



ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ  
Εθνικόν και Καποδιστριακόν  
Πανεπιστήμιον Αθηνών  
— ΙΔΡΥΘΕΝ ΤΟ 1837 —

## «ΑΠΟΛΛΩΝΙΣ»

*Γλωσσικοί πόροι στο ΕΚΠΑ: Υποδομή, χρήση και εργαλεία*

Διονύσης Γούτσος

ΕΚΠΑ, 01/03/2019

# Γλωσσικοί πόροι και εργαλεία

- Γλωσσικοί πόροι: σώματα κειμένων.

«εκτεταμένο δείγμα αυθεντικής χρήσης της υπό εξέταση γλώσσας, που συγκροτείται και χρησιμοποιείται ως πηγή στοιχείων για την παραγωγή ή εξέταση υποθέσεων για τη φύση της γλώσσας» (Sampson 2001: 6)

- όχι αρχείο, αποθετήριο, βιβλιοθήκη, αλλά με πρόσβαση στη γλώσσα

Γιατί;

- βασικό μεθοδολογικό εργαλείο σε όλους τους κλάδους
- εμπειρική θεμελίωση ερευνητικών υποθέσεων
- αλλαγή στην επιστημολογική θεώρηση της γλώσσας

# Σώματα κειμένων της Ελληνικής

- Τι έχουμε ποσοτικά;

Λίγο περισσότερα από 100 εκατ. λέξεις διαθέσιμα

|                     | Ιστοσελίδα                                                         | Αριθμός λέξεων |
|---------------------|--------------------------------------------------------------------|----------------|
| ΕΘΕΓ                | <a href="http://hnc.ilsp.gr">hnc.ilsp.gr</a>                       | 40.000.000     |
| ΣΕΚ                 | <a href="http://sek.edu.gr">sek.edu.gr</a>                         |                |
| ΣΚ Πύλης            | <a href="http://greeklanguage.gr">greeklanguage.gr</a>             | 7.000.000      |
| Προφορικό ΣΚ<br>ΙΝΣ | <a href="http://corpus-ins.lit.auth.gr">corpus-ins.lit.auth.gr</a> | 1.700.000      |
| Corpus of MG        | <a href="http://web-corpora.net">web-corpora.net</a>               | 35.700.000     |
| + GkWaC             | <a href="http://sketchengine.co.uk">sketchengine.co.uk</a>         | 100.000.000    |

# Σώματα κειμένων σε άλλες γλώσσες I

- Τι έχουν οι άλλοι;

Renouf (2007): English language corpora

- 1960-1970: one-million word Small Corpus: *LOB, Brown corpus*
- 1980: the multi-million word Large Corpus: super-corpus: *Bank of English, BNC*
- 1990: the 'Modern Diachronic' Corpus: dynamic, open-ended, chronological data flow: *FLOB, Frown*
- 2000: Web as corpus, cyber corpus

# Σώματα κειμένων σε άλλες γλώσσες II

- Τι έχουν οι άλλοι;

Αγγλικά:

- Oxford English Corpus: πάνω από 2,5 δισ.
- Cambridge English Corpus:
  - Written English πάνω από 1 δισ.
  - Spoken English 70 εκατ.
  - Business English 200 εκατ.
  - English for Specific Purposes
  - Academic English πάνω από 400 εκατ.
  - Learner English 200.000 γραπτά

# Σώματα κειμένων σε άλλες γλώσσες III



## corpus.byu.edu

corpora, size, queries = better resources, more insight

Overview  
**Corpora**  
Size, speed, queries  
Insight into variation

History / updates  
FAQ / questions  
Researchers  
Publications

Register  
Modify profile

Related resources  
[Full-text data](#)  
Word frequency  
Collocates  
N-grams  
WordAndPhrase  
Academic vocabulary

Problems  
Contact us

Created by Mark Davies, BYU. Overview, search types, researchers, publications, corpus-based resources.

