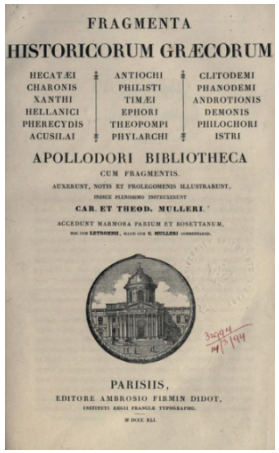
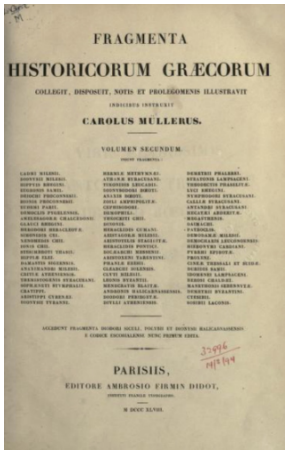
The Digital Fragmenta Historicorum Graecorum

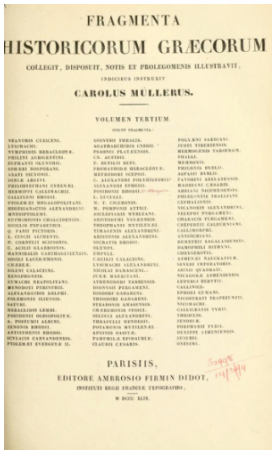
Tariq Yousef
University of Leipzig

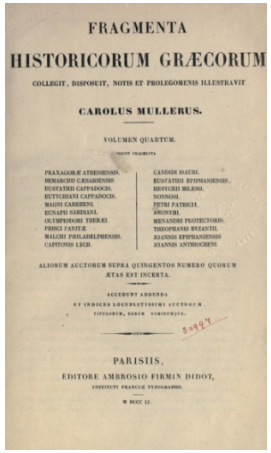
Sunoikisis DC 2016
December 16, 2015
Leipzig, Germany

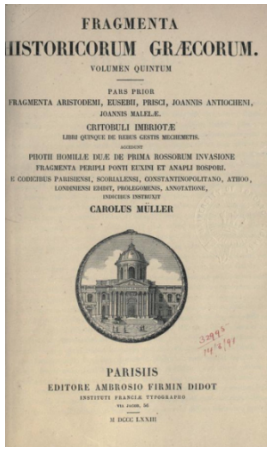
Digital Fragmenta Historicorum Graecorum (DFHG) Project


FRAGMENTA
HISTORICORUM GRÆCORUM
HECATÆI CHARONIS XANTHI HELLANICI PHERECYDIS ACUSILAI
ANTIOCHI PHILISTI TIMÆI EPHORI THEOPOMPI PHYLARCHI
CLITODEMI PHANODEMI ANDROTIONIS DEMONIS PHILOCHORI ISTRI
APOLLODORI BIBLIOTHECA
CUM FRAGMENTIS.
AUXERUNT, NOTIS ET PROLEGOMENIS ILLUSTRARUNT,
INDICE PLENISSIMO INSTRUXERUNT
CAR. ET THEOD. MULLERI.
ACCEDUNT MARMORA PARIUM ET ROSETTANUM.
PARISIIS,
EDITORE AMBROSIO FIRMIN DIDOT,


FRAGMENTA
HISTORICORUM GRÆCORUM
COLLEGIT, DISPOSUIT, NOTIS ET PROLEGOMENIS ILLUSTRAVIT
CAROLUS MULLERUS.
VOLUMEN SECUNDUM.
PARISIIS,
EDITORE AMBROSIO FIRMIN DIDOT,


FRAGMENTA
HISTORICORUM GRÆCORUM
COLLEGIT, DISPOSUIT, NOTIS ET PROLEGOMENIS ILLUSTRAVIT,
CAROLUS MULLERUS.
VOLUMEN TERTIUM
PARISIIS,
EDITORE AMBROSIO FIRMIN DIDOT,


FRAGMENTA
HISTORICORUM GRÆCORUM
COLLEGIT, DISPOSUIT, NOTIS ET PROLEGOMENIS ILLUSTRAVIT
CAROLUS MULLERUS.
VOLUMEN QUARTUM.
PARISIIS,
EDITORE AMBROSIO FIRMIN DIDOT,


FRAGMENTA
HISTORICORUM GRÆCORUM.
VOLUMEN QUINTUM
PARS PRIOR
FRAGMENTA ARISTODEMI, EUSEBII, PRISCI, JOANNIS ANTIOCHENI, JOANNIS MALELÆ.
CRITOBULI IMBRIOTÆ
LIBRI QUINQUE DE REBUS GESTIS MEHEMETIS.
PHOTII HOMILIÆ DUÆ DE PRIMA ROSSORUM INVASIONE
FRAGMENTA PERIPLI PONTI EUXINI ET ANAPLI BOSPORI.
CAROLUS MULLER
PARISIIS
EDITORE AMBROSIO FIRMIN DIDOT

www.dfhg-project.org

5 volumes - ca. 600 authors
6th century BC - 7th century CE
Greek-Latin alignment

Dfhg-dev

Digital Fragmenta Historicorum Graecorum (DFHG): a machine-corrected version of the five volumes of Karl Müller's Fragmenta Historicorum Graecorum.

View the Project on GitHub
OpenGreekAndLatin/dfhg-dev

Download **ZIP File** | Download **TAR Ball** | View On **GitHub**

This project is maintained by OpenGreekAndLatin

Hosted on GitHub Pages — Theme by orderedlist

Welcome to dfhg-dev!

About

This repository contains machine-corrected EpiDoc versions of the five volumes of Karl Müller's *Fragmenta Historicorum Graecorum* (FHG) as part of the Digital Fragmenta Historicorum Graecorum (DFG) project at the University of Leipzig. The files are encoded by the *Open Greek and Latin* research team (researchers and student assistants). While adequate, more can and should be done to improve the XML (the -dev(elopment) suffix indicates the repository is open to improvement).

If you want to help or contribute to the encoding process, please read the following documentation:

- DFHG Guidelines (1.0).
- DFHG Authors.

For a full description of the DFHG project, please visit our website.

Contents

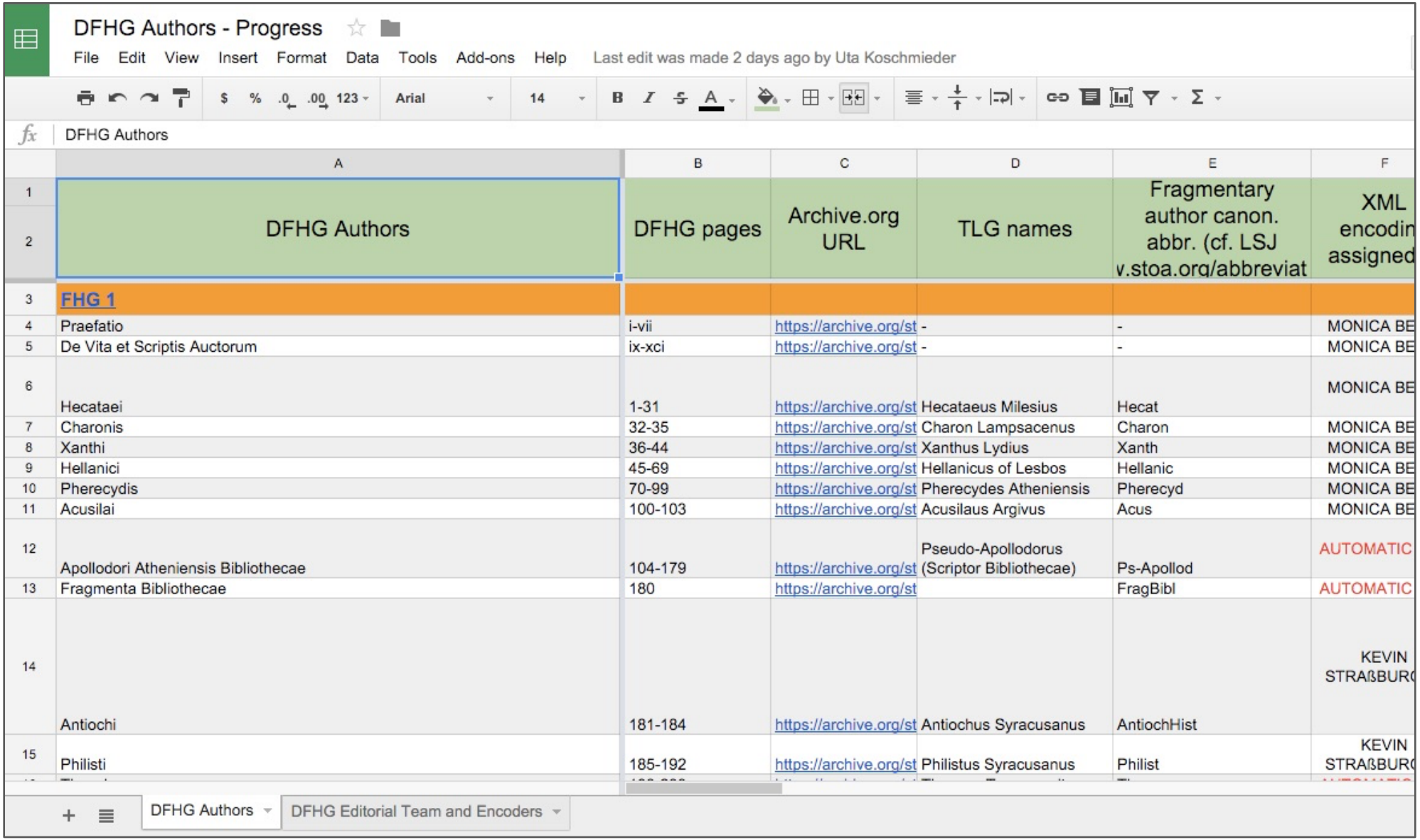
For a comprehensive view of the contents of the FHG, click here. Authors encoded thus far are (authors are listed according to the order of Müller's FHG):

