

# Filologia Risorse informatiche

Carnet de recherche and online journal – Italian Digital Humanities. ISSN: 2496-6223. DOI:

<https://doi.org/10.5281/zenodo.1193619>

## Note minime sul TLG: brevi cenni sulle 'origini' del TLG e piccolo vademecum (di Giovanna Battaglini)

bimestre novembre-dicembre 2018 (31 dicembre)

### **Per una definizione del TLG, *alias Treasury of the Greek Language*.**

Il TLG si configura come il più importante ed il più noto *corpus* digitale di testi greci, nonché il più ampio: i testi letterari greci digitalizzati afferiscono ad un lasso di tempo molto vasto, che va dall'epoca omerica (VIII secolo a.C.) alla caduta di Bisanzio (1453). Il TLG comprende 3300 autori e più di 11.000 opere.

TLG è un acronimo, che sta per *Treasury of the Greek Language* e che richiama consapevolmente il titolo del *lexicon* cinquecentesco (*Thesaurus Linguae Graeciae*) curato dal filologo Henri Estienne (Stephanus 1572). Benché il TLG nasca come progetto di digitalizzazione, ne esiste anche una sorta di edizione a stampa, a cura di Berkowitz-Squitier 1990 (*Thesaurus Linguae Graeciae. Canon of Greek Authors and Works*). Il corretto modo per citare la versione digitale (come suggerito sul sito del TLG) è il seguente: *Thesaurus Linguae Graeciae. Digital Library*. Ed. Maria C. Pantelia. University of California, Irvine.

Il maggiore punto di forza del TLG è la (concretizzata) possibilità di combinare le metodologie tradizionali (filologiche e letterarie) con le ITC: i testi greci sono digitalizzati a partire dalle principali edizioni critiche di riferimento (raramente, la versione digitalizzata non 'proviene' dalla migliore e più recente edizione critica; ma, talvolta, è possibile anche confrontare edizioni critiche differenti, con ricadute utilissime e rilevanti sul piano filologico) e sono interrogabili proprio sfruttando gli strumenti informatici della *textual search* e *lemmatic search*.

**Le origini del TLG Project: dall'idea di Marianne McDonald alla versione *online*.**

Il *TLG* è nato in seno ad uno speciale progetto di ricerca della University of California nel 1972. Tutto nacque da un'idea di Marianne McDonald, grecista ed allieva di Bruno Snell, la quale, dovendo realizzare una tesi relativa al lessico della felicità in Euripide, nel 1971 propose l'elaborazione di un database computerizzato della letteratura greca. Come si legge nella sezione *About* del sito del *TLG* – che riporta in maniera dettagliata le principali tappe di questo progetto, che ha cambiato lo studio della letteratura greca (anche valorizzando l'approccio di tipo lessicografico) – «the creation of the *TLG* came at the end of a long tradition of scholarly efforts to systematically collect and preserve extant Greek literature». Il primo di questi tentativi (o almeno, il primo notevole tentativo di allestimento di un *lexicon* della lingua greca) fu, appunto, il *TLG* dello Stephanus, a cui, in particolare nel XIX secolo, specie a seguito di una serie di ritrovamenti papiracei, seguì una serie di *lexica* sia greci che latini (si pensi, a mero titolo esemplificativo, al *Lexicon of the Greek Language* a cura di Pickering 1855 e al *Lexicon of the Latin Language* a cura di Leverett 1850). Lo stesso Snell si interessò a questo tipo di studi e fondò, nel 1953 ad Amburgo, l'*Istituto del Thesaurus Linguae Graeciae*, la cui *mission* era quella di studiare la lingua greca arcaica ed i successivi slittamenti semantici, al fine di comprendere lo sviluppo del pensiero greco. L'*Istituto* pubblicò, fra l'altro, l'edizione del *Lexicon für das frühgriechischen Epos* e l'edizione dell'*Index Hippocraticus*.