| English                                        | # words     | language/dialect | time period | compare   |
|------------------------------------------------|-------------|------------------|-------------|-----------|
| Global Web-Based English (GloWbE)              | 1.9 billion | 20 countries     | 2012-13     |           |
| Corpus of Contemporary American English (COCA) | 450 million | American         | 1990-2012   | * * * * * |
| Corpus of Historical American English (COHA)   | 400 million | American         | 1810-2009   | * *       |
| TIME Magazine Corpus                           | 100 million | American         | 1923-2006   |           |
| Corpus of American Soap Operas                 | 100 million | American         | 2001-2012   | *         |
| British National Corpus (BYU-BNC)*             | 100 million | British          | 1980s-1993  | * *       |
| Strathy Corpus (Canada)                        | 50 million  | Canadian         | 1970s-2000s |           |
| <b>Other languages</b>                         |             |                  |             |           |
| Corpus del Español                             | 100 million | Spanish          | 1200s-1900s | *         |
| Corpus do Português                            | 45 million  | Portuguese       | 1300s-1900s |           |
| <b>N-grams</b>                                 |             |                  |             |           |
| Google Books: American English                 | 155 billion | American         | 1500s-2000s | *         |
| Google Books: British English                  | 34 billion  | British          | 1500s-2000s |           |
| Google Books: One Million Books                | 89 billion  | Am/Br            | 1500s-2000s |           |
| Google Books: Spanish                          | 45 billion  | Spanish          | 1500s-2000s |           |

# Σώματα κειμένων σε άλλες γλώσσες IV

Modern | Parallel | Old Swedish | Litteraturbanken | Kubhist | Old texts | Fler ▾

Cite Korp Svenska | English  3 

 23 of 197 corpora selected — 769.88M of 8.43G tokens

1850 1900 1950 2000

Simple Extended

also as ☐ initial part ☐ final

KWIC: hits per page: 25 ▾

- ☐ Akademiska texter (2)
- ☐ August Strindberg (2)
- ☐ Finlandssvenska texter (56)
- ☐ Skyddade korpusar (7)
- ☐ Medicinska texter (12)
- ☐ Skönlitteratur (6)
- ☐ Sociala medier (60)
- ☐ Tidningstexter (37)
- ☐ Tidskrifter (1)
- ☐ Dramawebben (demo)
- ☐ LäSBarT – Lättläst svenska och barnbokstext
- ☐ PAROLE
- ☐ Psalmboken (1937)
- ☐ SNP 78–79 (Riksdagens snabbprotokoll)
- ☐ SUC 2.0
- ☐ SUC 3.0
- ☐ SALT svenska-nederländska
- ☐ Europarl svenska
- ☐ Svenska partiprogram och valmanifest 1887–2010
- ☐ Svensk författningssamling
- ☐ Svenskt frasnät (SweFN)
- ☐ Svenska Wikipedia (oktober 2014)
- ☐ Talbanken

61,423,858 sentences in the selected corpora

Search history ▾

☐ Show word picture

# Σώματα κειμένων σε άλλες γλώσσες IV



**Eastern Armenian National Corpus**

- HOME
- OBJECTIVE
- EANC 3.0
- COMPOSITION
- ANNOTATION
- TRANSLITERATION
- SOFTWARE
- EANC TEAM
- GALLERY
- ARMENIAN TEXTS ONLINE
- RESEARCH PROGRAM
- CONTACT US



- MAIN
- HELP
- LIBRARY
- STATISTICS

## Eastern Armenian National Corpus

Eastern Armenian National Corpus (EANC) is a comprehensive linguistic database of annotated texts in Standard Eastern Armenian (SEA), the language spoken in the Republic of Armenia.

EANC is:

- a comprehensive corpus with about 110 million tokens
- a powerful search engine for making complex lexical morphological queries
- a learner's corpus including English translations for frequent tokens
- a diachronic corpus covering SEA texts from the mid-19th century to the present
- a mixed corpus consisting of both written discourse and oral discourse
- an annotated corpus with morphological and metatext tagging
- an open access corpus
- an electronic library with full access to over 100 Armenian classic titles

[Learn what's new in EANC 3.0](#)

Wordform

Lexeme

Translat

[Gram. & Lexical Attributes](#)

Advanced ▾

Distance to the next token:  
From  to  words

Wordform

Lexeme

Translat

[Gram. & Lexical Attributes](#)

Advanced ▾

Advanced Distance ▾

Search

Clear

[Specify Subcorpus](#)  
[Display Options](#)  
[Search in New Window](#)  
[Error Report](#)

© 2007-2009 Corpus Technologies

Exact ☒ Loose ☐







# Σώματα κειμένων σε άλλες γλώσσες

- Εξειδικευμένα:
  - Διαχρονικά
  - Διαλεκτικά
  - Κοινωνιολέκτων
  - Πεδίων του λόγου
  - Μαθητικών κειμένων (learner corpora)
  - Επισημειωμένα για προσωδιακά φαινόμενα
  - Πολυμεσικά

# Γλωσσικοί πόροι: Σώματα κειμένων I

- Τι θέλουμε;

