

# Η συμμετοχή του Αριστοτελείου Πανεπιστημίου στην εθνική υποδομή CLARIN EL

Τιτίκα Δημητρούλια

# Πολυμορφία γλωσσικών και κειμενικών πόρων

- Τύπος πόρων (γλωσσάρια, σώματα κειμένων, κείμενα που μπορούν να γίνουν σώματα κειμένων, περιοδικά, δημοσιεύσεις κ.ά.)
- Τύπος αρχείων (word, excel, pdf, αρχεία εικόνας, html-xml, ενίοτε έντυπα)

# Πού βρίσκονται

- Κλειστοί σε σέρβερς
- Ελεύθερα διαθέσιμοι αλλά χωρίς προβολή και επομένως με ελάχιστη πρακτική χρήση
- Ελεύθερα διαθέσιμοι και χρησιμοποιούμενοι από μια ειδική κοινότητα ερευνητών

<http://echo.frl.auth.gr/emel1/index.php/el/>

ΑΡΧΙΚΗ ΕΡΕΥΝΗΤΙΚΑ ΠΡΟΓΡΑΜΜΑΤΑ ΕΠΙΣΤΗΜΟΝΙΚΕΣ ΣΥΝΕΡΓΑΣΙΕΣ ΜΕΛΗ ▾ ΝΕΑ ΕΠΙΚΟΙΝΩΝΙΑ



Τμήμα Γαλλικής Γλώσσας & Φιλολογίας  
**Εργαστήριο Μετάφρασης &  
Επεξεργασίας του Λόγου**  
Τομέας Μετάφρασης

## Αρχική

Το Εργαστήριο Μετάφρασης και Επεξεργασίας του Λόγου (EMEL), το οποίο ιδρύθηκε με το ΦΕΚ 1625/2-11-2004, απαρτίζεται από τις παρακάτω μονάδες:

- Μεταφρασεολογία και Γλώσσες Ειδίκευσης
- Διαπολιτισμική Μεσολάβηση και Μετάφραση
- Αυτόματη Επεξεργασία Φυσικών Γλωσσών.

Το EMEL εξυπηρετεί εκπαιδευτικές και ερευνητικές ανάγκες στα επιστημονικά πεδία:

- Θεωρία και Κριτική της Μετάφρασης
- Διαπολιτισμική Μεσολάβηση και Μετάφραση
- Μηχανική Μετάφραση
- Επεξεργασία και Ανάλυση Γραπτού και Προφορικού Λόγου
- Επεξεργασία και Μετάδοση της Πληροφορίας

# ΕΜΕΛ (σέρβερ)

- ηλεκτρονικά μορφολογικά λεξικά απλών και πολυλεκτικών σύνθετων λεξικών μονάδων κοινού λεξιλογίου και επιστημονικών όρων
- πεπερασμένα αυτόματα περιγραφής ονοματικών συνόλων,
- κλιτικούς γράφους απλών και πολυλεκτικών σύνθετων λεξικών μονάδων
- πεπερασμένα αυτόματα εξαγωγής ονοματικών οντοτήτων
- διγλωσση βάση δεδομένων με αθλητικούς και χρηματιστηριακούς όρους.

<http://niobe.frl.auth.gr/corpora>

FREL

Παράλληλο σώμα λογοτεχνικών κειμένων

Αρχική

Corpus

Δημοσιεύσεις

Βιβλιογραφία

Ημερίδα

Νέα

Επικοινωνία

Προβληματική

Μεταφράσεις

Αναζήτηση

### Αναζήτηση

Κείμενο προς αναζήτηση

Αναζήτηση

### Αποτελέσματα αναζήτησης

Αποτελέσματα αναζήτησης για το κείμενο [ παράθυρο ].

Βρέθηκαν 58 αποτελέσματα.