La prima versione *online* del *TLG* è stata resa disponibile nell'aprile del 2001 e consentiva agli utenti la consultazione di risorse bibliografiche e testuali. Ne esiste anche una versione in CD-rom, il miglior programma per consultare il quale è *Diogenes* (gratuito ed *open source*); l'ultima versione in CD-rom è il *TLG E*, realizzato nell'estate del 1999. Tuttavia, la versione in CD-rom da anni non è stata più aggiornata, sicché fa testo la versione *online*, caratterizzata da un costante arricchimento di digitalizzazioni testuali. La versione online è consultabile tramite abbonamento oppure (nel caso in cui si stia svolgendo un corso di laurea, un dottorato, un perfezionamento *et similia*) 'sfruttando' il *proxy* dell'Università di riferimento, (qualora suddetta Istituzione – come di norma accade – abbia sottoscritto l'abbonamento annuale). Ad ogni modo, a tal proposito due sono gli aspetti rilevanti: un *sub-corpus* del *TLG* (cioè una parte del *database*) è in *open access*: si tratta del *Canon of Greek Authors and Works (Abridged TLG)*; inoltre, le funzioni di ricerca offerte dal *Canon (Abridged TLG)* e dal *TLG (full corpus)* sono le stesse. Ciò che cambia è la mole dei testi digitalizzati interrogabili; inoltre, va sottolineato che il *Canon* contiene la digitalizzazione di tutti i testi dei principali autori della letteratura greca, quali, ad esempio, Omero, Esiodo, i tragici, gli oratori. Pertanto, già nella sola versione gratuita il *TLG* si presenta come un *tool* di straordinaria importanza ed utilità.

Nel 1972 ci fu la prima *planning Conference*, alla quale furono invitati classicisti provenienti sia dall'America che dall'Europa. Il contributo più importante fu quello offerto da David W. Packard, inventore dell'*Ibycus System*, nome dell'*hardware* e del *software* usati inizialmente per correggere le digitalizzazioni dei testi greci e per effettuare ricerche nel *TLG corpus*. Packard creò anche l'*Alpha Code*, la prima convenzione di codifica e formattazione dei caratteri greci; l'*Alpha Code* prese

successivamente il nome di *Beta Code* (tutt'oggi usato per la codifica del greco politonico, anche nell'ambito di importanti progetti, quali il *Perseus Project* ed il *Packard Humanities Institute*), che assegna una posizione *ASCII* a ciascuno dei 24 caratteri dell'alfabeto greco, indicando i segni diacritici con caratteri non alfabetici.

### **Breves notulae sulle funzioni del TLG.**

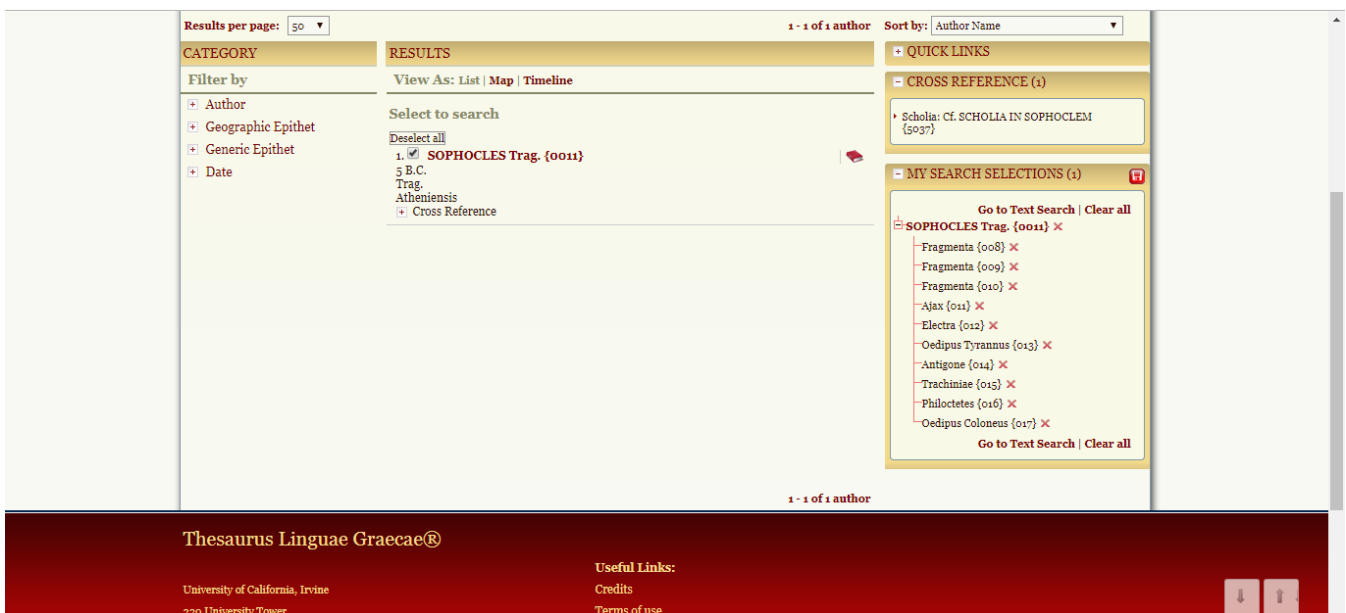
Il TLG consente anche semplicemente di 'sfogliare' un singolo testo (*Brows*) oppure di visualizzare sinotticamente due testi (*Brows two texts*); ma, chiaramente, la principale funzione offerta dal TLG è quella di ricerca testuale (nonché lemmatica e sintagmatica) all'interno dell'ampio *corpus* di testi digitalizzati, sia nell'ambito del *Canon* (*Abridged TLG*) che nell'ambito del *TLG* (*full corpus*). Per fare ciò, occorre dapprima inserire il nome dell'autore, ricordando che i nomi degli autori sono indicizzati in latino (*e.g. Sophocles*), con la possibilità di precisare l'epiteto geografico (*Geographic Epithet*) o un eventuale epiteto generico (*Generic Epithet*) o un altro tipo di epiteto (*Non-Standard Gen. Epith.*) e naturalmente il secolo di appartenenza (*Data*):