# Περισσότερα, μεγαλύτερα και εξειδικευμένα Σώματα Κειμένων

## Προτάσεις/κατευθύνσεις:

- Έμφαση σε διαχρονικά/ανοιχτά σώματα κειμένων
- Αλλαγή αντίληψης για τα ψηφιακά δεδομένα
- Ανάπτυξη εθνικής στρατηγικής
- Συλλογική προσπάθεια συγκέντρωσης και διαμοιρασμού δεδομένων

# Γλωσσικοί πόροι: Σώματα κειμένων III

- Έμφαση σε διαχρονικά/ανοιχτά σώματα κειμένων



ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ  
Εθνικών και Καποδιστριακών  
Πανεπιστημίων Αθηνών

Διαχρονικό σώμα ελληνικών κειμένων του 20ού αιώνα

• Homepage • Εκδηλώσεις • Αναζήτηση • Χάρτης ιστοτόπου • Επικοινωνία

Greek Corpus 20 Ερευνητική ομάδα Συμβουλευτική επιτροπή Ημερίδα Δημοσιεύσεις Σώμα κειμένων

**Εκδηλώσεις**

Παρουσίαση στο 11ο Διεθνές Συνέδριο Ελληνικής Γλωσσολογίας  
Στο 11ο Διεθνές Συνέδριο Ελληνικής Γλωσσολογίας παρουσιάστηκε η ανακοίνωση "Πρόσφατη γλωσσική..."

Πρόγραμμα ημερίδας για τα διαχρονικά σώματα κειμένων  
Ο Τομέας Γλωσσολογίας του Τμήματος Φιλολογίας διοργανώνει στις 28 Ιουνίου ημερίδα για τη συγκρότηση...

[Περισσότερα »](#)

**Greek Corpus 20**

Το πρόγραμμα **Διαχρονικό σώμα ελληνικών κειμένων του 20ού αιώνα**, που χρηματοδοτείται στο πλαίσιο της πράξης ΑΡΙΣΤΕΙΑ από το Ευρωπαϊκό Κοινωνικό Ταμείο και από Εθνικούς Πόρους (Γενική Γραμματεία Έρευνας και Τεχνολογίας), περιλαμβάνει τη συγκρότηση και ανάλυση ενός διαχρονικού σώματος κειμένων (corpus) της Ελληνικής του 20ού αιώνα. Η Ελληνική, σε αντίθεση με άλλες γλώσσες, δεν έχει επωφεληθεί επαρκώς έως τώρα από τις τεράστιες προόδους στο ευρύτερο ερευνητικό πεδίο της ανάπτυξης και αξιοποίησης γλωσσικών σωμάτων κειμένων. Το πρόγραμμα στοχεύει να καλύψει αυτό το κενό με την ανάπτυξη ενός σώματος κειμένων της Ελληνικής 20 εκατ. λέξεων, που θα ενσωματωθεί στο προϋπάρχον Σώμα Ελληνικών Κειμένων ([ΣΕΚ](#), μέγεθος: 30 εκατ. λέξεις, περιλαμβάνει κείμενα από το 1990 έως το 2010).

Ευχαριστίες:

ΜΝΗΜΕΣ

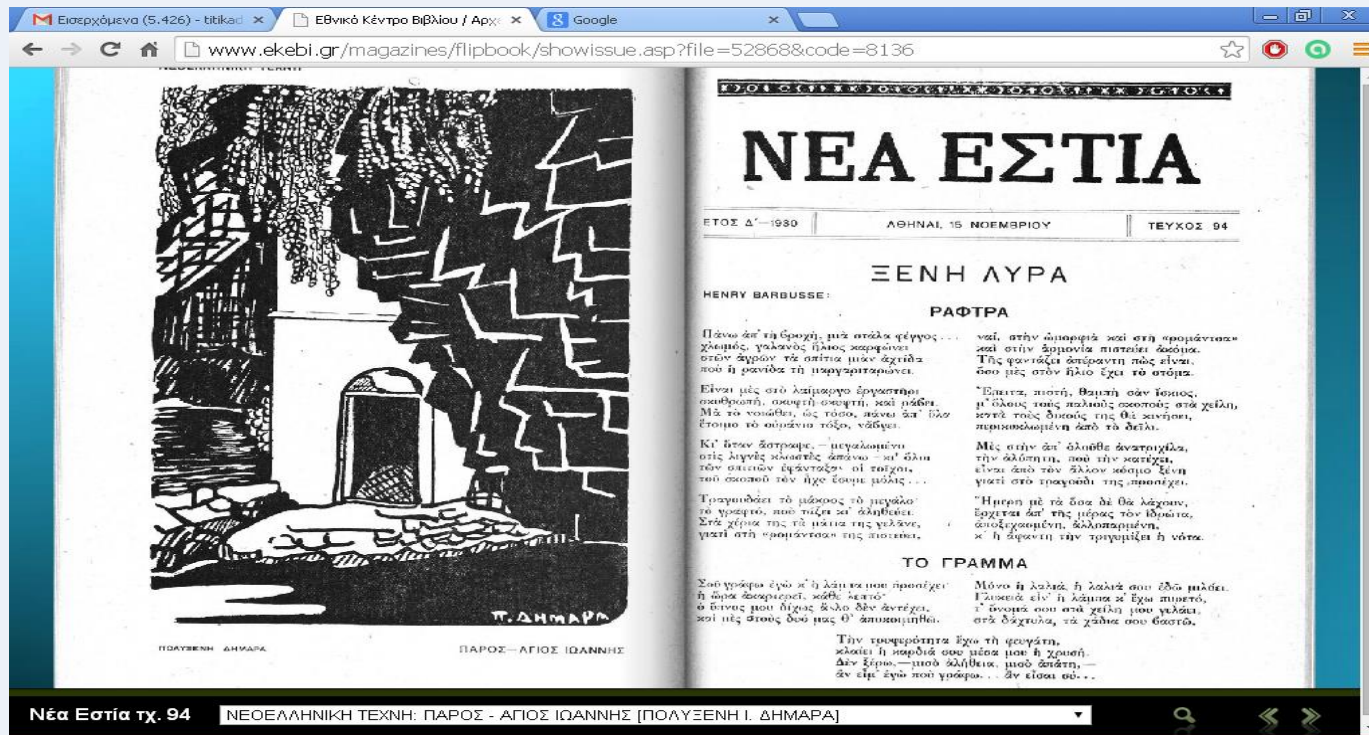
Ευρωπαϊκή Ένωση  
Ευρωπαϊκό Κοινωνικό Ταμείο

ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ  
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ  
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ  
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