#	Πηγή	Κείμενο στα Ελληνικά	Κείμενο στα Γαλλικά
1.	ALFRED JARRY Υμνου Τύραννος	Εσύ, όμως, για δεν πας στο <b>παράθυρο</b> να δεις αν έρχονται οι καλεσμένοι μας.;	Mère Ubu, va donc voir à la fenêtre si nos invités arrivent.
2.	Fourmier, Alain - Ο μεγάλος Μολν	Κάποιος είχε ανοίξει το <b>παράθυρο</b> και είχαν στερεώσει στο άνοιγμα δυο πράσινα Βενετσιάνικα φανάρια.	Quelqu'un avait ouvert la fenêtre et l'on avait attaché dans l'embrasure deux lanternes vénitienes vertes.
3.	Fourmier, Alain - Ο μεγάλος Μολν	Το ανοιχτό <b>παράθυρο</b> χτύπησε από τον αέρα.	Le vent fit battre la fenêtre ouverte.
4.	Fourmier, Alain - Ο μεγάλος Μολν	Διάβαζα καθισμένος σε μια παλιά πλάστιγγα, δίπλα σε ένα <b>παράθυρο</b> που έβλεπε στον κήπο.	et je lisais assis sur une vieille bascule, auprès d'une fenêtre qui donnait sur le jardin.
5.	Fourmier, Alain - Ο μεγάλος Μολν	με κίνηση πολύ αργή, που σιγά σιγά επιβραδύνονταν περισσότερο, ώπου σταμάτησε κάτω από το <b>παράθυρο</b> της	Cela ralentit le pas et finalement vint s'arrêter sous la fenêtre de la salle à manger qui donnait sur la route, mais qui



AUTH Research Comitee



Laboratory of Translation and  
Speech Processing (EMEL)  
School of French


http://www.populismus.gr/category/clipping-en/printed-greek-press-clipping/

Twitter Facebook About Orientation Publicity Contact English

**POPULISMUS** POPULIST DISCOURSE AND DEMOCRACY Home Activities Media Monitoring Food For Thought

## Printed Greek Press Clipping

MAR 10 **Το στίγμα** Printed Greek Press Clipping 0



Download PDF

Tags: Ελληνική Κυβέρνηση Ευρώπη Εφημερίδα Των Συντακτών Λαός

SEARCH Search...

RECENT POSTS

Book Review Symposium on Nadia Urbinati's Democracy Disfigured: Opinion, Truth, and the People, Lasse Thomassen, Jan Biba, Jeffrey Edward Green and John P. McCormick

The Problem with Podemos, Alberto Garzón, Jacobin

What has gone wrong for Podemos?, Miguel-Anxo Murado, TheGuardian

# Ποιον ενδιαφέρουν

- ◉ Ως μαζικά δεδομένα: όλους!
- ◉ Τον καθένα ξεχωριστά, αναλόγως με τις ανάγκες του
- ◉ Το ίδιο υλικό μπορεί να χρησιμοποιηθεί για εντελώς διαφορετικές εφαρμογές (γλωσσικής ανάλυσης, διδακτικής, μηχανικής μετάφρασης, ορολογίας, ανάλυσης λόγου σε όλα τα πεδία κ.λπ.)



# Κατηγορίες γλωσσικών πόρων

- ◉ Σώματα κειμένων, προφορικά και γραπτά
- ◉ Γλωσσάρια
- ◉ Αρχεία
- ◉ Επιστημονικές δημοσιεύσεις
- ◉ Διπλωματικές εργασίες
- ◉ Περιοδικά
- ◉ Φωτογραφίες με λεζάντες
- ◉ Βίντεο

http://ins.web.auth.gr/index.php?lang=el&Itemid=165



ΑΡΙΣΤΟΤΕΛΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΟΝΙΚΗΣ  
ΙΝΣΤΙΤΟΥΤΟ ΝΕΟΕΛΛΗΝΙΚΩΝ ΣΠΟΥΔΩΝ  
(ΙΔΡΥΜΑ ΜΑΝΟΛΗ ΤΡΙΑΝΤΑΦΥΛΛΙΔΗ)



