

# STATISTICS

HOME CANON SEARCH TEXT SEARCH BROWSE LEXICA N-GRAMS STATISTICS VOCABULARY TOOLS HELP

## STATISTICS:

Full Corpus  Author  Lemma

From  to



B.C. | A.D. | All

Include Varia and Incerta

The statistics view allows you to review word usage in the entire corpus (**Full corpus statistics**), in individual authors or works (**Author statistics**) or view the usage of a particular lemma (**Lemma statistics**).

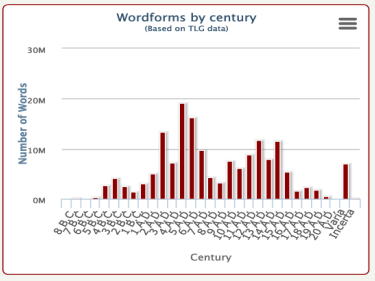
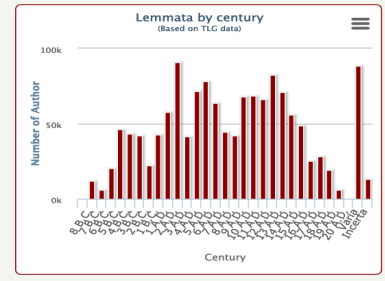
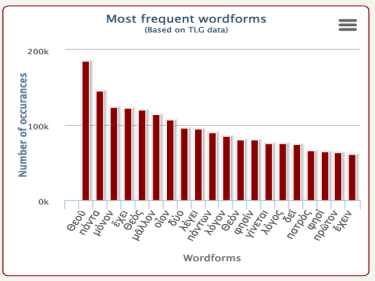
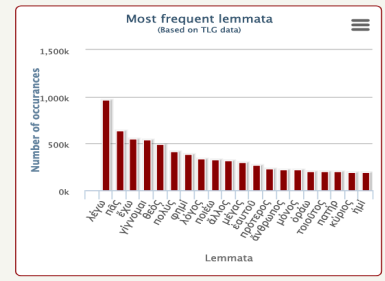
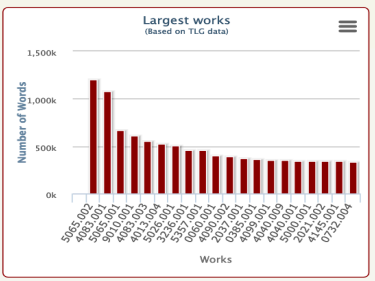
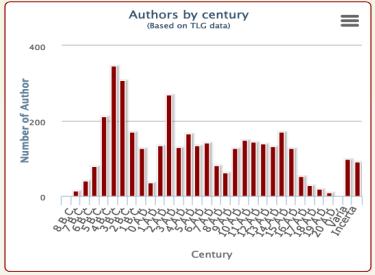
From 8 B.C. to A.D. 20



SUMMARY

- There are 10,869 documents in this corpus with a total of 115,244,474 words, 262,436 lemmata and 1,775,946 unique wordforms
- Author distribution by century is: 4 B.C. (346), 3 B.C. (309), A.D.2 (269) **(more)**
- The largest authors in the corpus are: JOANNES CHRYSOSTOMUS, GALENUS, BASILICA **(more)**
- The largest works in the corpus are: BASILICA, *Scholia in Basilicorum libros I-XI* [5065.002], EUSTATHIUS Thessalonicensis, *Commentarii ad Homeri Iliadem* [4083.001], BASILICA, *Basilica* [5065.001] **(more)**
- The most frequent lemmata in the corpus are: λέγω, πᾶς, πᾶσα, πᾶν, ἔχω **(more)**
- The most frequent wordforms in the corpus are: Θεοῦ, πάντα, μόνω **(more)**
- Lemmata are distributed by century as follows: A.D.2 (91,111), Varia (88,332), A.D.12 (82,519) **(more)**
- Wordforms are distributed by century as follows: A.D.4 (19,219,885), A.D.5 (16,174,864), A.D.2 (13,303,391) **(more)**

- AUTHORS BY CENTURY
- LARGEST AUTHORS
- LARGEST WORKS
- MOST FREQUENT LEMMATA
- MOST FREQUENT WORDFORMS
- LEMMATA BY CENTURY
- WORDFORMS BY CENTURY
- WORD COUNTS BY CENTURY



# FULL CORPUS STATISTICS

This page includes a summary of the data included in the TLG corpus. Information is presented in the forms of diagrams and expandable menus.