Volume 1

- Charon Lampsacenus
- Hellanicus Lesbius
- Acusilaus Argivus
- Antiochus Syracusanus

GitHub, Inc. [US] https://github.com/OpenGreekAndLatin/dfhg-dev/blob/master/Volume_1/Ister%20Cyrenaeus/dfhg1ister_grc.xml

```
132                         <ab><title xml:lang="la">LIBER I.</title></ab>
133
134                         <cit n="1" xml:lang="grc">
135                             <bibl>Suidas v. Τιτανίδα γῆν</bibl>
136                             <quote>... <persName role="author">Ἴστρος</persName> ἐν <num value="1"
137                                     >α</num>
138                                 <title>Ἀττικῶν</title>, Τιτᾶνας βοᾶν· ἐ&#976;οήθουν γὰρ τοῖς
139                                 ἀνθρώποις ἐπακούοντες, ὡς <persName role="author"
140                                     >Νίκανδρος</persName> ἐν <num value="1">α</num>
141                                 <title>Αἰτωλικῶν</title>.</quote>
142                             <note xml:lang="la">Idem legitur in <bibl>Photii Lexico</bibl>.</note>
143                         </cit>
144
145                         <cit n="2" xml:lang="grc">
146                             <bibl>Apostolius XVIII , 77</bibl>
147                             <quote>Τιτανίδα παροικεῖς, ἐπὶ τῶν φιλοθέων. Οἱ μὲν τὴν πᾶσαν γῆν, οἱ δὲ
148                                 τὴν Ἀττικὴν ἀπὸ Τιτηνίου, ἑνὸς τῶν Τιτάνων ἀρχαιοτέρου, οἰκήσαντος
149                                 περὶ Μαραθῶνα, ὃς μόνος οὐκ ἐστράτευσεν ἐπὶ τοὺς θεούς, ὡς <persName
150                                     role="author">Φιλόχορος</persName> ἐν <title>Τετραπόλει</title>
151                                 καὶ <persName role="author">Ἴστρος</persName> ἐν <num type="ordinal"
152                                     value="1">πρώτῃ</num> τῶν <title>Ἀττικῶν</title>.</quote>
153                             <note xml:lang="la">Vide <bibl>Pliavonn. s. Τιταν. γῆν</bibl>, ubi eadem
154                                 leguntur, paucis omissis et suppresso Istri nomine.</note>
155                         </cit>
156
157                         <cit n="3" xml:lang="grc">
158                             <bibl>Harpocrat.: Λαμπάς.</bibl>
159                             <quote>... Τρεῖς ἄγουσιν Ἀθηναῖοι ἑορτὰς λαμπάδας (λαμπάδος <foreign
160                                     xml:lang="la">cod. Angl.</foreign>), Παναθηναίοις, καὶ
161                                 Ἡφαιστείοις, καὶ Προμηθείοις, ὡς <persName role="author"
162                                     >Πολέμων</persName> φησὶν ἐν τῷ περὶ τῶν ἐν τοῖς προπυλαίοις
163                                 πινάκων. <persName role="author">Ἴστρος</persName> δ' ἐν <num
164                                     type="ordinal" value="1">πρώτῃ</num> τῶν <title type="alt"
165                                     >Ἀτθίδων</title>, εἰπὼν ὡς ἐν τῇ τῶν Ἀπατουρίων ἑορτῇ Ἀθηναίων
166                                 οἱ καλλίστας στολὰς ἐνδεδυκότες, λα&#976;όντες ἡμμένας λαμπάδας ἀπὸ
167                                 τῆς ἑστίας, ὑμνοῦσι τὸν Ἥφαιστον θύοντες (<foreign xml:lang="la">sic
168                                     codd.</foreign>; <bibl>Valesius</bibl> θέοντες), ὑπόμνημα τοῦ
169                                 κατανοήσαντα τὴν χρείαν τοῦ πυρὸς διδάξαι τοὺς ἄλλους.</quote>
```



DFHG Authors - Progress

File Edit View Insert Format Data Tools Add-ons Help Last edit was made 2 days ago by Uta Koschmieder

Arial 14

fx DFHG Authors

	A	B	C	D	E	F
1	DFHG Authors	DFHG pages	Archive.org URL	TLG names	Fragmentary author canon. abbr. (cf. LSJ v.stoa.org/abbreviat	XML encodin assigned
2						
3	FHG 1					
4	Praefatio	i-vii	https://archive.org/st	-	-	MONICA BE
5	De Vita et Scriptis Auctorum	ix-xci	https://archive.org/st	-	-	MONICA BE
6	Hecataei	1-31	https://archive.org/st	Hecataeus Milesius	Hecat	MONICA BE
7	Charonis	32-35	https://archive.org/st	Charon Lampsacenus	Charon	MONICA BE
8	Xanthi	36-44	https://archive.org/st	Xanthus Lydius	Xanth	MONICA BE
9	Hellanici	45-69	https://archive.org/st	Hellanicus of Lesbos	Hellanic	MONICA BE
10	Pherecydis	70-99	https://archive.org/st	Pherecydes Atheniensis	Pherecyd	MONICA BE
11	Acusilai	100-103	https://archive.org/st	Acusilaus Argivus	Acus	MONICA BE
12	Apollodori Atheniensis Bibliothecae	104-179	https://archive.org/st	Pseudo-Apollodorus (Scriptor Bibliothecae)	Ps-Apollod	AUTOMATIC
13	Fragmenta Bibliothecae	180	https://archive.org/st		FragBibl	AUTOMATIC
14	Antiochi	181-184	https://archive.org/st	Antiochus Syracusanus	AntiochHist	KEVIN STRAßBUR
15	Philisti	185-192	https://archive.org/st	Philistus Syracusanus	Philist	KEVIN STRAßBUR

DFHG Authors | DFHG Editorial Team and Encoders

Automatic XML Tagging

Automatic XML Tagging

XML markup levels:

- **Macro Level Markup:** Chapters, sections, paragraph, ..etc

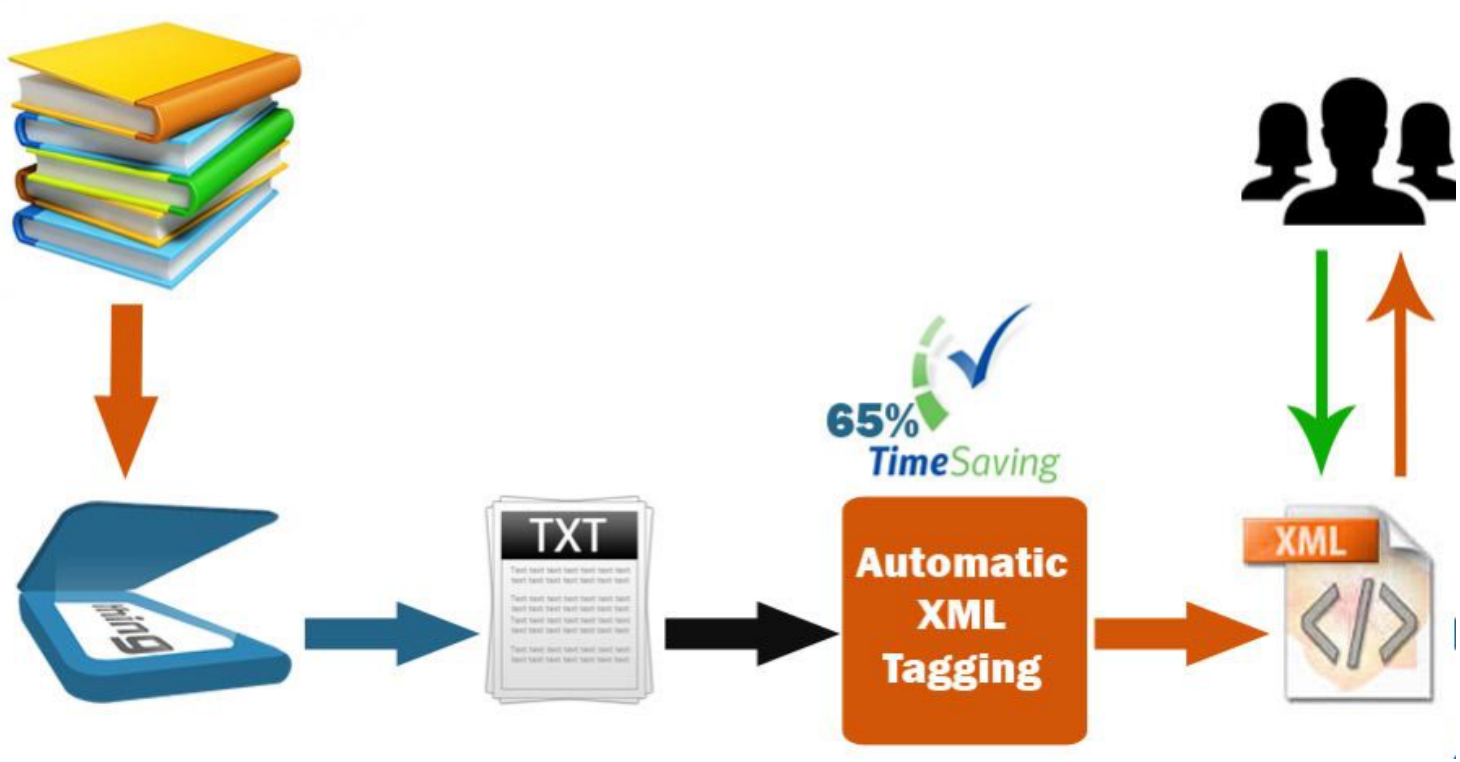
- **Micro Level Markup:** used to determine the data types of the elements (date, time, person name, ..etc) or the role of the elements (author, editor, work name, ..etc).