The screenshot shows the TLG website interface. At the top, there is a navigation bar with the logo and the text 'THESAURUS LINGVAE GRAECAE' and 'A Digital Library of Greek Literature'. Below the navigation bar, there is a search bar with the text 'SOPHOCLES' and a 'GO' button. The search criteria are set to 'Author'. The results show 'SOPHOCLES Trag. {0011}' with a red book icon. The interface includes navigation tabs like HOME, CANON, TEXT SEARCH, BROWSE, LEXICA, N-GRAMS, STATISTICS, VOCAB. TOOLS, and HELP. There is also a sidebar with 'QUICK LINKS' and 'CROSS REFERENCE'.

Cliccando sull'icona raffigurante un libro rosso, è possibile visualizzare tutti i riferimenti bibliografici digitalizzati con riferimento a tutte le opere di Sofocle:



A questo punto, occorre selezionare l'autore richiesto (nel caso esemplificato, la ricerca è stata condotta in maniera da avere un unico risultato: *Sophocles Trag. Atheniensis*), le cui opere compariranno nella sezione *My search selection*. Prima di procedere alla ricerca testuale (*Go to Text search*), è possibile de-selezionare alcune opere, nel caso in cui le si voglia escludere intenzionalmente dalla ricerca. (Ad esempio, nel caso esemplificato, si potrebbe pensare di effettuare la ricerca solo nell'ambito delle tragedie 'integre', escludendo, dunque, la produzione 'frammentaria'.)



È possibile effettuare ricerche non solo per autore, ma anche per editore (e.g. Lloyd-Jones & Wilson), titolo dell'opera – sempre rammentando che anche le opere sono indicizzate in latino – (e.g. *Ajax*), titolo della pubblicazione (e.g. *Sophoclis Fabulae*) oppure anche tramite il titolo della serie (e.g. *Poetarum Graecorum Fragmenta*, Testi a fronte) e per anno di pubblicazione.