Information included under the expandable menu is linked to Text search or the TLG Canon.

# Statistical information for Homer's Iliad

## AUTHOR STATISTICS

### STATISTICS:

Full Corpus
  Author
  Lemma

HOMERUS



Or Ilias {001}

### SUMMARY

- There are 111,909 words in this author and 7,942 lemmata
- The most frequent wordforms are: αὐτάρ, Αχαιῶν, Ἔκτωρ ([more](#))
- The most frequent lemmata are: Αχαιός, -οῦ, ὁ, Τρῶες, -ῶν, οἱ, ἀνὴρ ([more](#))
- The most over-represented lemmata in this author are: Αχαιός, -οῦ, ὁ, Τρῶες, -ῶν, οἱ, ἀνὴρ ([more](#))
- The most under-represented lemmata in this author/work are: αἶ, ικνέομαι, Ἀπόλλων, -νος, ὁ ([more](#))
- There are 11 unique occurrences in this author ([less](#))

### OVER-REPRESENTED LEMMATA

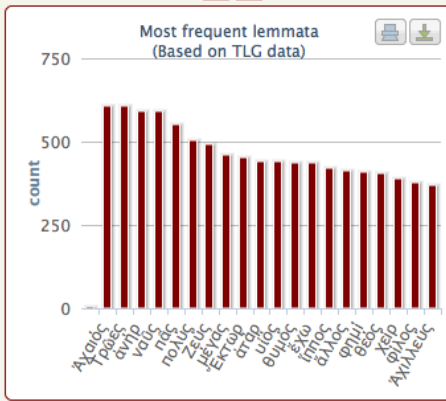
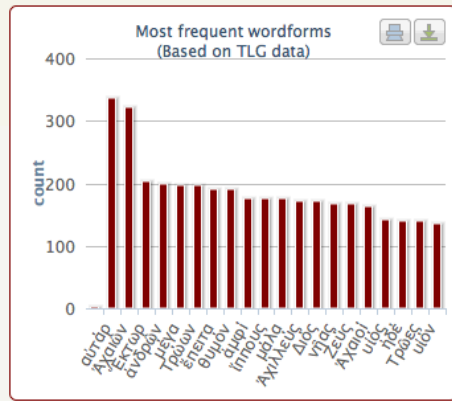
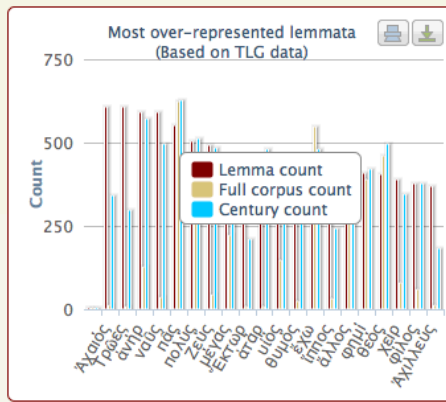
### UNDER-REPRESENTED LEMMATA

### MOST FREQUENT WORDFORMS

### MOST FREQUENT LEMMATA

### UNIQUE OCCURRENCES

Lemmata	count
λλεγηγορίδης, -ου, ὁ	1
λαστύπυλος, -ου, ὁ	1
δαίτορα, -ας, ἡ	1
ἐπίστωρ, -ορος, ὁ	1
εὐσσωρος, -ου, ὁ	1
κρειοντιάδης, -ου, ὁ	1
προθῶων, -νος, ὁ	1
προτιάων, -ονος, ὁ	1
πύρις, -ίδος, ὁ	1
φασσιάδης, -ου, ὁ	1
φιληγορίδης, -ου, ὁ	1



Click to go to text search

Click to enlarge

Obtain list of 100 most frequent lemmata

This page provides statistical information about a specific author or work.

The information is given through diagrams or expandable menus (left column).