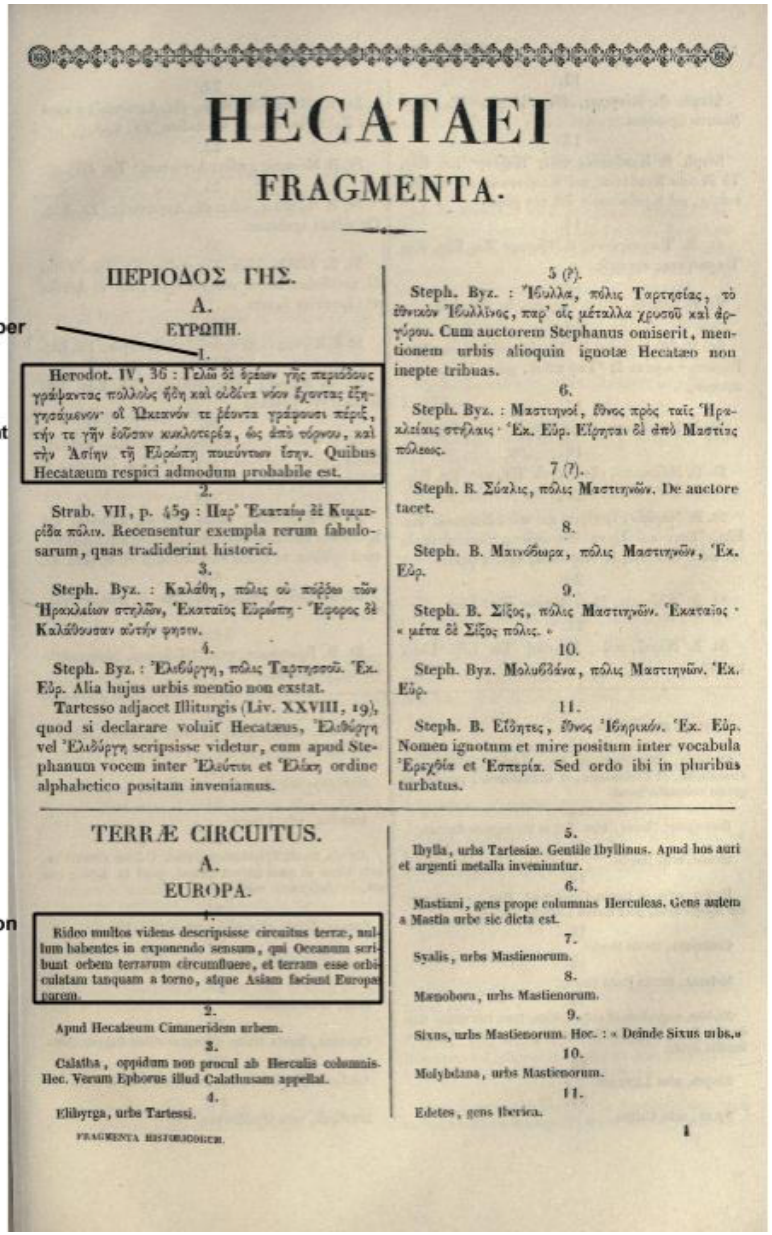
DFHG Collection

- 5 Volumes (available as image files)
- 600 Authors
- 3000 pages

- **Manual tagging is impractical in term of time and effort, especially when the data collection is huge.**

Automatic XML Tagging





HECATAEI

FRAGMENTA.

ΠΕΡΙΟΔΟΣ ΓΗΣ.

A.

ΕΥΡΩΠΗ.

1.

Herodot. IV, 36 : Γελῶ δὲ ὁρέων γῆς περιόδους γράψαντας πολλοὺς ἤδη καὶ οὐδένα νόον ἔχοντας ἐξηγησάμενον· οἱ Ὠκεανόν τε ῥέοντα γράφουσι πέριξ, τήν τε γῆν ἐοῦσαν κυκλοτερέα, ὡς ἀπὸ τόρνου, καὶ τὴν Ἀσίην τῇ Εὐρώπῃ οὐδὲν ἐλάσσω ποιεύντων ἴσην. Quibus Hecatæum respici admodum probabile est.

2.

Strab. VII, p. 459 : Παρ' Ἑκαταίῳ δὲ Κιμμερίδα πόλιν. Recensentur exempla rerum fabulosarum, quas tradiderint historici.

3.

Steph. Byz. : Καλάθη, πόλις οὐ πόρρω τῶν Ἡρακλείων στηλῶν, Ἑκαταῖος Εὐρώπῃ· Ἔφορος δὲ Καλάθουσαν αὐτὴν φησιν.

4.

Steph. Byz. : Ἐλιβύργη, πόλις Ταρτησσοῦ. Ἐκ Εὐρ. Alia hujus urbis mentio non exstat.

Tartesso adjacet Illiturgis (Liv. XXVIII, 19), quod si declarare voluit Hecatæus, Ἐλιβύργη vel Ἐλιβύργη scripsisse videtur, cum apud Stephanum vocem inter Ἐλεύτυα et Ἐλίκη ordine alphabetico positam inveniamus.

5 (?).

Steph. Byz. : Ἴδυλλα, πόλις Ταρτησίας, τὸ ἐθνικὸν Ἰδυλλῖνος, παρ' οἷς μέταλλα χρυσοῦ καὶ ἀργύρου. Cum auctorem Stephanus omiserit, mentionem urbis alioquin ignotæ Hecatæo non inepte tribuas.

6.

Steph. Byz. : Μαστιηνοί, ἔθνος πρὸς ταῖς Ἡρακλείαις στήλαις· Ἐκ. Εὐρ. Εἰρῆναι δὲ ἀπὸ Μαστίης πόλεως.

7 (?).

Steph. B. Σύαλις, πόλις Μαστιηνῶν. De auctore tacet.

8.

Steph. B. Μαινόβωρα, πόλις Μαστιηνῶν, Ἐκ. Εὐρ.

9.

Steph. B. Σίξος, πόλις Μαστιηνῶν. Ἑκαταῖος· « μετὰ δὲ Σίξος πόλις. »

10.

Steph. Byz. Μολυβδάνα, πόλις Μαστιηνῶν. Ἐκ. Εὐρ.

11.

Steph. B. Εἴδητες, ἔθνος Ἰβηρικόν. Ἐκ. Εὐρ. Nomen ignotum et mire positum inter vocabula Ἑρεχθία et Ἑσπερία. Sed ordo ibi in pluribus turbatus.

TERRÆ CIRCUITUS.

A.

EUROPA.

1.

Rideo multos videns descripsisse circuitus terræ, nullum habentes in exponendo sensum, qui Oceanum scribunt orbem terrarum circumfluere, et terram esse orbiculatam tanquam a torno, atque Asiam faciunt Europæ parem.

2.

Apud Hecatæum Cimmeridem urbem.

3.

Calatha, oppidum non procul ab Herculis columnis. Hec. Verum Ephorus illud Calathusam appellat.

4.

Elibyrga, urbs Tartessi.

5.

Ibylla, urbs Tartesiæ. Gentile Ibyllinus. Apud hos auri et argenti metalla inveniuntur.

6.

Mastiani, gens prope columnas Herculeas. Gens autem a Mastia urbe sic dicta est.

7.

Syalis, urbs Mastienorum.

8.

Mænobora, urbs Mastienorum.

9.

Sixus, urbs Mastienorum. Hec. : « Deinde Sixus urbs. »

10.

Molybdana, urbs Mastienorum.

11.

Edetes, gens Iberica.

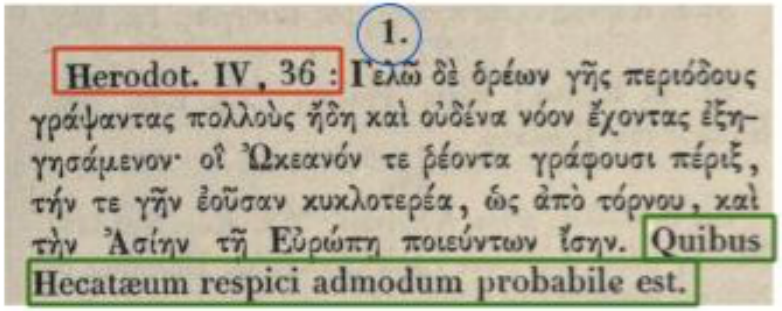
Fragment Number

Greek Fragment

Latin Translation

The plain texts were produced by an OCR.