ΕΣΠΑ 2007-2013  
Ευρωπαϊκό Κοινωνικό Ταμείο

# Γλωσσικοί πόροι: Σώματα κειμένων IV

- Αλλαγή αντίληψης για τα ψηφιακά δεδομένα



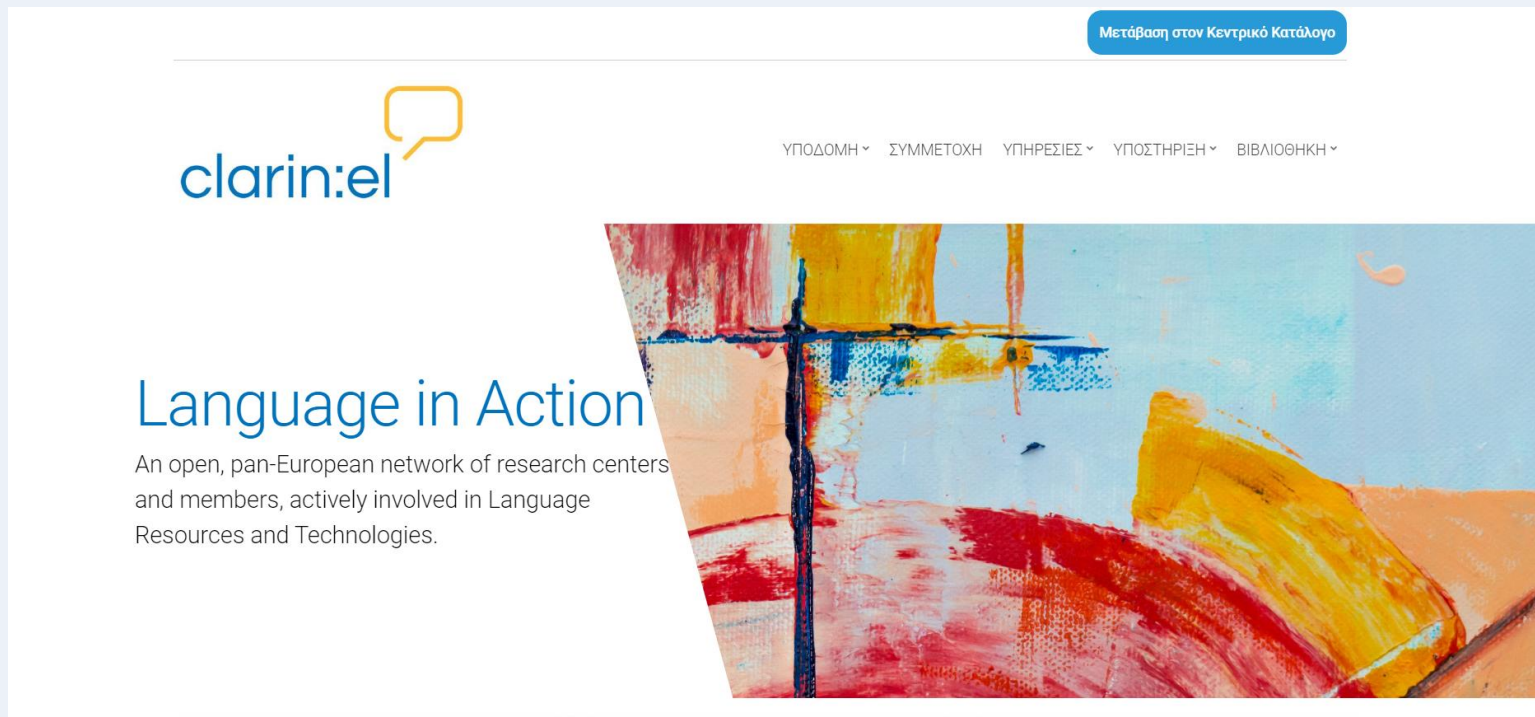
- Ανάπτυξη εθνικής στρατηγικής





# Γλωσσικοί πόροι: Σώματα κειμένων II

- Συλλογική προσπάθεια συγκέντρωσης και διαμοιρασμού δεδομένων





# Γλωσσικά εργαλεία

Βασικά εργαλεία:

- Διαχωρισμός & εντοπισμός λεκτικών μονάδων (segmentation / tokenization)
- Λημματοποίηση (lemmatization)
- Επισημείωση με μέρος του λόγου (PoS tagging)
- Μορφοσυντακτική επισημείωση (below-PoS tagging)
- Συντακτική ανάλυση (constituency/dependency parsing)
- Σημασιολογική επισημείωση (semantic annotation)
- Κειμενική επισημείωση (text annotation)
- Στοίχιση δί-/πολύγλωσσων κειμένων (alignment)

Εξειδικευμένα εργαλεία/Οπτικοποίηση

# Γλωσσικά εργαλεία της Ελληνικής

- Τι έχουμε;



National Centre for Scientific Research "Demokritos"



Member of  
**W3C**

## Institute of Informatics and Telecommunications

[Home](#)
[Location](#)
[Contact](#)

Ελληνικά

### The Institute

- News
- ▶ About IIT
- People
- Reports
- ▶ Cooperations
- Jobs

### Research

- Laboratories
- Publications
- R&D Projects
- Awards

### Education

- Overview
- Joint programs (with Greek Universities)
- Joint PhD programs / Research Collaborations
- Lecture Series
- Training
- Schools

Home > D-SCRIBE - System for Digitisation and Processing of Greek Manuscripts

## D-SCRIBE - System for Digitisation and Processing of Greek Manuscripts

**Name:** System for Digitisation and Processing of Greek Manuscripts

**Acronym:** D-SCRIBE

The objective of this project is the development of an integrated system for management and processing of old Greek manuscripts. The system includes tools for the indexing and recognition of texts from the early Christian period, as well as educational software for training of paleographers in reading and interpretation of the old manuscripts.

**Programme:** GSRT/Ministry of Development

**Project type:** EPAN/EPAN Consortium

Dates

**Start to End date:** 01/09/2003 - 28/02/2006

**Duration:** 30 months

NCSR Related

**NCSR role:** Partner

# Γλωσσικά εργαλεία σε άλλες γλώσσες

## I

- Τι έχουν οι άλλοι;

**Diachronic corpus DiaCORIS - Corpus query form**

|                                                                                                                                                                 |                                                                                                                                                                                             |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>User Authentication</b><br><br>DiaCORIS access is now free for research purposes<br>(Please, read the footnote carefully).                                   | <b>Query</b><br><br>( <a href="#">Query Language Help</a> ).<br>"Partenone"<br><br>Section <input type="text" value="All"/> SubCorpus <input type="text" value="All"/>                      |
| <b>Concordance Options</b><br><br>Show <input type="radio"/> 30 <input type="radio"/> 100 <input type="radio"/> 300 <input checked="" type="radio"/> all lines. | Sort position: <input type="text" value="1st left"/>                                                                                                                                        |
| <b>Collocations</b><br><br>Get Collocates? <input type="radio"/> NO! <input checked="" type="radio"/> Yes.                                                      | Sort using<br><input checked="" type="radio"/> Log-Likelihood Ratio.<br><input type="radio"/> Mutual Information.<br><input type="radio"/> T-score.<br><input type="radio"/> Raw frequency. |

Interrogando il corpus DiaCORIS tramite questa procedura, l'utente dichiara e accetta che l'interrogazione è volta unicamente a scopi di ricerca scientifica e che non ne sarà tratto alcun beneficio economico. L'accesso al corpus è concesso esclusivamente per scopi di ricerca scientifica.