M. Τριανταφυλλίδης

Το Ινστιτούτο

Ερευνητικά προγράμματα

**Αρχεία**

Εκδηλώσεις & Δραστηριότητες

Επικοινωνία

e-Shop



#### Οι εκδόσεις μας

Απαντα του Μανόλη Τριανταφυλλίδη

Άλλα έργα του Μανόλη Τριανταφυλλίδη

Λεξικά της ελληνικής γλώσσας

Περιγραφή της ελληνικής γλώσσας

Ιστορική γλωσσολογία και ιστορία της ελληνικής γλώσσας

Γενική γλωσσολογία

Μελέτες για την ελληνική γλώσσα

Διδασκαλία της ελληνικής ως μητρικής γλώσσας

Διδασκαλία της ελληνικής ως δεύτερης/ξένης γλώσσας

Η νεοελληνική γραμματική σε ξένες γλώσσες

Αρχαιογνωσία και αρχαιογλωσσία στη Μέση Εκπαίδευση

Αρχική > Αρχεία > Corpus Προφορικού Λόγου

### Corpus Προφορικού Λόγου

Το Corpus Προφορικού Λόγου του Ινστιτούτου Νεοελληνικών Σπουδών (Ιδρύματος Μ. Τριανταφυλλίδη) εμπίπτει στο ερευνητικό πρόγραμμα "Γλωσσική Διεπίδραση και Ανάλυση Συνομιλίας" του Ινστιτούτου με επιστημονικά υπεύθυνα την καθηγήτρια Θ.-Σ. Παυλίδου. Αναπτύχθηκε πρωταρχικά με στόχο την ποιοτική ανάλυση της γλώσσας και της γλωσσικής επικοινωνίας, ειδικότερα από τη σκοπιά της Ανάλυσης Συνομιλίας, γεγονός από το οποίο απορρέουν και τα ιδιαίτερα χαρακτηριστικά του. Μέρος όμως του Corpus μπορεί να χρησιμοποιηθεί on line και για ποσοτική ανάλυση.

#### Χαρακτηριστικά του Corpus Προφορικού Λόγου του Ινστιτούτου

Η αναγκαιότητα ύπαρξης ενός corpus προφορικού λόγου για την ελληνική γλώσσα προκύπτει καταρχήν από τη σημασία που αποδίδει η σύγχρονη γλωσσολογία στον προφορικό λόγο γενικά έναντι του γραπτού. Με βάση όμως και τα πορίσματα της κοινωνιογλωσσολογίας, η μελέτη του προφορικού λόγου οφείλει να στηρίζεται σε γλωσσικό υλικό που προέρχεται από φυσικές περιστάσεις επικοινωνίας, τέτοιες δηλαδή που επιτρέπουν την αυθόρμητη και αβίαστη παραγωγή του. Κατά συνέπεια, η δημιουργία ενός corpus προφορικού λόγου θέτει στις/στους ερευνητριες/τές μια σειρά από προκλήσεις (από τη διασφάλιση της φυσικότητας του υλικού, την υπέρβαση δηλαδή του λεγόμενου 'παράδοξου του παρατηρητή', έως τη συγκατάθεση των συμμετεχόντων ατόμων για μαγνητοφώνηση, βιντεοσκόπηση κτλ.), οι οποίες δεν υφίστανται όταν πρόκειται για corpora γραπτού λόγου αποτελούμενα από γραπτά και δημοσιευμένα κείμενα.

Το Corpus Προφορικού Λόγου του Ινστιτούτου αναπτύχθηκε πρωταρχικά με στόχο την

Q αναζήτηση...

Οι τελευταίες μας εκδόσεις

**Μανόλης Σκλάβος,  
Της Κρήτης ο  
χαλασμός**

Τ. Μαρκομιχελάκη (επιμ.)