Greek Fragment (Original Image)



1.

Herodot. IV, 36 : Γελῶ δὲ ὁρέων γῆς περιόδους γράψαντας πολλοὺς ἤδη καὶ οὐδένα νόον ἔχοντας ἐξηγησάμενον· οἳ Ὠκεανόν τε ῥέοντα γράφουσι πέριξ, τήν τε γῆν ἐοῦσαν κυκλοτερέα, ὡς ἀπὸ τόρνου, καὶ τὴν Ἀσίην τῇ Εὐρώπῃ ποιεύντων ἴσην. Quibus Hecatæum respici admodum probabile est.

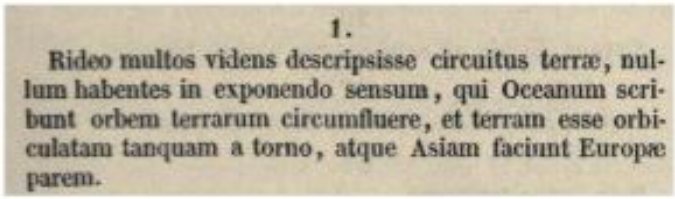
OCR Output

```
1.
<p>Herodot. IV, 36: Γελῶ δὲ ὁρέων γῆς περιόδους
γράψαντας πολλοὺς ἤδη καὶ οὐδένα νόον ἔχοντας ἐξηγησάμενον·
οἳ Ὠκεανόν τε ῥέοντα γράφουσι πέριξ,
τήν τε γῆν ἐοῦσαν κυκλοτερέα, ὡς ἀπὸ τόρνου, καὶ
τὴν Ἀσίην τῇ Εὐρώπῃ ποιεύντων ἴσην. Quibus
Hecatæum respici admodum probabile est.</p>
```

```
<cit n="1" xml:lang="grc">
        <bibl>Herodot. IV, 36</bibl>
        <quote>Γελῶ δὲ ὁρέων γῆς περιόδους γράψαντας
        πολλοὺς ἤδη καὶ οὐδένα νόον ἔχοντας ἐξηγησάμενον· οἳ
        Ὠκεανόν τε ῥέοντα γράφουσι πέριξ, τήν τε γῆν ἐοῦσαν
        κυκλοτερέα, ὡς ἀπὸ τόρνου, καὶ τὴν Ἀσίην τῇ Εὐρώπῃ
        ποιεύντων ἴσην.</quote>
        <note>Quibus Hecat&#230;um respici admodum
        probabile est. </note>
</cit>
```

XML Output

Latin Fragment (Original Image)



1.

Rideo multos videns descripsisse circuitus terræ, nullum habentes in exponendo sensum, qui Oceanum scribunt orbem terrarum circumfluere, et terram esse orbiculatam tanquam a torno, atque Asiam faciunt Europæ parem.

OCR Output

```
1.
<p>Rideo multos videns descripsisse circuitus
terræ, nullum habentes in exponendo sensum,
qui Oceanum scribunt orbem terrarum
circumfluere, et terram esse orbiculatam
tanquam a torno, atque Asiam faciunt Europæ
parem.</p>
```

```
<p n="1">
        Rideo multos videns descripsisse
        circuitus terræ, nullum habentes in
        exponendo sensum, qui Oceanum
        scribunt orbem terrarum circumfluere,
        et terram esse orbiculatam tanquam a
        torno, atque Asiam faciunt Europæ
        parem.
</p>
```

XML Output

Fragment Boundary Identification

OCR output:

- <body> .... </body>
- ...
- <pb n=”287”>

```
<p>Ante Ogygen nihil apud Græcos narratur dignum quod
memoretur, excepto Phoroneo, illius æquali, atque Inacho,
Phoronei patre, qui primus Argolidis rex erat, ut narrat
Acusilaus.
14.
<p>Acusilaus dicit Phoroneum fuisse primum hominum.
... Hinc Plato in Timæo secutus Acusilaum, scribit:</p>
</body>
<pb n="101">
```

```
<body>
ex Codd. Mss. edita: Δευκαλίων, ἐφ' οὗ ὁ κατακλυσμὸς
γέγονε, Προμηθέως μὲν ἦν υἱὸς, μητρὸς δὲ ὡς
πλεῖστοι λέγουσι Κλυμένης, ὡς δὲ Ἡσίοδος Πρυνείης,
ὡς δὲ Ἀκουσίλαος Ἠσιόνης τῆς Ὠκεανοῦ καὶ τοῦ
Προμηθέως.</p>
<p>Sturz. pro Ἠσιόνης, quæ h. l. Oceani filia dicitur,
legendum putat Ἀσίας et pro corrupto
illo Πρυνείης reponi vult Πανδώρας.</p>
<p>Schol. Pind. Ol. 9, 70: Κοινὰ τὰ περὶ Δευκαλίωνα
καὶ Πύῤῥαν, καὶ ὅτι τοὺς λίθους κατόπιν ῥίπτοντες
ἀνθρώπους ἐποίουν, μαρτυρεῖ Ἀκουσίλαος.</p>
8.
<p>Schol. Apollon. III, 1123: Ἀκουσίλαος καὶ Ἡσίοδος
ἐν ταῖς μεγάλαις Ἠοίαις φασὶν ἐξ Ἰοφώσσης τῆς
Αἰήτου (Phrixum genuisse Argum).</p>
9.
<p>Schol. Apoll. Rh. IV, 1147: Περὶ δὲ τοῦ δέρους,
ὅτι ἦν χρυσοῦν, οἱ πλεῖστοι ἱστοροῦσιν. Ἀκουσίλαος δὲ
ἐν τῷ περὶ γενεαλογιῶν πορφυρευθῆναι φησὶν ἀπὸ τῆς
θαλάσσης.</p>
10.
```

Fragment Boundary Identification

```
2.
<p>Etymol. M.: Κοῖος· ὁ πατὴρ Λητοῦς· παρὰ τὸ
κοεῖν, ὅ ἐστι νοεῖν καὶ συνιέναι, ὅ ἐστι συνετός. Κοῖόν
θ' Ὑπερίονά τ' Ἰαπετόν τε (ex Hesiodi Theog. 134).
Οἱ Αἰολεῖς τῷ κ ἀντὶ τοῦ ν κέχρηνται. Οὗτοι δὲ Τιτᾶνες
καὶ Τιτανίδες καλοῦνται, ὡς Ἀκουσίλαος.</p>
```

Language Detection

```
2.
<p>Etymol. M.: Κοῖος· ὁ πατὴρ Λητοῦς· παρὰ τὸ
κοεῖν, ὅ ἐστι νοεῖν καὶ συνιέναι, ὅ ἐστι συνετός. Κοῖόν
θ' Ὑπερίονά τ' Ἰαπετόν τε (ex Hesiodi Theog. 134).
Οἱ Αἰολεῖς τῷ κ ἀντὶ τοῦ ν κέχρηνται. Οὗτοι δὲ Τιτᾶνες
καὶ Τιτανίδες καλοῦνται, ὡς Ἀκουσίλαος.</p>
```

5 Latin words

36 Greek words

```
2.
<p>Cœus, pater Latonæ, sic nominatus a κοεῖν, quod idem
est ac νοεῖν et συνιέναι (scire, intelligere), ita ut κοῖος sit
quasi νοῖος prudens, intelligens. Nam Æoles litera κ pro
ν utuntur. Hesiodus: Cœum et Hyperiona et Iapetum.
Hi vero Titanes et Titanides vocantur, ut ait Acusilaus.</p>
```

39 Latin words

6 Greek words

Language Detection

```
2.
<p>Etymol. M.: Κοῖος· ὁ πατὴρ Λητοῦς· παρὰ τὸ
κοεῖν, ὅ ἐστι νοεῖν καὶ συνιέναι, ὅ ἐστι συνετός. Κοῖόν
θ' Ὑπερίονά τ' Ἰαπετόν τε (ex Hesiodi Theog. 134).
Οἱ Αἰολεῖς τῷ κ ἀντὶ τοῦ ν κέχρηνται. Οὗτοι δὲ Τιτᾶνες
καὶ Τιτανίδες καλοῦνται, ὡς Ἀκουσίλαος.</p>
```

```
<cit n="2" xml:lang="grc">Etymol. M.: Κοῖος· ὁ πατὴρ Λητοῦς· παρὰ τὸ
κοεῖν, ὅ ἐστι νοεῖν καὶ συνιέναι, ὅ ἐστι συνετός. Κοῖόν
θ' Ὑπερίονά τ' Ἰαπετόν τε (ex Hesiodi Theog. 134).
Οἱ Αἰολεῖς τῷ κ ἀντὶ τοῦ ν κέχρηνται. Οὗτοι δὲ Τιτᾶνες
καὶ Τιτανίδες καλοῦνται, ὡς Ἀκουσίλαος.</cit>
```

```
2.
<p>Cœus, pater Latonæ, sic nominatus a κοεῖν, quod idem
est ac νοεῖν et συνιέναι (scire, intelligere), ita ut κοῖος sit
quasi νοῖος prudens, intelligens. Nam Æoles litera κ pro
ν utuntur. Hesiodus: Cœum et Hyperiona et Iapetum.
Hi vero Titanes et Titanides vocantur, ut ait Acusilaus.</p>
```

```
<p n="2">Cœus, pater Latonæ, sic nominatus a κοεῖν, quod idem
est ac νοεῖν et συνιέναι (scire, intelligere), ita ut κοῖος sit
quasi νοῖος prudens, intelligens. Nam Æoles litera κ pro
ν utuntur. Hesiodus: Cœum et Hyperiona et Iapetum.
Hi vero Titanes et Titanides vocantur, ut ait Acusilaus.</p>
```

- We enclose every fragment in a <cit> element.
- We encode every fragment number by adding a @n attribute to the <cit> element.