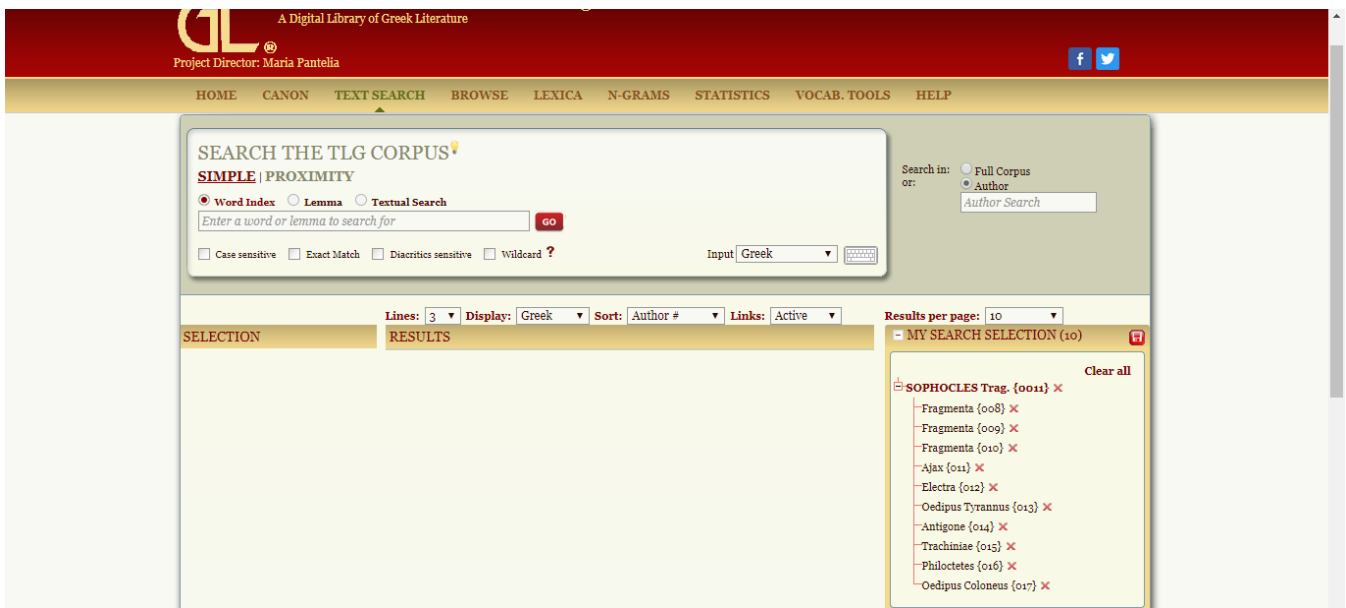
La ricerca testuale può essere effettuata a vari livelli:

~ *word index*: la parola digitata è 'trattata' come parte iniziale di un possibile lemma (aggettivo, sostantivo, verbo) da cercare, tenendo conto del fatto che suddetta parola può essere digitata o in Beta