**Αναζήτηση και  
νόστος του  
Οδυσσέα**

Δ. Ν. Μαρωνίτης

**Ημερίδα με θέμα «Η  
εκδοτική των  
κειμένων της  
νεοελληνικής  
γραμματείας»**

# Pedagogic corpus



Τμήμα Αγγλικής  
Γλώσσας και Φιλολογίας

Διομηγεία  
Κινηματογράφος  
Αμερικανική Μετάφραση



ΑΡΙΣΤΟΤΕΛΕΙΟ  
ΠΑΝΕΠΙΣΤΗΜΙΟ  
ΘΕΣΣΑΛΟΝΙΚΗΣ

## Pedagogic Corpus

The **pedagogic corpora** or "**corpora of language teaching coursebooks**" (as Gabrielatos, 2005, calls them) was created by *Thomas Zapounidis* as part of his MA thesis under the supervision of the Assistant Prof. *Marina Mattheoudakis*. The corpora include:

1. all the material (Student Book and Workbook) of the *Funway* series - the old coursebooks for the teaching of English at the 4th, 5th and 6th grades of primary school
2. all the material (Student Book and Workbook) of the New English coursebooks for the 4th, 5th and 6th grades of primary school
3. all the material (Student Book and Workbook of both the advanced and beginner levels of each grade) used at the three grades of Junior High School.

These corpora provide important information about the language of the English coursebooks as taught in Greek State Schools in both the Primary and Secondary Sectors. They are meant to be utilised for research purposes with a number of available software such as Paul Nation's *RANGE* and *FREQUENCY*, Laurence Anthony's *AntWordProfiler*, *MONOCONC* and so on. In particular, the corpora could be used:

- to examine the **size** of vocabulary in each unit, coursebook or series of coursebooks (unveiling potential lexical overburden in certain units, coursebooks)
- to examine the **frequency** of lexical items in each unit, coursebook or whole series of coursebooks (in an attempt to decide which are the high frequency/ low frequency words)
- to examine and compare the **degree of recycling** of lexical items between units or between coursebooks of the same series, or between these and other coursebooks
- to examine the *different meanings* of words in **concordances**
- to investigate the number of **phrases** and **collocates** taught
- to **compare** the old (*Funway* series) and new coursebooks in the primary sector
- to **compare** the quantity and quality of the language taught in the **primary** sector with the corresponding data from the **secondary** sector and determine the degree of continuity
- to compare the **quantity** and **quality** of the language taught in the coursebooks with that of other coursebooks in the market
- to compare the **quality** of the language taught in the coursebooks with that of a native corpus
- to compare the language of the *student book* with that of the corresponding *workbook* or *audio material* and decide on their suitability and relevance
- by teachers as online reference tools indicating where they have encountered certain lexical items and whether these items will be met again in the other coursebooks (within the same series or the next sector)
- by coursebook materials in evaluating the material
- by secondary sector learners in **data driven learning** (learners as investigators)

The above suggestions are not exhaustive as researchers will utilize these data in a variety of other ways to suit their own research.



skype



Maria Efstathiadi  
είναι συνδεδεμένη

# Σώματα κειμένων μαθητών και άλλα

- ◉ Τμήμα Φιλολογίας
- ◉ Τμήμα Γαλλικής Γλώσσας και Φιλολογίας
- ◉ ΕΣΚΕΙΜΑΘ ?
- ◉ Σώμα προφορικού λόγου με διασφαλισμένα δικαιώματα (συνεντεύξεις με άξονα τον χώρο, υπό επεξεργασία, Τμήμα Γαλλικής Γλώσσας και Φιλολογίας)

# Σώματα κειμένων πολιτικού λόγου / twitter (Τμήμα Πολιτικών Επιστημών)

Twitter Facebook About Orientation Publicity Contact English

**POPULISMUS** POPULIST DISCOURSE AND DEMOCRACY Home Activities Media Monitoring Food For Thought

## Printed Greek Press Clipping

MAR 10 **Το στίγμα**  
Printed Greek Press Clipping

0



Download PDF

Tags: Ελληνική Κυβέρνηση Ευρώπη Εφημερίδα Των Συντακτών Λαός

SEARCH

Search... Search

RECENT POSTS

Book Review Symposium on Nadia Urbinati's Democracy Disfigured: Opinion, Truth, and the People, Lasse Thomassen, Jan Biba, Jeffrey Edward Green and John P. McCormick

