (ΠΑ)

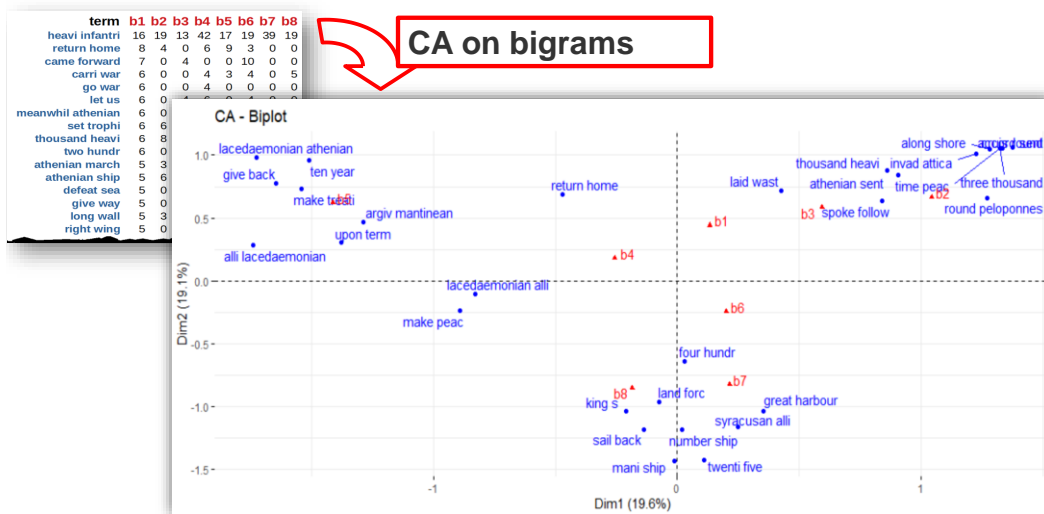
Book	Bi-grams
b1	13581
b2	10934
b3	10603
b4	12408
b5	8889
b6	10517
b7	9116
b8	10094

term	b1	b2	b3	b4	b5	b6	b7	b8
heavi infantri	16	19	13	42	17	19	39	19
return home	8	4	0	6	9	3	0	0
came forward	7	0	4	0	0	10	0	0
carri war	6	0	0	4	3	4	0	5
go war	6	0	0	4	0	0	0	0
let us	6	0	4	6	0	4	0	0
meanwhil athenian	6	0	5	7	0	5	0	4
set trophi	6	6	3	12	0	6	6	5
thousand heavi	6	8	4	3	0	0	0	0
two hundr	6	0	9	5	0	6	5	0
athenian march	5	3	0	0	0	0	0	0
athenian ship	5	6	7	5	0	0	4	7
defeat sea	5	0	0	0	0	0	3	0
give way	5	0	0	6	0	0	0	0
long wall	5	3	0	5	3	0	0	0
right wing	5	0	0	4	5	3	0	0

Σώμα | Συχνότητες | Εξόρυξη | Ngrams | Συμπεράσματα

Ο Πελοποννησιακός Πόλεμος:
 Προεπισκόπηση με N-grams & Παραγοντική Ανάλυση

- Ngrams & ΠΑ



Σώμα | Συχνότητες | Εξόρυξη | Ngrams | Συμπεράσματα

Ο Πελοποννησιακός Πόλεμος:
Προεπισκόπηση με N-grams & Παραγοντική Ανάλυση

Συμπεράσματα & Επόμενα Βήματα


- Τα bigrams συνδιασμό με την ΠΑ παρεχουν πληρέστερη εικόνα σύνοψης ιστορικού κειμενου σε σχέση με τα wordclouds και την ΠΑ σε unigrams
- Επομενα βήματα:
 - διερευνηση με 3-,4- ... grams
 - απεικόνιση με βάση επιμέρους ενότητες (Chapters)
 - αξιοποίηση της Ιερ. Ταξινόμηση στα αποτελέσματα της ΠΑ