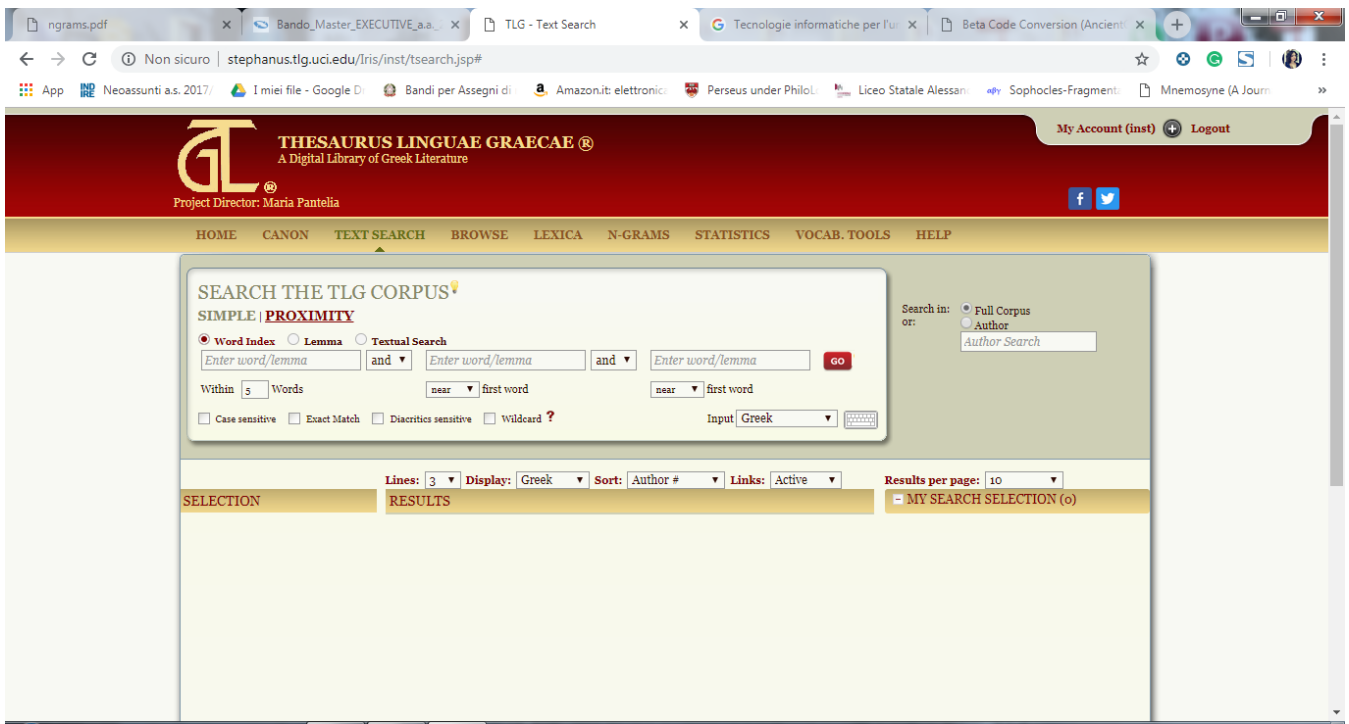
Code o in traslitterazione oppure direttamente in caratteri greci, o usando la ‘tastierina’ che il sito mette a disposizione oppure utilizzando la propria tastiera di greco politonico (qualora la si sia già attivata/installata sul proprio pc);

~ *lemma*: tale funzione consente di ricercare la specifica parola digitata (sia pure variamente declinata/coniugata), che viene dunque considerata come lemma a sé;

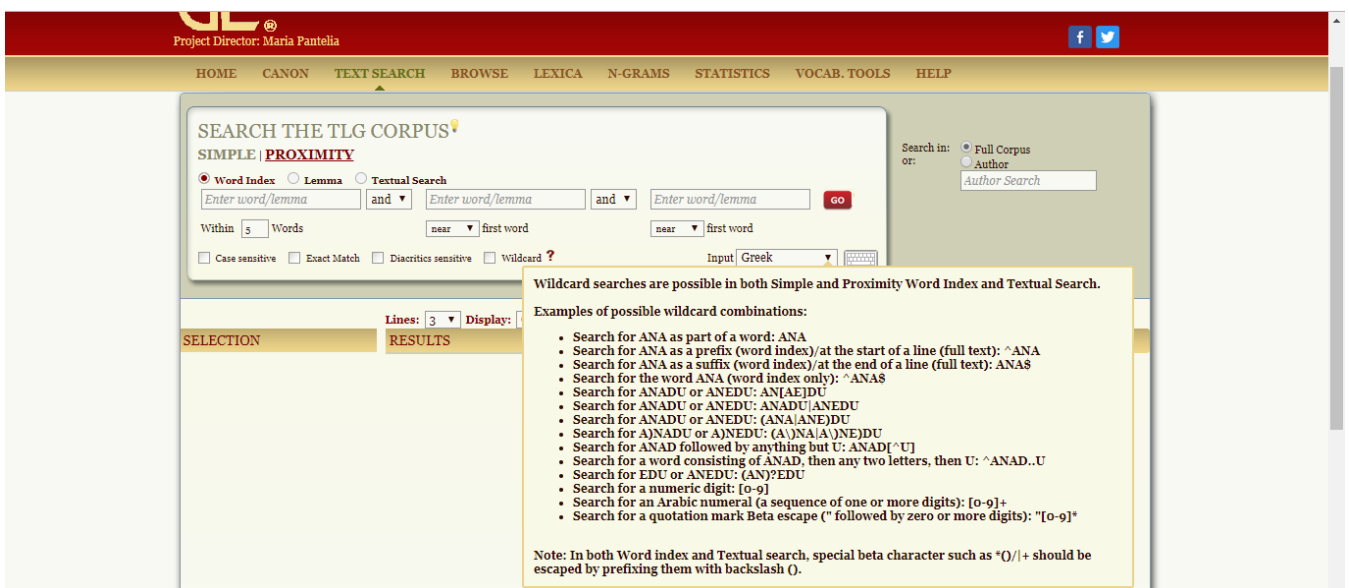
~ *textual search*: ricerca testuale per sintagmi:



Quanto sin qui accennato vale per la ricerca semplice (*Search the TLG Corpus – Simple*); molto utile è anche la funzione *Proximity*, che consente di inserire la parola/il lemma/il sintagma cercata/-o con elementi di contesto, vale a dire unitamente a parole che si trovano in sua prossimità. La parola (*word index*), il lemma (*lemma*) o il sintagma (*textaul search*) rappresentano la prima ‘parola’ (*first word*) ed è possibile precisare se la seconda ed eventualmente la terza si trovino prima (*before*), dopo (*after*) o semplicemente in prossimità (*near*). È anche possibile precisare il numero di parole di ‘prossimità’ (di *default*, viene suggerito il numero di 5 parole):



Nel caso delle funzioni di ricerca *Word index* e *Textual search* è possibile utilizzare particolari metacaratteri (*Wildcards*) che consentono di precisare se la parola che intendiamo cercare rappresenta, ad esempio, un prefisso o un suffisso o se vogliamo cercarla nella *facies* nella quale la digitiamo (nel caso di una preposizione, un avverbio, una particella) o ancora se vogliamo cercare non solo la parola digitata, ma anche sue eventuali varianti dialettali:



Il *TLG* consente anche di salvare le ricerche effettuate (sino ad un massimo di dieci ricerche), ma anche di stampare (*printable version*) tali risultati.

Tramite la funzione *Lexica*, il *TLG* consente di usare anche risorse lessicografiche (*lexicographical resources*), quali *LSJ* (*The Online Liddel Scott – Jones Greek-English Lexicon*), il *Middle LS*, *Cunliffe* (*A Lexicon of the Homeric Dialect*), *DGE* (*Diccionario Griego-Español*), *Slater* (*Lexicon to Pindar*),

Powell (*A Lexicon to Herodotus*), Trapp-LBG (*Lexicon zur byzantinischen Gräzität*), Kriaras (*Ancient Greek-Modern Greek*), Triantafyllides (*Dictionary of Standard Modern Greek*):

Utilissima è la funzione *N-Grams*, che consente di identificare passaggi comuni fra due testi dello stesso autore o di autori diversi. Comparando due autori diversi, è possibile precisare le opere all'interno delle quali effettuare la ricerca o se il confronto debba coinvolgere tutte le opere dei due autori in questione. Le sezioni comuni vengono evidenziate in giallo. La comparazione consente anche piccole 'interpolazioni' all'interno dei passaggi comuni (se esse compaiono nel testo di partenza):

La comparazione (*compare texts*) viene effettuata per lemmi, il che vuol dire che sono 'consentite' anche differenze di flessione fra i due testi; ma è, altresì, possibile selezionare la funzione *wordforms* (in luogo della funzione *lemma*), per individuare solo le esatte ripetizioni di una parola e/o di un sintagma

e/o di una intera espressione. Questa interessantissima ed utilissima funzione è resa possibile da un metodo di analisi di tipo statistico, basato sulla distribuzione della frequenza delle lettere (e, più precisamente, dei digrammi e dei trigrammi): si tratta della *N-Grams Analysis*. Questo tipo di analisi non tiene conto delle *Stop-words*, che non sono semanticamente significative, né dell'ordine delle parole. Nella comparazione di testi di due autori diversi, l'analisi è basata sui trigrammi e i risultati visualizzati contengono un minimo di due (ed un massimo di quattro) trigrammi. Questa ricerca può presentare difficoltà o non essere fruttuosa nel caso di testi particolarmente brevi, come i frammenti (per alcuni dei quali disponiamo del solo titolo).