The Problem with Podemos, Alberto Garzón, Jacobin

What has gone wrong for Podemos?, Miguel-Anxo Murado, TheGuardian

www.populismus.gr/clipping-en/populism-in-world-media-en/what-has-gone-wrong-for-podemos-miguel-anxo-murado-the-guardian/

# Γλωσσάρια

- ◉ Γλωσσολογικοί όροι
- ◉ Ιατρικοί όροι
- ◉ Παρασιτολογία
- ◉ Σημασιολογικός Ιστός
- ◉ Θεολογία
- ◉ Χρηματιστηριακοί όροι
- ◉ Αθλητισμός
- ◉ ΕΙΝΑΙ ΒΕΒΑΙΟ 'ΟΤΙ ΥΠΑΡΧΟΥΝ ΚΑΙ ΠΟΛΛΑ  
ΆΛΛΑ ΚΛΕΙΣΤΑ Η ΑΦΑΝΗ

# ΑΡΧΕΙΑ

- ◉ ΙΚΕΕ: Ιδρυματικό Καταθετήριο Επιστημονικών Εργασιών+γκρίζα βιβλιογραφία
- ◉ Αρχείο νεοελληνικής λογοτεχνίας
- ◉ Αναγνωστικά
- ◉ Θεατρικές κριτικές

# Περιοδικά – Συλλογές Περιοδικών (ενδεικτικά)

- ◉ *Fragmenta Hellenoslavica*
- ◉ ΚΟΣΜΟΣ
- ◉ Πολιτισμός και Έρευνα
- ◉ *iNTERCULTURAL tRANSLATION iNTERSEMIOTIC*
- ◉ Σύνθεσις
- ◉ *Studying Humour - International Journal*
- ◉ ΥΔΡΟΤΕΧΝΙΚΑ
- ◉ BYZANTINA
- ◉ Παναθήναια



# Άλλοι πόροι

- ◉ Λογοτεχνικά κείμενα
- ◉ Άρθρα εφημερίδων για την αρχαιολογία 1832-1932
- ◉ Άρθρα εφημερίδων 1800-2000
- ◉ Αρχείο θεατρικών προγραμμάτων
- ◉ Ψηφιακή συλλογή Παύλου Καλλιγιά (νομικά κείμενα)

# Μορφότυποι και εργαλεία

- Χρειάζονται συχνά κάποιες περαιτέρω τροποποιήσεις (π.χ. Μετατροπή pdf σε txt)
- Χρειάζονται σενάρια αξιοποίησης των δεδομένων που μόνο οι επιστήμονες μπορούν να δημιουργήσουν
- Η ανάγκη για δεδομένα είναι τεράστια

# Για να έχουμε μια εικόνα στις άλλες γλώσσες

Εχετε μεταβεί σε πλήρη οθόνη. [Εξοδος από πλήρη οθόνη \(F11\)](#)

**W O R T S C H A T Z**  
UNIVERSITÄT LEIPZIG

## Corpus français - Université de Leipzig

Mot-clé:  [Recherche!](#) [?](#)

Texte de base:  distinction entre minuscules et majuscules au début d'un mot

[demandant](#) [découvertes](#) [ferroviaire](#) [séminaire](#) [Presse](#)

Changez le corpus:

[WWW](#) [Wikipédia](#) [journaux](#) [tout](#)

Le *Corpus français* est une base de données composée de près de 37 millions de phrases, soit environ 700 millions de mots. Il a été extrait et constitué par le groupe de recherche TAL de l'Université de Leipzig / Allemagne, et aménagé avec le concours de Daniel Elmiger et Alain Kamber (Université de Neuchâtel / Suisse). Les interfaces et outils de recherche ont été développés dans le cadre du projet Leipzig Corpora Collection of Computer Science de l'Université de Leipzig.