Designed by Fabio Tamburini.

# Γλωσσικά εργαλεία σε άλλες γλώσσες II

## Corpus query result

Query: "Partenone"

**Sorted: 0**

**Number of concordances: 19 / 19**

|                          |                                |             |                               |
|--------------------------|--------------------------------|-------------|-------------------------------|
| <a href="#">context:</a> | erie , come l ' opistodomo del | <Partenone> | , nelle quali i turchi che pa |
| <a href="#">context:</a> | ? - , per es . , le linee del  | <Partenone> | , i quadri di Raffaello , i v |
| <a href="#">context:</a> | un greco nato nell ' ombra del | <Partenone> | . Gli uomini distratti , miop |
| <a href="#">context:</a> | in quei secoli ha costruito il | <Partenone> | e Nôtre Dame ; ha scritto i   |
| <a href="#">context:</a> | canti degli uomini nuovi . Il  | <Partenone> | avrebbe schiacciato il nostro |
| <a href="#">context:</a> | e rovine dell ' Acropoli , del | <Partenone> | , fuori per lo Stadio , ricer |
| <a href="#">context:</a> | lture più belle del fregio del | <Partenone> | : all ' ingresso della Iliade |
| <a href="#">context:</a> | non toccheranno certamente il  | <Partenone> | . Non hanno nessuna ragione p |
| <a href="#">context:</a> | sero sarebbe una sventura : il | <Partenone> | è tuttavia un segno di opere  |
| <a href="#">context:</a> | lumina con il Suo splendore il | <Partenone> | mirabile , il Campidoglio aus |
| <a href="#">context:</a> | nia . Dal colonnato dorico del | <Partenone> | al colonnato d ' acciaio dell |
| <a href="#">context:</a> | to con la biacca una specie di | <Partenone> | : una sagoma puerile di colon |
| <a href="#">context:</a> | ntano la notte fra i marmi del | <Partenone> | . ( L ' uccello sacro a Palla |
| <a href="#">context:</a> | o pensare che gli ateniesi del | <Partenone> | , se fossero vissuti nei temp |
| <a href="#">context:</a> | all ' ombra delle colonne del  | <Partenone> | . La mitologia accompagnò per |
| <a href="#">context:</a> | al Santo Sepolcro , davanti al | <Partenone> | , eccetera . Attendo da uomin |
| <a href="#">context:</a> | ano più luminosi e sublimi del | <Partenone> | , e qualsiasi costruzione ci  |
| <a href="#">context:</a> | ere di Fidia , lo scultore del | <Partenone> | , il massimo artista della Gr |
| <a href="#">context:</a> | ' è l ' antica Grecia col suo  | <Partenone> | e la sua scoperta della Demo  |

**Collocates** (Corpus size = 28320192 tokens)

| Collocate | Log-Likelihood Ratio | f(node,colloc) | f(colloc) |
|-----------|----------------------|----------------|-----------|
| del       | 91.750000            | 12             | 219209    |
| ,         | 41.000000            | 12             | 1888152   |

## Γλωσσικά εργαλεία σε άλλες γλώσσες III

Wmatrix2

http://ucrel.lancs.ac.uk/wmatrix2.html

Wmatrix BBC LU ISS MOTD Paul Rayson's Home Page iGoogle Cation Village Amazon.co.uk ICAME Apple Yahoo! Google Maps YouTube

## View collocation list

You are logged in as: paul

[Tagging > Tag Wizard...](#) | 
 [My Tag Wizard...](#) | 
 [Domain Tag Wizard...](#) | 
 [Load file without tagging...](#) | 
 [Tracker tag wizard...](#) | 
 [Folders > My folders](#) | 
 [Details](#) | 
 [Create...](#) | 
 [Delete...](#) | 
 [Archive...](#) | 
 [Extract...](#) | 
 [Join...](#) | 
 [Share...](#) | 
 [Empty TRASH](#) | 
 [Emode > Tag wizard...](#) | 
 [View Lexicon & MWEs & Merged Lexicon & Merged MWEs & Reg table](#) | 
 [Admin...](#) | 
 [Merge Lexicon](#) | 
 [Merge MWEs](#) | 
 [Options > Switch to Simple Interface](#) | 
 [Edit user options...](#) | 
 [Admin](#) | 
 [Help > Contents](#) | 
 [Availability](#) | 
 [Tagsets: POS & Semantic](#) | 
 [USAS: Lexicon & MWEs & Context rules](#) | 
 [Updates](#) | 
 [Feedback](#) | 
 [You are here > My folders > LabourManifesto2005](#) | 
[File details](#) | 
[Summary sheet](#)