Each word in each list is linked to Text Search.

Two icons at the bottom of each diagram allow you to enlarge the diagram or obtain a list of the most frequent lemmata in the selected author or work.

## Important notes

- There are certain lemmata which we ignore when looking at the most frequent lemmata in a subcorpus. Stop-words are ignored in N-grams and Statistics pages. Milesian numerals are also ignored in lists of statistics.
- The total number of words listed in the Summary section is the raw count which includes non-Greek words. Lemmata, wordforms and total word counts--broken down by century--do not include non-Greek words.
- Unique occurrences vs. hapaxes  
It is easy for the TLG search engine to identify lemmata that are unique to a specific author in the corpus. The statistics display lists of those lemmata as unique occurrences.

Unique occurrences of lemmata are known as “hapaxes”. We do not use the word hapax, because our corpus is not the same as the classical corpus against which hapax was defined. A word is a hapax in Homer if it was not used elsewhere in the classical literary canon. But the TLG is not limited to the classical literary canon: it also includes Homeric dictionaries, Homeric scholia, grammarians, and literary theorists, all of whom discussed the words of Homer at length. So while there are numerous Homeric hapaxes by the classical definition, only 14 lemmata in Homer are restricted to Homer in the TLG corpus.

## STATISTICS

### Exclusions

### Unique occurrences vs. hapaxes

## Over-represented and under-represented lemmata per century

Relative frequencies of lemma use, rather than raw counts, are particularly important when we are comparing usage across different centuries. Through accidents of scribal preference and cultural preservation, the quantity of preserved Greek literature varies widely by century. The corpus for A.D. 4 is around 18 million words; the corpus for 8 B.C. is 0.2 million words, which is 90 times smaller. So it is meaningless to say that there are more instances of *καλός* in 4 A.D. than B.C. 8 (17,547 vs. 402); with 90 times more words, we would expect *any* word to occur more often in iv A.D. than viii B.C.

One way of looking at this is by working out what the expected count of the word would be per century, if the word were distributed evenly across the TLG. *καλός* occurs 111,604 times in the 98 million words of the TLG. Extrapolating from that, it should occur 20,230 times in iv A.D., and 274 times in viii B.C. There are over 2,300 instances of *καλός* less than we would expect in iv A.D.; so we can say that *καλός* is underrepresented by 2300 instances for that century. There are 127 instances of *καλός* more than we would expect in viii B.C.; so we can say that *καλός* is overrepresented by 127 instances for that century.

STATISTICS

Over- and  
under-  
representati  
on

# STATISTICS:

Full Corpus  Author  Lemma

λέξις, -εως, ἡ

Input Greek

GO

From 8 B.C. to A.D. 20

-8 -7 -6 -5 -4 -3 -2 -1 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

B.C. | A.D. | All

Include Varia and Incerta

## SUMMARY

- This lemma is attested in the corpus 23,164 times
- The highest use is found in authors: EUSTATHIUS Thessalonicensis, GALENUS, ORIGENES ([more](#))
- The highest use is found in works: EUSTATHIUS Thessalonicensis, *Commentarii ad Homeri Iliadem*, EUSTATHIUS Thessalonicensis, *Commentarii ad Homeri Odysseam*, JOANNES, *Commentarium in Hermogenis librum περὶ ἰδεῶν* ([less](#))
- Its highest attestation by century is: A.D.2 3,506, varia 3,234, A.D.12 2,911, ([more](#))