**Le corpus, dédié à l'étude du français contemporain écrit, est composé de trois parties:**

- Informations tirées de journaux francophones (plus de 19 millions de phrases)
- Pages web (plus de 11 millions de phrases)
- Wikipédia (près de 6 millions de phrases)

Si vous souhaitez obtenir des informations supplémentaires ou des instructions sur la façon d'utiliser le corpus, veuillez cliquer [ici](#).

[Conditions d'utilisation](#) · [Université de Leipzig](#) · [Department of Computer Science](#) · [Groupe de recherche TAL](#)

© 1998 - 2015 Deutscher Wortschatz · tous droits réservés · [contact](#)

#### ■ Frantext

Issu de la base Frantext, le corpus Frantext « textes libres de droits » offre à la communauté scientifique, un large champ d'investigation où sont réunies 500 œuvres de la littérature française couvrant la période du 18e au 20e siècle. Le traitement informatique des données textuelles en format TEI XML a été réalisé par le laboratoire ATILF. L'interface de recherche permet d'effectuer des sélections au sein du corpus par genre de texte, auteur, période...

☛ [Accès au corpus Frantext](#)

#### ■ Corpus journalistique de l'Est Républicain

Dans le cadre d'un accord de collaboration avec L'Est Républicain, le CNRTL offre après en avoir assuré le traitement informatique, l'accès à un nouveau corpus de type journalistique. Ce corpus est constitué des données textuelles correspondant à deux années de toutes les éditions intégrales du quotidien régional.

☛ [Accès au corpus de l'Est Républicain](#)

#### ■ Projet d'Etiqueteur Robuste pour l'Ecrit et pour l'Oral (PERCEO)

PERCEO est un ensemble de ressources destiné à servir à l'annotation automatique en parties du discours et en lemmes. Nous rendons disponible sur cette page les fichiers paramètres obtenus à l'aide du module d'entraînement du logiciel [Tree Tagger](#) ainsi que l'ensemble des ressources qui nous ont permis de les obtenir.

☛ [Accès au projet PERCEO](#)

#### ■ Traitement de Corpus Oraux en Français (TCOF)

Le projet « Traitement de Corpus Oraux en Français » (TCOF) est né de la volonté de conserver des corpus oraux constitués dans les années 80-90 à des fins de recherches personnelles. L'équipe constituée au sein du laboratoire ATILF (UMR CNRS 7118) a élaboré l'architecture d'une première base de données de corpus alignés texte/son avec Transcriber. Celle-ci s'est progressivement enrichie à partir des années 2000 grâce à la collaboration d'autres (enseignants-)chercheurs, d'ITA et d'étudiants en Sciences du langage de l'université de Nancy. Aujourd'hui, l'équipe met à disposition de la communauté scientifique une partie de ses ressources.

☛ [Accès au projet TCOF](#)

#### ■ Corpus d'articles de linguistiques issus de la revue "Sciences Humaines"

Dans le cadre d'un projet visant à la détection et la désambiguïsation de termes en texte intégral, un partenariat a été signé avec la revue "[Sciences Humaines](#)". Ce partenariat autorise à diffuser ces articles sous la licence creative Commons (attribution à l'auteur, pas d'utilisation commerciale, rediffusion aux mêmes conditions).

☛ [Accès au corpus de la revue "Sciences Humaines"](#)

#### ■ La Pèlerinage de Vie humaine de Guillaume de Dicuville



# Και παράλληλα σώματα κειμένων

Home / Query / WordAlign / Wiki [books] [DGT] [DOGC] [ECB] [EMEA] [EUbooks] [EU] [Europarl] [GNOME] [hren] [JRC] [KDE4/doc] [MBS] [MultiUN] [NCv9] [OO/OO3] [subs/12/13] [ParCor] [PHP] [SETIMES] [SPC] [Tatoeba] [TEP] [TedTalks] [TED] [Tanzil] [Ubuntu] [UN] [WikiSource] [WMT]

## OPUS ... the open parallel corpus

OPUS is a growing collection of translated texts from the web. In the OPUS project we try to convert and align free online data, to add linguistic annotation, and to provide the community with a publicly available parallel corpus. OPUS is based on open source products and the corpus is also delivered as an open content package. We used several tools to compile the current collection. All pre-processing is done automatically. No manual corrections have been carried out.