Save this table as a [tab-delimited file](#)

Change statistic:

Search this list:  
Enter the word you wish to search for here:  
   
 (you can also search for part of a word;  
enter '.', or leave blank for complete list)

|    | MI    | T-score | Collocation          |
|----|-------|---------|----------------------|
| 1  | 12.27 | 1.73    | Sure Start           |
| 2  | 11.85 | 2.00    | Tony Blair           |
| 3  | 11.85 | 2.00    | Preface Tony         |
| 4  | 11.85 | 2.00    | Preface Blair        |
| 5  | 11.85 | 2.00    | ID cards             |
| 6  | 11.85 | 1.73    | maternity paternity  |
| 7  | 11.85 | 1.73    | Going instead        |
| 8  | 11.85 | 1.73    | Common Policy        |
| 9  | 11.85 | 1.73    | Prime Minister       |
| 10 | 11.85 | 1.73    | Penalty Disorder     |
| 11 | 11.53 | 2.00    | Royal Mail           |
| 12 | 11.53 | 2.00    | electronic tagging   |
| 13 | 11.53 | 2.00    | Secure today         |
| 14 | 11.44 | 1.73    | firm foundations     |
| 15 | 11.27 | 2.45    | Northern Ireland     |
| 16 | 11.27 | 2.00    | law?abiding majority |
| 17 | 11.12 | 1.73    | European Union       |

# Γλωσσικά εργαλεία σε άλλες γλώσσες IV

The screenshot displays the EBO Trained software interface. The main window shows a text document with various words highlighted in different colors (yellow, green, blue, red) to represent different linguistic features. The interface includes several toolbars and panels:

- Top Bar:** Contains the text "undo/redo" and "F-score".
- Left Panel:** Includes a "Setup" section with a file path "/Users/.../EBO Trained" and a "text toolbar" with icons for undo, redo, and other text editing functions.
- Right Panel:** Features a "Sidebar" with a "Display" section showing "Variants (357 tokens)", "Normalised (0 tokens)", and "Not variants (4448 tokens)". Below this is a "Types List" showing various linguistic types like "themselves (21)", "som (19)", "being (16)", etc.
- Bottom Panel:** Includes a "Step Complete" section with a progress bar and a "Threshold (%)" set to 50.
- Text Area:** The central part of the window displays a large block of text, likely a historical document, with words like "LONDON", "William", "Bourn", "Cornhill", "Honor", "Credit", "Repute", "Honor", "Credit", "Repute", "Honor", "Credit", "Repute", etc., highlighted in different colors.
- Context Menu:** A right-click context menu is visible over the text area, showing options like "Normalise instance", "Normalise all", "Mark instance as Not variant", "Select type", and "Mark all as Not variant".