Σώμα | Συχνότητες | Εξόρυξη | Ngrams | **Συμπεράσματα**

Ο Πελοποννησιακός Πόλεμος:
Προεπισκόπηση με N-grams & Παραγοντική Ανάλυση

Βιβλιογραφία

- DeNoyelles, A., & Reyes-Foster, B. (2015). Using word clouds in online discussions to support critical thinking and engagement. *Online Learning*, 19(4), n4.
- Felix, C., Franconeri, S., & Bertini, E. (2017). Taking word clouds apart: An empirical investigation of the design space for keyword summaries. *IEEE transactions on visualization and computer graphics*, 24(1), 657-666. DOI: [10.1109/TVCG.2017.2746018](https://doi.org/10.1109/TVCG.2017.2746018)
- Koutsoupas, N., & Mikelis, K. (2019). *Exploring International Relations Journal Articles: A Multivariate Approach*. SAGE Publications Ltd. DOI: 10.4135/9781526486189
- Koutsoupas, N., & Mikelis, K. (2019, August). Text, Content and Data Analysis of Journal Articles: The Field of International Relations. In *Conference of the International Federation of Classification Societies* (pp. 113-120). Springer, Cham. DOI: [10.1007/978-3-030-60104-1_13](https://doi.org/10.1007/978-3-030-60104-1_13)
- Leginus, M., Zhai, C., & Dolog, P. (2016). Personalized generation of word clouds from tweets. *Journal of the Association for Information Science and Technology*, 67(5), 1021-1032. DOI: [10.1002/asi.23494](https://doi.org/10.1002/asi.23494)
- Xie, Y., & Lin, S. Y. (2019). Using word clouds to support students' knowledge integration from online inquiry: An investigation of the process and outcome. *Interactive Learning Environments*, 27(4), 478-496. DOI: [10.1080/10494820.2018.1484774](https://doi.org/10.1080/10494820.2018.1484774)



Ο Πελοποννησιακός Πόλεμος:
Προεπισκόπηση με N-grams & Παραγοντική Ανάλυση

Νίκος Κουτσουπιάς & Κυριάκος Μικέλης
Πανεπιστήμιο Μακεδονίας

Ερωτήσεις / Παρατηρήσεις ?

The Peloponnesian War: Preview with N-grams and Correspondence Analysis Nikos Koutsoupias & Kyriakos Mikelis
University of Macedonia

Abstract

The work of Athenian historian and general Thucydides "The History of the Peloponnesian War" is examined in English with Correspondence Analysis and a combination of text mining methods and techniques as implemented in R-language software packages. The possibility of extracting a preview and a visual summary of the body of the historical text is explored using term frequencies in each of the eight books of the work, while visualizing the clouds of unigrams and bigrams.

Ο Πελοποννησιακός Πόλεμος: Προεπισκόπηση με N-grams και Παραγοντική Ανάλυση
Νικόλαος Κουτσουπιάς & Κυριάκος Μικέλης
Πανεπιστήμιο Μακεδονίας

Περίληψη

Το έργο του Αθηναίου ιστορικού και στρατηγού Θουκυδίδη "Η Ιστορία του Πελοποννησιακού Πολέμου" εξετάζεται σε αγγλική μετάφραση με την Παραγοντική Ανάλυση και συνδυασμό μεθόδων για την εξόρυξη κειμένου όπως υλοποιούνται σε λογισμικά πακέτα της γλώσσας R. Διερευνάται η δυνατότητα εξαγωγής προεπισκόπησης και σύνοψης του σώματος του ιστορικού κειμένου χρησιμοποιώντας συχνότητες όρων σε κάθε ένα από τα οκτώ βιβλία του έργου, με ταυτόχρονη οπτικοποίηση των νεφών "μονόλεκτων" (unigram) και "δίλεκτων" (bigram) λεκτικών μονάδων.