Fragment’s Elements Identification

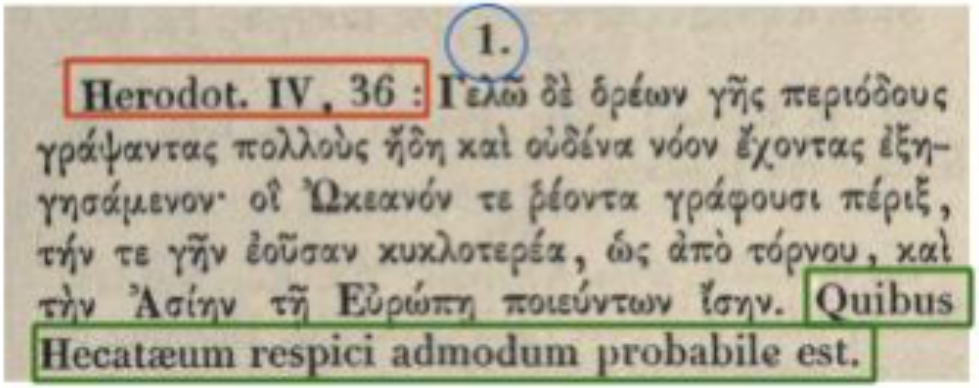
The Greek fragments usually contain the following elements:

- Fragment’s number;
- Bibliographic reference to the source text that preserves the fragment;
- Greek text of the fragment (= quotation from the source text);
- Brief note in Lat

Fragment’s Elements Identification

- We enclose the bibliographic reference to the source text within a <bibl> element.
- The Greek text of the fragment is in <quote>.
- If the Greek contains a word or a sentence in Latin, we encode it with <foreign xml:lang="la">...</foreign>.
- Latin notes at the end of a fragment sit in <note> with an attribute @xml:lang="la".

Fragment's Elements Identification


1.

Herodot. IV, 36 : Γελῶ δὲ ὁρέων γῆς περιόδους γράψαντας πολλοὺς ἤδη καὶ οὐδένα νόον ἔχοντας ἐξηγησάμενον· οἳ Ὠκεανόν τε ῥέοντα γράφουσι πέριξ, τήν τε γῆν ἐοῦσαν κυκλοτερέα, ὡς ἀπὸ τόρνου, καὶ τὴν Ἀσίην τῇ Εὐρώπῃ ποιεύντων ἴσην. Quibus Hecatæum respici admodum probabile est.

OCR Output

```
1.
<p>Herodot. IV, 36: Γελῶ δὲ ὁρέων γῆς περιόδους
γράψαντας πολλοὺς ἤδη καὶ οὐδένα νόον ἔχοντας ἐξηγησάμενον·
οἳ Ὠκεανόν τε ῥέοντα γράφουσι πέριξ,
τήν τε γῆν ἐοῦσαν κυκλοτερέα, ὡς ἀπὸ τόρνου, καὶ
τὴν Ἀσίην τῇ Εὐρώπῃ ποιεύντων ἴσην. Quibus
Hecatæum respici admodum probabile est.</p>
```

```
<cit n="1" xml:lang="grc">
        <bibl>Herodot. IV, 36</bibl>
        <quote>Γελῶ δὲ ὁρέων γῆς περιόδους γράψαντας
        πολλοὺς ἤδη καὶ οὐδένα νόον ἔχοντας ἐξηγησάμενον· οἳ
        Ὠκεανόν τε ῥέοντα γράφουσι πέριξ, τήν τε γῆν ἐοῦσαν
        κυκλοτερέα, ὡς ἀπὸ τόρνου, καὶ τὴν Ἀσίην τῇ Εὐρώπῃ
        ποιεύντων ἴσην.</quote>
        <note>Quibus Hecat&#230;um respici admodum
        probabile est. </note>
</cit>
```

XML Output

DEMO DEMO2

The Greek - Latin Dynamic Lexicon

Goals

- Create a Greek-Latin lexicon using English as pivot language.

- Identify the translation relationships among the words in the parallel fragments of the DFHG bilingual corpus.

- Provide training data for a word alignment system.

Large digitized Ancient Greek-Latin lexica are publicly unavailable.

Greek-Latin Dynamic Lexicon

Data source(Perseus Digital Library):

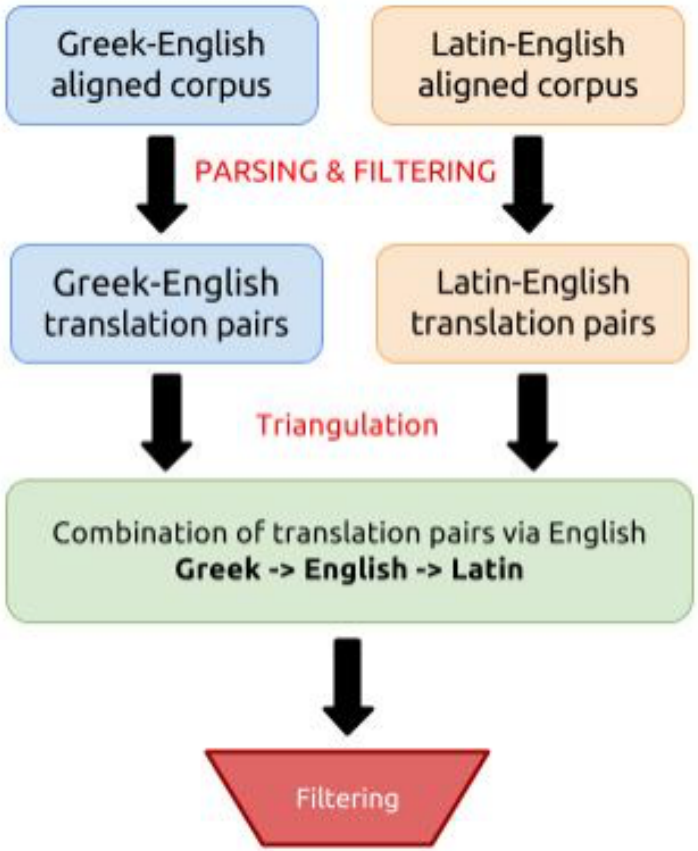
- Greek-English aligned texts (104 files)
- Latin-English aligned texts (59 files)

PERSEUS DIGITAL LIBRARY

The parallel texts are aligned on a sentence level using **Moore's Bilingual Sentence Aligner**. Then the **Giza++** toolkit is used to align the sentence pairs at the level of individual words.

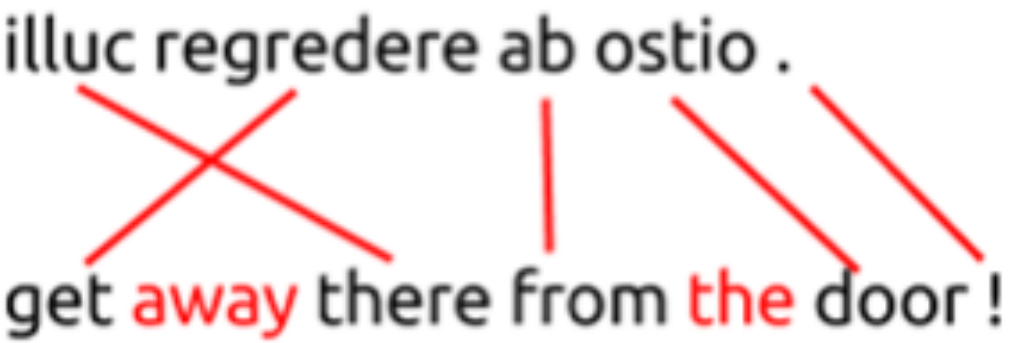
	Files	Sentences	Words	Distinct pairs
Ancient Greek	104	210 K	4,32 M	43 K
Latin	59	132 K	2,33 M	39 K