The OPUS collection is growing! Check this page from time to time to see new data arriving ...  
Contributions are very welcome! Please contact <jorg.tiedemann@lingfil.uu.se >

Search & download resources:

### Search & Browse

- OPUS multilingual search interface
- Europarl v7 search interface
- Europarl v3 search interface
- OpenSubtitles search interface
- EUconst search interface
- Word Alignment Database

### Tools & Info

- OPUS Wiki
- Tools for tagging and parsing
- Downloads (tools and models)
- Other annotation and corpus tools
- Experimental visualization tool for monolingual and parallel treebanks (demo)
- Uplug at bitbucket
- A reliable Language Identifier
- Scripts for OpenSubtitles2012/2013

### Links to other Resources

### Sub-corpora (downloads & infos):

- Books - A collection of translated literature (DOGC2014-07-17.tar.gz - 236 MB)
- DGT - A collection of EU Translation Memories provided by the JRC
- DOGC - Documents from the Catalan Government (DOGC2014-07-17.tar.gz - 702 MB)
- ECB - European Central Bank corpus
- EMEA - European Medicines Agency documents (EMEA0.3.tar.gz - 5.0 GB)
- The EU bookshop corpus (EUbookshop/EUbookshop0.2.tar.gz - 33 GB)
- EUconst - The European constitution (EUconst0.1.tar.gz - 67 MB)
- EUROPARL v7 - European Parliament Proceedings (Europarlv7.tar.gz - 8.4 GB)
- EUROPARL - European Parliament Proceedings (Europarl3.tar.gz - 3.6 GB)
- GNOME - GNOME localization files (GNOME2014-08-20.tar.gz - 9 GB)
- The Croatian - English WaC corpus (hrenWaC1.tar.gz - 48 MB)
- JRC-Acquis- legislative EU texts
- KDE4 - KDE4 localization files (v.2) (KDE4.tar.gz - 1.4 GB)
- KDEdoc - the KDE manual corpus (KDEdoc.tar.gz - 35 MB)
- MBS - Belgisch Staatsblad corpus
- MultiUN - Translated UN documents
- News Commentary (News-Commentary9.tar.gz - 2.2 GB)
- OO - the OpenOffice.org corpus (OpenOffice.tar.gz - 34 MB)
- OfisPublik - Breton - French parallel texts (OfisPublik0.1.tar.gz - 19MB)
- OpenOffice.org 3 corpus
- OpenSubtitles - the open subtitles corpus

### Latest News

- 2014-10-24: New: JRC-Acquis
- 2014-10-20: NCv9, TED talks, DGT, WMT
- 2014-08-21: New: Ubuntu, GNOME
- 2014-07-30: New: Translated Books
- 2014-07-27: New: DOGC, Tanzil
- 2014-05-07: Parallel coref corpus ParCor
- 2013-02-08: New version: EUbookshop
- 2013-02-01: New version: Tatoeba
- 2013-11-18: Extended corpus: Subs2013
- 2013-09-16: New version: Europarl v7
- 2013-08-22: Improved: OpenSubs2012
- 2013-07-01: New corpus EUbookshop
- 2013-02-26: Added UN and MultiUN
- 2012-12-20: New corpus: OfisPublik

# Πολλαπλές χρήσεις

- ◉ Αυτόματη εξαγωγή ορολογίας
- ◉ Συστήματα αυτόματης μετάφρασης
- ◉ Μεταφραστικές μνήμες
- ◉ Ανάλυση και διδασκαλία γλωσσών
- ◉ Συγκρίσιμα σώματα κειμένων και εξαγωγή πληροφορίας
- ◉ Ανάλυση γνώμης
- ◉ Μελέτες πεδίου