Infine, altra sezione utile del *TLG* è la funzione *Statistics*, che consente di individuare le parole più ricorrenti o nell'intero corpus del *TLG* o in uno specifico autore o in una specifica opera o in una specifica serie o in uno specifico secolo. I risultati sono evidenziati anche graficamente, soprattutto tramite istogrammi:



STATISTICS

Full Corpus Author Lemma

SOPHOCLES

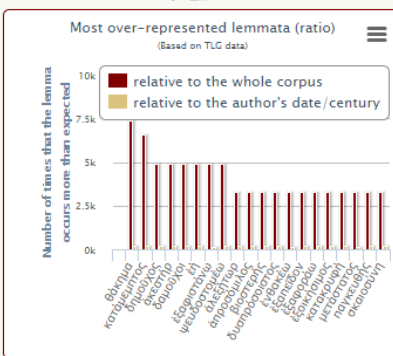
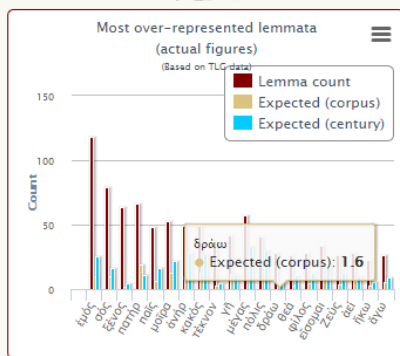
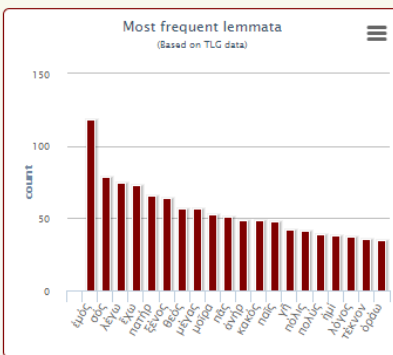
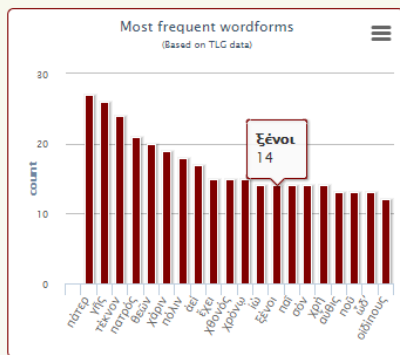
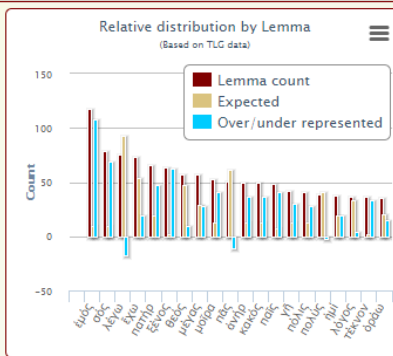
GO

Or Oedipus Coloneus {017}

SUMMARY

- There are 11,007 words in this work corresponding to 4,211 unique forms and 2,332 lemmata
- The most frequent wordforms are: πάτερ, γῆς, τέκνον (**more**)
- The most frequent lemmata are: ἐμός, -ή, -όν, σός, σή, σόν, λέγω (**more**)
- The most over-represented lemmata in this work are: ἐμός, -ή, -όν, σός, σή, σόν, ξένος, -η, -ον (**more**)
- The most under-represented lemmata in this work are: γίνωμαι, φημί, ποιέω (**more**)
- There are 20 unique occurrences in this work (**more**)

- OVER-REPRESENTED LEMMATA
- UNDER-REPRESENTED LEMMATA
- MOST FREQUENT WORDFORMS
- MOST FREQUENT LEMMATA
- UNIQUE OCCURRENCES



Credo che già da queste brevi note – che, naturalmente, non hanno alcuna pretesa di esaustività – emerga lo straordinario potenziale di uno strumento come il TLG, divenuto oramai insostituibile per gli studi letterari, filologici, lessicografici sui testi greci.

Bibliografia dei lavori citati

Berkowitz-Squitier 1990 =

L. Berkowitz and K. A. Squitier, *Thesaurus Linguae Graeciae. Canon of Greek Authors and Works* (3<sup>rd</sup> edition), Oxford University Press 1990.