The proposed method



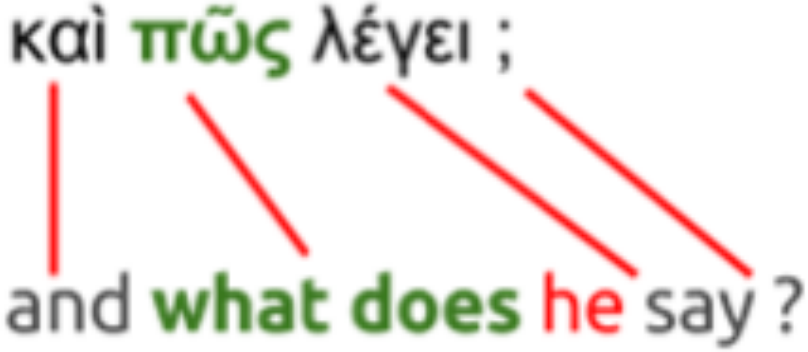
One-to-One Alignment

```
<sentence id="6">
  <wds lnum="L1">
     <w n="6-1" nrefs="6-3">illuc </w>
     <w n="6-2" nrefs="6-1">regredere</w>
     <w n="6-3" nrefs="6-4">ab</w>
     <w n="6-4" nrefs="6-6">ostio </w>
     <w n="6-5" nrefs="6-7">. </w>
  </wds>
  <wds lnum="L2">
     <w n="6-1" nrefs="6-2">get</w>
     <w n="6-2" nrefs="">away</w>
     <w n="6-3" nrefs="6-1">there </w>
     <w n="6-4" nrefs="6-3">from</w>
     <w n="6-5" nrefs="">the</w>
     <w n="6-6" nrefs="6-4">door</w>
     <w n="6-7" nrefs="6-5">!</w>
  </wds>
</sentence>
```



One-to-many Alignment

```
<sentence id="8" >
   <wds lnum="L1">
      <w n="61-1" nrefs="8-1">καὶ</w>
      <w n="61-2" nrefs="8-2 8-3">πῶς</w>
      <w n="61-3" nrefs="8-5">λέγει</w>
      <w n="61-4" nrefs="8-6">;</w>
   </wds>
   <wds lnum="L2">
      <w n="61-1" nrefs="8-1">and</w>
      <w n="61-2" nrefs="8-2">what</w>
      <w n="61-3" nrefs="8-2">does</w>
      <w n="61-4" nrefs="">he</w>
      <w n="61-5" nrefs="8-3">say</w>
      <w n="61-6" nrefs="8-4">?</w>
    </wds>
</sentence>
```



Alignment errors

The Latin word **(et)** occurred 55034 times in the text and is aligned to:

Word	Frequency	Percentage
and	48185	88%
;	1422	2,5%
; and	1145	2%
,	818	1,5%
, and	545	1%
other translations	1919	4%

Source of Errors:
The alignment is noisy and done automatically by the Giza++ software

Alignment errors filtering

- Delete all entries that are punctuation or special characters on one side aligned to words on the other side.

- Delete translation pairs with probability less than 1%, because those pairs are noisy and incorrect.

Word	Frequency	Percentage
and	48185	88%
;	1422	2,5%
; and	1145	2%
,	818	1,5%
, and	545	1%
other translations	1919	4%

Lemmatization

Latin and Greek are highly inflected languages, whereas English is considered a weakly inflected language.

eius cupio filiam virginem mihi desponderi.
i' d like his maiden daughter to be promised me in marriage.

piliae et filiae salutem.
love to pilia and your daughter.

mea haec erilis gestitavit filia.
this daughter of my mistress had them.

Lemmatization

Latin and Greek are highly inflected languages, whereas English is considered a weakly inflected language.

eius cupio filiam virginem mihi desponderi.
i' d like his maiden daughter to be promised me in marriage.

piliae et filiae salutem.
love to pilia and your daughter.

mea haec erilis gestitavit filia.
this daughter of my mistress had them.

Lemmatization solves this problem

filiam, filiae, filia → daughter

filia → daughter

Lemmatization

Latin and Greek are highly inflected languages, whereas English is considered a weakly inflected language.

καὶ ἐπειδὴ ἔτεκεν υἱόν, ἔξαρνος ἦν μὴ εἶναι ἐξ αὐτοῦ τὸ παιδίον
but when she gave birth to a son, (callias) denied that the child was his

υἱὸν δὲ Τεισίου τοῦ Ῥαμνουσίου
the son of teisias of rhamnos

οὑμὸς υἱός
my son

Lemmatization solves this problem

υἱὸν, υἱός, υἱόν → son

υἱός → son

MORPHEUS

Lemmatization of Latin and Greek words is done using **Morpheus**.

Morpheus is a morphological parsing and lemmatizing tool for Ancient Greek and Latin produced by the Perseus Project and available as a **web service.**

It takes as input a token (Greek or Latin) and returns the lemma from which this token derives, and the full morphological information (part of speech, case, number, gender, person, mood, etc.).

Lemmatization

Lemmatization of English translations will produce better results.

E.g., translation candidates of the Greek word **λέγειν**

Translation	Freq	Precentage
say	551	36%
speak	492	32%
tell	149	9.7%
speaking	110	7%
said	89	6%
saying	54	3.5%
mention	45	2.9%
says	25	1.5%
spoke	19	1.2

Translation probabilities

Translation	Freq	Precentage
say	551	36%
speak	492	32%
tell	149	9.7%
speak	110	7%
say	89	6%
say	54	3.5%
mention	45	2.9%
say	25	1.5%
speak	19	1.2

Lemmatization of English words

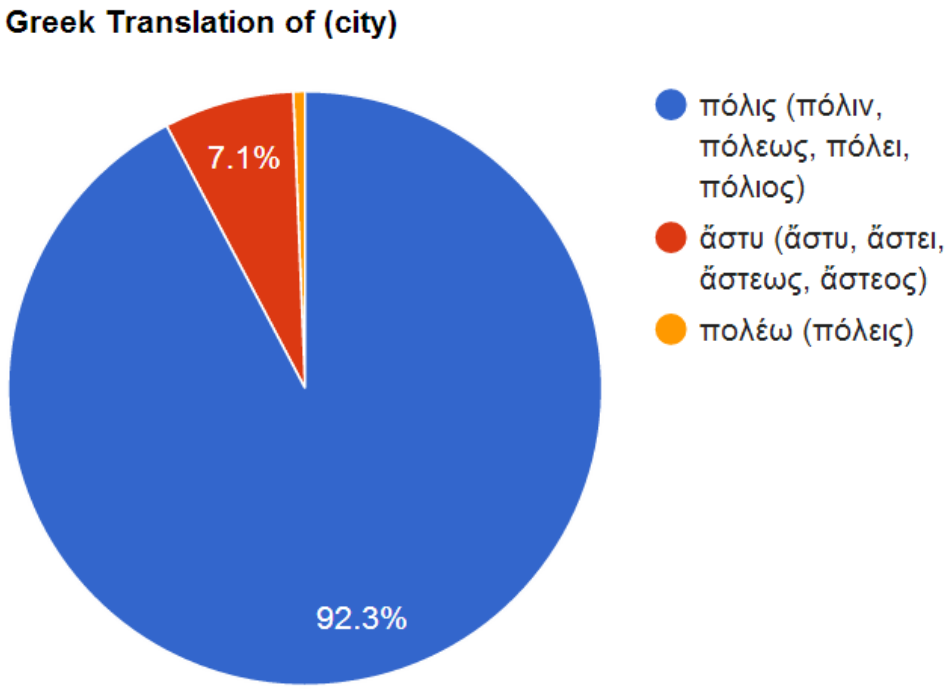
Translation	Freq	Precentage
say	710	46.8%
speak	621	40.6%
tell	149	9.7%
mention	45	2.9%

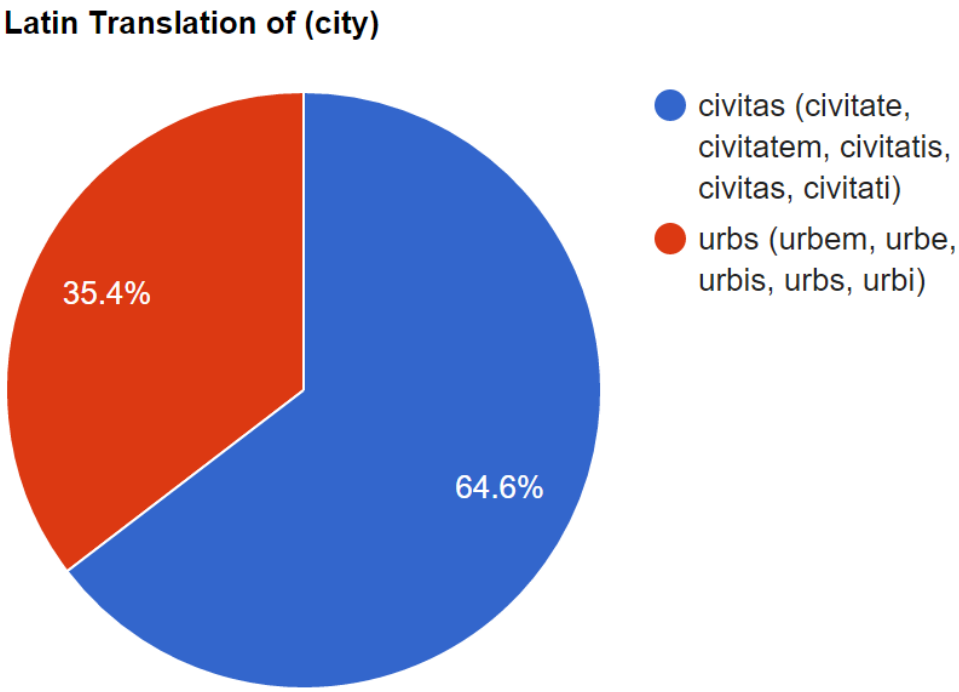
Group the results and recalculate the probabilities

Triangulation

Triangulation is based on the assumption that two expressions are likely to be translations if they are translations of the same word in a third language.