# Δεδομένα και εργαλεία

- ◉ Δεδομένα σε μορφή κειμένου ανοιχτά σε όλες και όλους
- ◉ Εργαλεία ελεύθερης πρόσβασης για ληματοποίηση, επισημείωση, μορφολογικοί και συντακτικοί αναλυτές, εργαλεία δημιουργίας συμφραστικών πινάκων μονόγλωσσων και πολύγλωσσων, εξαγωγής ορολογίας κ.λπ.
- ◉ Ιδιαιτερότητα της ελληνικής (πολυτονικό)



# Για μια θέση της Ελλάδας στον ψηφιακό χάρτη

## CLARIN-EL

Common Language Resources and Technology Infrastructure



[Αρχική Σελίδα](#) [CLARIN](#) [CLARIN-EL](#) [Χαρτογράφηση](#) [Συχνές Ερωτήσεις](#) [Δημοσιεύσεις](#) [Παραδοτέα](#) [Επικοινωνία](#)

Καλωσήρθατε στον ιστότοπο του CLARIN-EL, του δικτύου Γλωσσικών Πόρων και Τεχνολογιών (ΓΠΤ) για την ελληνική γλώσσα

Το CLARIN EL είναι το ελληνικό σκέλος του έργου CLARIN (<http://www.clarin.eu>), μιας πανευρωπαϊκής προσπάθειας

- να **συγκεντρωθούν**
- και να **διατεθούν** στην ερευνητική κοινότητα
- μέσω μιας **διαδικτυακής Ερευνητικής Υποδομής γλωσσικοί πόροι** (κειμένα, πολυμεσικά δεδομένα, λεξικά κτλ.) σε όλες τις γλώσσες,
- και **εργαλεία γλωσσικής τεχνολογίας** για την επεξεργασία τους (μορφολογικοί / συντακτικοί αναλυτές, εργαλεία επισημείωσης, στατιστικά εργαλεία κτλ.).

Το CLARIN-EL στην τρέχουσα κατασκευαστική φάση στοχεύει

- να διαθέσει στην ερευνητική κοινότητα την υποδομή και τον μηχανισμό
  - για τη συγκέντρωση και την τεκμηρίωση γλωσσικών πόρων (κειμένων, καταγραφών προφορικού λόγου, λεξικών, ορολογικών γλωσσαρίων κτλ.) και εργαλείων γλωσσικής τεχνολογίας για τα Ελληνικά
  - σε ένα δίκτυο αποθετηρίων τα οποία θα εμπλουτίζονται διαρκώς από τα μέλη του δικτύου
- να παρέχει διαδικτυακές υπηρεσίες πρόσβασης σε πόρους και εργαλεία
- να διευρύνει το νομικό καθεστώς πρόσβασης σε ΓΠΤ και να παρέχει νομικές συμβουλευτικές υπηρεσίες σε παρόχους και χρήστες ΓΠΤ
- να ενημερώνει και να εκπαιδεύει ερευνητές και ευρύ κοινό για τις δυνατότητες στη χρήση των ΓΠΤ.

είστε ερευνητής;

• **έχετε αναπτύξει ή συγκεντρώσει γλωσσικούς πόρους σε ψηφιακή μορφή;**

### Προσέχί Γεγονότα

Δεν υπάρχουν γεγονότα

Κ	Δ	Τ	Τ	Π	Π	Σ
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31	1	2	3	4

Κ	Δ	Τ	Τ	Π	Π	Σ
29	30	31	1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30	1	2

### Σύνδεσμοι

- ▶ Ευρωπαϊκό CLARIN
- ▶ Ολλανδικό CLARIN
- ▶ Ισπανικό CLARIN

### Είσοδος

Όνομα Χρήστη

Κωδικός

Να με θυμάσαι

**Σύνδεση**

- Εξάσστε τον κωδικό σας
- Εξάσστε το όνομα χρήστη;
- Δημιουργία λογαριασμού

Σας ευχαριστώ πολύ για την προσοχή σας!