## HIGHEST USE BY AUTHOR

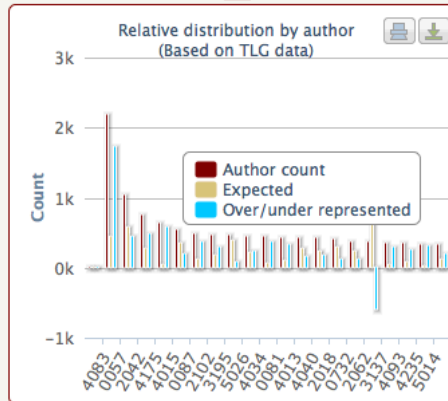
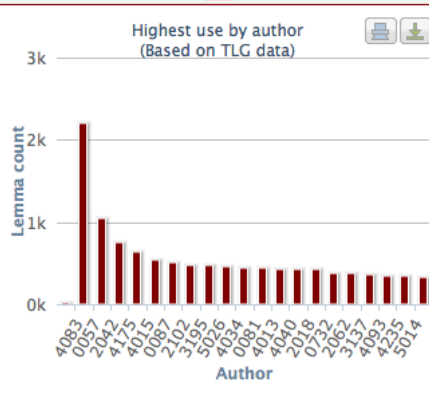
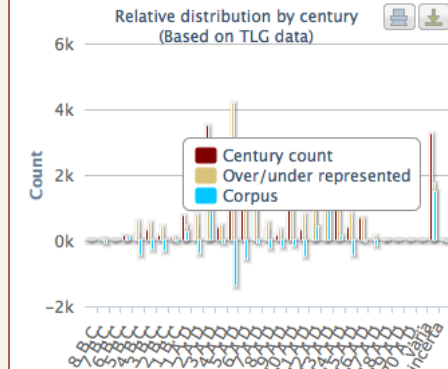
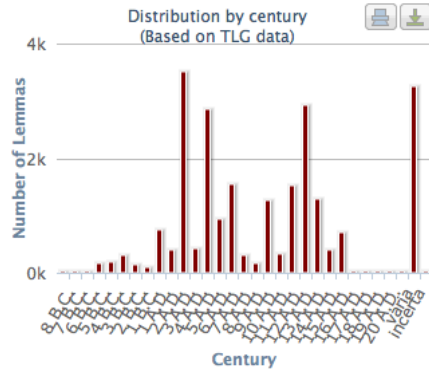
### HIGHEST USE BY WORK

Author/Work	count
EUSTATHIUS Thessalonicensis <i>Commentarii ad Homeri Iliadem</i> [4083.001]	1,472
EUSTATHIUS Thessalonicensis <i>Commentarii ad Homeri Odysseam</i> [4083.003]	637
JOANNES <i>Commentarium in Hermogenis librum περὶ ἰδεῶν</i> [4235.002]	317
ETYMOLOGICUM MAGNUM <i>Etymologicum magnum</i> [4099.001]	276
ANONYMI IN ARISTOTELIS	251

## DISTRIBUTION BY CENTURY

## RELAT. DISTRIB. BY CENTURY

## GEOGRAPHIC DISTRIBUTION



Links to the TLG Canon

# LEMMA STATISTICS

This page provides statistical information about a particular word.

The information is given through diagrams or expandable menus (left column).

Click on Geographic Distribution to view this information on the map.

## STATISTICS:

Full Corpus
  Author
  Lemma

λέξις, -εως, ἡ

GO

Input

From  to

-8 -7 -6 -5 -4 -3 -2 -1 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

B.C. | A.D. | All

Include Varia and Incerta

### SUMMARY

- This lemma is attested in the corpus 23,164 times
- The highest use is found in authors: EUSTATHIUS Thessalonicensis, GALENUS, ORIGENES ([more](#))
- The highest use is found in works: EUSTATHIUS Thessalonicensis, *Commentarii ad Homeri Iliadem*, EUSTATHIUS Thessalonicensis, *Commentarii ad Homeri Odysseam*, JOANNES, *Commentarium in Hermogenis librum περὶ ἰδεῶν* ([less](#))
- Its highest attestation by century is: A.D.2 3,506, varia 3,234, A.D.12 2,911, ([more](#))

### HIGHEST USE BY AUTHOR

### HIGHEST USE BY WORK

### DISTRIBUTION BY CENTURY

### RELAT. DISTRIB. BY CENTURY

### GEOGRAPHIC DISTRIBUTION

Location	count
Abdera	17
Achrida	24
Aegina	3
Aegyptus	26
Agrigentum [vel Acragas]	15
Alexandria	3845
Amasia	8
Amida	1
Amisus	8
Anchialus	2
Ancyra	53
Antinoupolis	35
Antiochia	506
Antiochia (Pisidiae)	3



Go back to diagrams

# LEMMA STATISTICS

Click on Geographic Distribution to view information about a specific lemma on the map.