DEMO





Triangulation

- The English word (**city**) is translated to (92.3% **πόλις)**, (7.1% **ἄστυ)**, (0.6% **πολέω)**
  (**city**) is translated to (64,6% **civitas)**, (35.4% **urbs**)

- The extracted pairs via triangulation:
  (**πόλις** -**civitas**), (**πόλις** -**urbs**), (**ἄστυ**-**civitas**), (**ἄστυ**-**urbs**),(**πολέω**-**civitas**), (**πολέω**-**urbs**).

  those pairs don't have the same level of relatedness,
  therefore we have to filter the results to keep only strong related pairs

Translation-Pairs filtering

We used **Jaccard coefficient** as a similarity metric to measure the similarity or the relatedness between every Greek-Latin pairs.

$$J(A, B)=\frac{A \cap B}{A \cup B}$$

The relatedness between the Greek word (**πόλις**) and the Latin word (**civitas**) can be calculated as follows:

Translation-Pairs filtering

πόλις		
city	7432	74 %
state	1911	19 %
town	366	3.5 %
athens	357	3.5 %

πολέω		
city	6673	64.3 %
of	2352	22.6 %
state	823	7.9 %
town	530	5.2 %

ἄστυ		
city	6747	91 %
citizen	399	4 %
town	236	3 %
others	96	1%

urbs		
city	44395	94.2 %
rome	2190	4,6 %
town	531	1.2 %

civitas		
city	8748	72.9 %
state	2340	19.5 %
citizenship	420	3.5 %
citizen	372	3.1 %

Translation-Pairs filtering

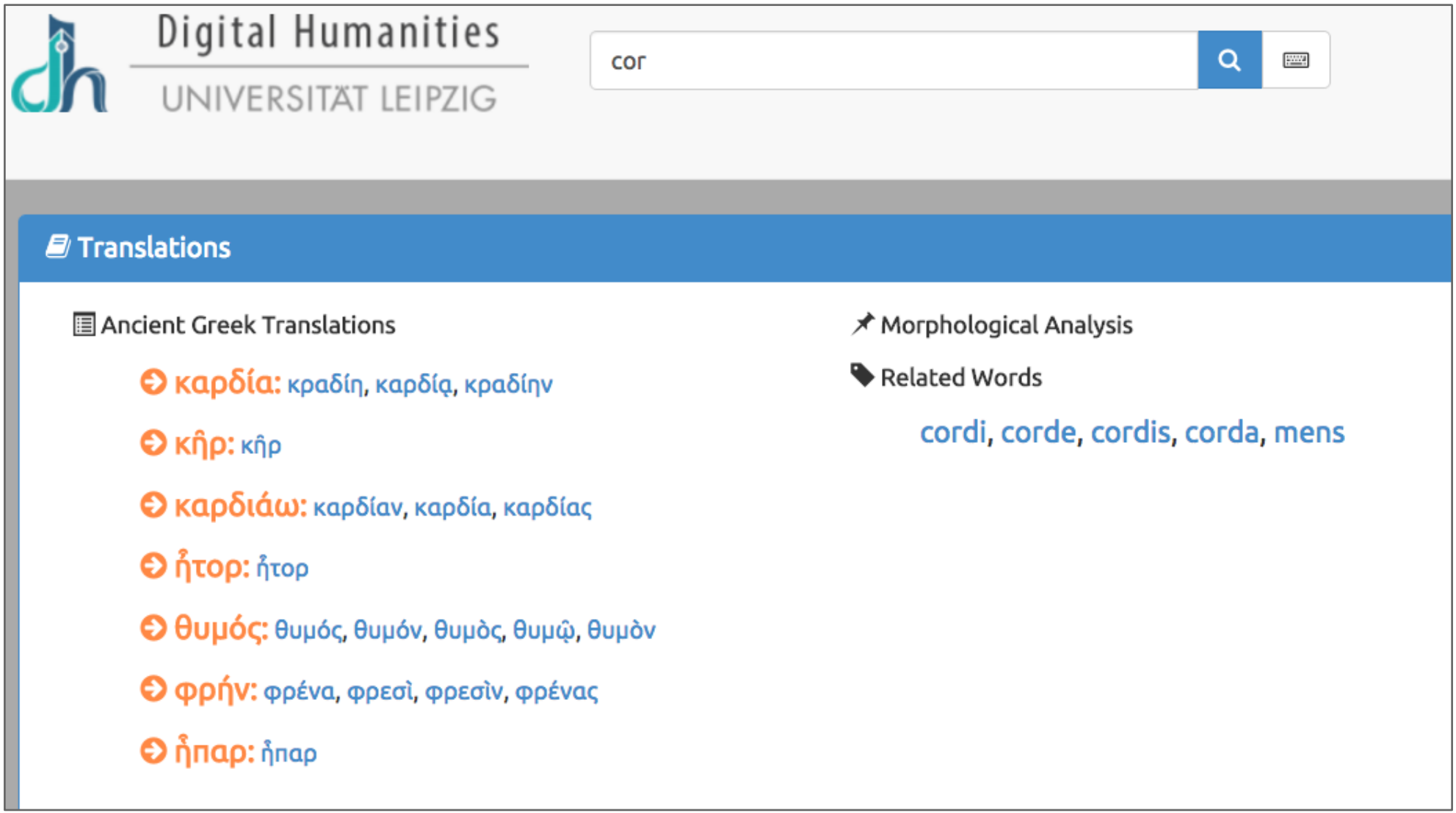
πόλις - civitas	city, state	(74 + 19 + 72 + 19.5)/200 = 92.25 %		92.25 %
πόλις - urbs	city, town	(74 + 3.5 + 94.2 + 1.2)/200= 86.45 %		86.45 %
πολέω - civitas	city, state	(64.3 + 7.9 + 72 + 19.5)/200 = 81.85 %		81.85 %
πολέω - urbs	city, town	(64.3 + 5.2 + 94.2 + 1.2)/200= 82.45 %		82.45 %
ἄστυ - civitas	city, citizen	(91+ 4 + 72 + 3.1)/200 = 85.05 %		85.05 %
ἄστυ - urbs	city, town	(91 + 3 + 94.2 + 1.2)/200= 94.7 %		94.7 %

Translation-Pairs filtering

Ρώμη		
rome	8044	100 %

urbs		
city	44395	94.2 %
rome	2190	4,6 %
town	531	1.2 %

J(**Ρώμη, urbs**) = (100+4.6)/200 = 51.3 %


Digital Humanities
UNIVERSITÄT LEIPZIG
cor
Translations
Ancient Greek Translations
καρδία: κραδίη, καρδίᾳ, κραδίην
κῆρ: κῆρ
καρδιάω: καρδίαν, καρδία, καρδίας
ἦτορ: ἦτορ
θυμός: θυμός, θυμόν, θυμὸς, θυμῷ, θυμὸν
φρήν: φρένα, φρεσὶ, φρεσὶν, φρένας
ἧπαρ: ἧπαρ
Morphological Analysis
Related Words
cordi, corde, cordis, corda, mens

heart

✔ cor - κῆρ [κῆρ] (0,02)

ἤχθετο γὰρ κῆρ .
on this he groaned aloud ,and called upon his friend by name .

✔ cor - καρδίαν [καρδιάω] (0,04)

κἀμὲ καρδίαν ἀμύσσει φροντίς ·
my heart ,too ,is racked with anxiety ,and to you ,my friends ,will i make a disclosure .

✔ cor - κραδίην [καρδία] (0,07)

νηπύτι᾽ ὡς ἄνοον κραδίην ἔχες ·
fool ,how witless is the heart thou hast !

✔ cor - καρδίᾳ [καρδία] (0,07)

τί ὅτι ἔθου ἐν τῇ καρδίᾳ σου τὸ πρᾶγμα τοῦτο ;
how is it that you have conceived this thing in your heart ?

✔ cor - κραδίη [καρδία] (0,07)

" τέτλαθι δή ,κραδίη ·
" endure ,my heart ;

Evaluation

Mean reciprocal rank (MRR)

$$MRR = \frac{1}{|Q|}\sum_{i=1}^{|Q|} RR_i$$

We selected randomly 200 translation pairs obtained via the proposed method with different frequencies (high and low) and different JACCARD Co values.

Each pair should be assigned into four categories: **Correct**, **small difference, big difference and incorrect**.

Evaluation

Category	Reciprocal Rank
Correct	1
Small difference	0.75
Big difference	0.25
Uncorrect	0

Jaccard co	60% <	70% <	80% <	90% <
MMR	61.25 %	74 %	87.5 %	94.5 %

Conclusion

The quality of the method depends on two factors:

- **The size of aligned-parallel corpora**

Bigger corpora produce better translation probability distribution and more translation candidates.

- **The quality of the aligner**

Manually aligned corpora yield more accurate results.

Named-Entities Recognition

Named-Entities Recognition

Named-Entity Recognition (NER) as the task of **locating** and **classifying** single words and phrases in the text into predefined categories such as **person names**, **locations**, **organizations** as well as quantities, percentage, time expressions and monetary values.

Jim bought 300 shares of Acme Corp. in 2006.

$[\text{Jim}]_{\text{Person}}$ bought 300 shares of $[\text{Acme Corp.}]_{\text{Organization}}$ in $[\text{2006}]_{\text{Time}}$.