Leverett 1850 =

*A new and copious Lexicon of the Latina Language: compiled chiefly from the Magnum totius latinitatis lexicon of Facciolati and Forcellini, and the German works of Scheller and Luenemann*, edited by F. P. Leverett, vol. 1, Boston 1850. – link per il download del lexicon (vol. 1) in pdf: <<https://archive.org/details/newcopiouslexico00leveuoft/page/n7>>

Pickering 1855 =

J. Pickering, *A comprehensive Lexicon of the Greek Language*, Boston 1855. – link per il download del lexicon in pdf: <<https://archive.org/details/12487584.1705.emory.edu/page/n3>>

Stephanus 1572 =

ΘΗΣΑΥΡΟΣ ΤΗΣ ΕΛΛΗΝΙΚΗΣ ΓΛΩΣΣΗΣ. *Thesaurus Graeciae Linguae ab Henrico Stephano constructus* 1572. – link per il download del Thesaurus in pdf: <<https://books.google.it/books?id=IXFpAAAAcAAJ&hl=it&pg=PA1#v=onepage&q&f=false>>

## Sitografia

\* ASCII (s.v.) in *GloDIUM*: <<https://fri.hypotheses.org/glossario-di-informatica-umanistica>>

\* *Beta Code* (s.v.), in *Wikipedia*: <[https://en.wikipedia.org/wiki/Beta\\_Code](https://en.wikipedia.org/wiki/Beta_Code)>

\* *Biblioteca digitale* (s.v.) in *GloDIUM*: <<https://fri.hypotheses.org/glossario-di-informatica-umanistica>>

\* *Diogenes* (software): <<http://community.dur.ac.uk/p.j.heslin/Software/Diogenes/>>

\* *Filologia digitale* (s.v.) in *GloDIUM*: <<https://fri.hypotheses.org/glossario-di-informatica-umanistica>>

\* *Packard Humanities Institute* (s.v.), in *Wikipedia*: <[https://en.wikipedia.org/wiki/Packard\\_Humanities\\_Institute](https://en.wikipedia.org/wiki/Packard_Humanities_Institute)>

\* *Packard Humanities Institute* (official website): <<https://packhum.org/>>

\* *Perseus Project* (s.v.), in *Wikipedia*: < [https://en.wikipedia.org/wiki/Perseus\\_Project](https://en.wikipedia.org/wiki/Perseus_Project)>

\* *Perseus Project* (official website): < <http://www.perseus.tufts.edu/hopper/>>

\* Snell, Bruno (s.v.) in *Treccani, Enciclopedia italiana – V appendice* (1994): < [http://www.treccani.it/enciclopedia/bruno-snell\\_%28Enciclopedia-Italiana%29/](http://www.treccani.it/enciclopedia/bruno-snell_%28Enciclopedia-Italiana%29/)>

\* *TLG* (ultima versione digitale del *TLG: Thesaurus Linguae Graeciae. Digital Library*. Ed. Maria C. Pantelia. University of California, Irvine): < <http://stephanus.tlg.uci.edu/>>

\* *TLG* (sub-corpus del) in *open access: Canon of Greek Authors and Works (Abridged TLG)*: <<http://stephanus.tlg.uci.edu/Iris/canon/csearch.jsp>>

\* *TLG, Istituto del* (s.v.) in *Wikipedia*: <[https://it.wikipedia.org/wiki/Thesaurus\\_Linguae\\_Graecae\\_\(Amburgo\)](https://it.wikipedia.org/wiki/Thesaurus_Linguae_Graecae_(Amburgo))>

\* *Thesaurus Linguae Graeciae* (s.v.), in *Wikipedia*: < [https://en.wikipedia.org/wiki/Thesaurus\\_Linguae\\_Graecae](https://en.wikipedia.org/wiki/Thesaurus_Linguae_Graecae)>

\* *Thesaurus* (s.v.) in *GloDIUm*: < <https://fri.hypotheses.org/glossario-di-informatica-umanistica>>



This entry was posted in Posts on 02/01/2019 [<https://fri.hypotheses.org/1391>] .

---

---