## Γλωσσικά εργαλεία σε άλλες γλώσσες V

Historical Thesaurus of English

historicalthesaurus.arts.gla.ac.uk/category/?id=37860&h=Y

## Equus caballus/horse

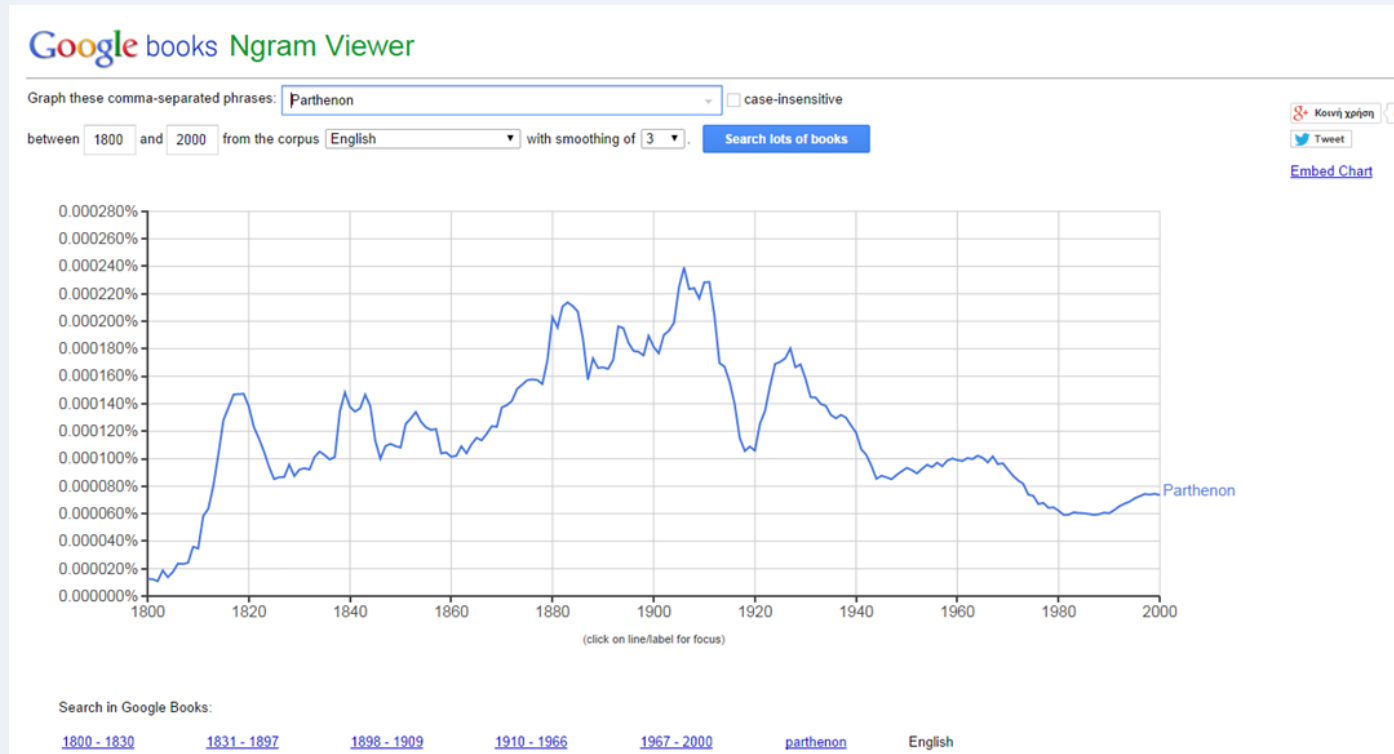
↑ Up the hierarchy to: [The world](#) > [Animals](#) > [Mammals](#) > [Group Ungulata \(hoofed\)](#) > [Family Equidae \(general equines\)](#)

↔ Other parts of speech: [aj \(Of/pertaining to a horse/horses\)](#) [av \(Like/as becomes a horse\)](#)

There are 41 words at this level:

|                                                  |                      |
|--------------------------------------------------|----------------------|
| <b>friphengest</b> OE                            | <a href="#">Cite</a> |
| <b>mearh</b> OE                                  | <a href="#">OED</a>  |
| <b>wicg</b> OE                                   | <a href="#">OED</a>  |
| <b>stot</b> < <b>stot</b> OE–c1440               | <a href="#">OED</a>  |
| <b>blonk</b> < <b>blanca</b> OE–1535             | <a href="#">OED</a>  |
| <b>horse</b> < <b>hors</b> OE–                   | <a href="#">OED</a>  |
| <b>caple/capul</b> c1290–1819 also dial. & poet. | <a href="#">OED</a>  |
| <b>foal</b> a1300–1513                           | <a href="#">OED</a>  |
| <b>rouncy</b> c1300–a1529 + 1875– also arch.     | <a href="#">OED</a>  |
| <b>scot</b> c1386(2) obs. exc. dial.             | <a href="#">OED</a>  |
| <b>brock</b> c1386 + 1586 + 1847/78 Dict.        | <a href="#">OED</a>  |

# Γλωσσικά εργαλεία σε άλλες γλώσσες VI





# Γλωσσικά εργαλεία

- Τι θέλουμε;
- Διαθέσιμα εργαλεία σε φιλική μορφή για τους χρήστες
- Ολοκληρωμένες πλατφόρμες γλωσσικής ανάλυσης
- Εργαλεία για παλαιότερες και διαλεκτικές ποικιλίες Ελληνικής  
π.χ. πολυτονικό OCR

# Γλωσσικά εργαλεία

- Προσανατολισμός στους χρήστες των γλωσσικών εργαλείων
- Οπτικοποίηση γλωσσολογικών ευρημάτων

# ΕΥΧΑΡΙΣΤΩ!