Entities Identification

Τὸ **Σικελικὸν** γένος ἐξέλιπεν **Ἰταλίαν**· ὡς μὲν **Ἑλλάνικος** ὁ **Λέσβιός** φησι, τρίτῃ γενεᾷ πρότερον τῶν Τρωϊκῶν, **Ἀλκυόνης** ἱεραμένης ἐν **Ἄργει** κατὰ τὸ ἕκτον καὶ εἰκοστὸν ἔτος. Δύο γὰρ ποιεῖ στόλους **Ἰταλικοὺς** διαβάντας εἰς **Σικελίαν**· τὸν μὲν πρότερον, **Ἐλύμων**, οὕς φησιν ὑπὸ Οἰνώτρων ἐξαναστῆναι· τὸν δὲ μετὰ τοῦτον ἔτει πέμπτῳ γενόμενον, **Αὐσονίων Ἰάπυγας** φυγόντων· βασιλέα δὲ τούτων ἀπαφαίνει **Σικελὸν**, ἀφ᾽ οὗ τοὔνομα τοῖς τε ἀνθρώποις καὶ τῇ νήσῳ τεθῆναι.

Genus **Siculum Italiam** reliquit, ut refert **Hellanicus Lesbius**, tribus ætatibus ante bellum Trojanum, anno vigesimo sexto sacerdotii quod **Alcyone Argis** exercebat. Duas enim classes **Italicas** in **Siciliam** trajecisse tradit: priorem **Elymorum**, quos ab Œnotris pulsos dicit: posteriorem vero **Ausonum Iapygas** fugientium, quæ quinto post anno eo trajecit. Horum autem regem **Siculum** fuisse dicit, a quo tam ipsi genti quam insulæ nomen inditum.

Entities Identification

Τὸ **Σικελικὸν** γένος ἐξέλιπεν **Ἰταλίαν**· ὡς μὲν **Ἑλλάνικος** ὁ **Λέσβιός** φησι, τρίτῃ γενεᾷ πρότερον τῶν Τρωϊκῶν, **Ἀλκυόνης** ἱεραμένης ἐν **Ἄργει** κατὰ τὸ ἕκτον καὶ εἰκοστὸν ἔτος. Δύο γὰρ ποιεῖ στόλους **Ἰταλικοὺς** διαβάντας εἰς **Σικελίαν**· τὸν μὲν πρότερον, **Ἐλύμων**, οὕς φησιν ὑπὸ Οἰνώτρων ἐξαναστῆναι· τὸν δὲ μετὰ τοῦτον ἔτει πέμπτῳ γενόμενον, **Αὐσονίων Ἰάπυγας** φυγόντων· βασιλέα δὲ τούτων ἀπαφαίνει **Σικελὸν**, ἀφ᾽ οὗ τοὔνομα τοῖς τε ἀνθρώποις καὶ τῇ νήσῳ τεθῆναι.

Genus **Siculum Italiam** reliquit, ut refert **Hellanicus Lesbius**, tribus ætatibus ante bellum Trojanum, anno vigesimo sexto sacerdotii quod **Alcyone Argis** exercebat. Duas enim classes **Italicas** in **Siciliam** trajecisse tradit: priorem **Elymorum**, quos ab Œnotris pulsos dicit: posteriorem vero **Ausonum Iapygas** fugientium, quæ quinto post anno eo trajecit. Horum autem regem **Siculum** fuisse dicit, a quo tam ipsi genti quam insulæ nomen inditum.

Entities Identification

- All proper names are capitalized, but not all capitalized words are proper names

- The POS of a NE must be **Noun** or **UNKNOWN**

- Since the Latin fragments are translations of the Ancient Greek fragments, each proper name in the Ancient Greek fragment has a correspondence in the Latin parallel fragment, and they are similar from a syntactic or phonetic point of view.

Entities Identification

In order to calculate the similarity between the translation pairs, we need a metric to measure the similarity. **Levenshtein Edit Distance** is the most popular metric

F1: Token begins with a capital letter

F2: POS tag

F3: Levenshtein distance

Words pair	F1	F2	F3
Ἀθήνας - Athenas	1	Noun - Noun	0
Ἄργος - Argos	1	Noun - Noun	0
Ἑλλάνικος - Hellanico	1	Noun / Unknown	1

Transliteration

converting the text from one writing system to another.

GRC	EN	GRC	EN	GRC	EN
Α, α	A, a	**Ν, ν**	N, n	**Ι, ι**	I, i
Β, β	B, b	**Ξ, ξ**	X, x	**Κ, κ**	K, k
Γ, γ	G, g	**Ο, ο**	O, o	**Λ, λ**	L, l
Δ, δ	D, d	**Π, π**	P, p	**Μ, μ**	M, m
Ε, ε	E, e	**Ρ, ρ**	R, r	**Φ, φ**	Ph, ph
Ζ, ζ	Z, z	**Σ, σ**	S, s	**Χ, χ**	Ch, ch
Η, η	Ē, ē	**Τ, τ**	T, t	**Ψ, ψ**	Ps, ps
Θ, θ	Th, th	**Υ, υ**	Y, y	**Ω, ω**	Ō, ō

Source Text (Ancient Greek)	**Τύραητα**, πόλις **Σαυνιτῶν**. **Φίλιστος** ἐνδεκάτῳ.
Latin Transliteration	**Týraētа**, pólis **Saunitō̂n**. **Phílistos** hendekátōi.
Transliteration (without accent)	**Tyraeta**, polis **Sauniton**. **Philistos** hendekatoi.
Target Text (Latin)	**Tyrseta**, urbs **Samnitum**. **Philistus** undecimo.

Levenshtein Distance

Levenshtein distance is the most common metric for measuring the distance between two words or strings in terms of minimum number of insertion, deletion and substitution required to change one word or sequence into the other

$$\operatorname{lev}_{a,b}(i,j) = \begin{cases} \max(i,j) \\ \min \begin{cases} \operatorname{lev}_{a,b}(i-1,j)+1 \\ \operatorname{lev}_{a,b}(i,j-1)+1 \\ \operatorname{lev}_{a,b}(i-1,j-1)+1_{(a_i \neq b_j)} \end{cases} \end{cases}$$

Ancient Greek	Latinized Greek	Latin	Levenshtein Distance
Φίλιστος	Philistos	Philistus	1
Ἑλλάνικος	Hellanikos	Hellanicus	2
Δαμασκὸν	Damaskos	Damascus	2
Κύπρος	kypros	Cyprus	2

Modified Levenshtein Distance

Levenshtein distance is not very helpful in our case, because it is binary and there is no tolerance with transliteration errors.

the distance between letters is not binary, but it is on scale. The cost of insertion or deletion depends on:

- Letter position (part of the prefix or suffix)
- Letter type (vowel or consonant)

Group elements	Distance within group elements
E, e Ē, ē I, i Y, y	0.15
Ō, ō O, o U,u Y, y	0.15
C, c K, k Q, q	0.15
S, s C, c	0.25
M, m N, n	0.35

Modified Levenshtein Distance

Ancient Greek	Latinized Greek	Latin	Levenshtein Distance	Modified Levenshtein	*Dist/Length*
Φίλιστος	Philistos	Philistus	1	0.15	0,0167
Ἑλλάνικος	Hellanikos	Hellanicus	2	0.30	0.03
Δαμασκὸν	Damaskos	Damascus	2	0.30	0.0375
Κύπρος	kypros	Cyprus	2	0.30	0.05

Relations between Entities

Τὸ **Σικελικὸν** γένος ἐξέλιπεν **Ἰταλίαν**· ὡς μὲν **Ἑλλάνικος** ὁ **Λέσβιός** φησι, τρίτῃ γενεᾷ πρότερον τῶν Τρωϊκῶν, **Ἀλκυόνης** ἱεραμένης ἐν **Ἄργει** κατὰ τὸ ἕκτον καὶ εἰκοστὸν ἔτος. Δύο γὰρ ποιεῖ στόλους **Ἰταλικοὺς** διαβάντας εἰς **Σικελίαν**· τὸν μὲν πρότερον, **Ἐλύμων**, οὕς φησιν ὑπὸ Οἰνώτρων ἐξαναστῆναι· τὸν δὲ μετὰ τοῦτον ἔτει πέμπτῳ γενόμενον, **Αὐσονίων Ἰάπυγας** φυγόντων· βασιλέα δὲ τούτων ἀπαφαίνει **Σικελὸν**, ἀφ᾽ οὗ τοὔνομα τοῖς τε ἀνθρώποις καὶ τῇ νήσῳ τεθῆναι.

Genus **Siculum Italiam** reliquit, ut refert **Hellanicus Lesbius**, tribus ætatibus ante bellum Trojanum, anno vigesimo sexto sacerdotii quod **Alcyone Argis** exercebat. Duas enim classes **Italicas** in **Siciliam** trajecisse tradit: priorem **Elymorum**, quos ab Œnotris pulsos dicit: posteriorem vero **Ausonum Iapygas** fugientium, quæ quinto post anno eo trajecit. Horum autem regem **Siculum** fuisse dicit, a quo tam ipsi genti quam insulæ nomen inditum.

Relations between Entities

Phrase	Frequency in the text	Category
Hellanicus Lesbius	6	Author Name
Charon Lampsacenus	5	Author Name
Antiochus Syracusanus	4	Author Name
Acusilaus Argivus	1	Author Name

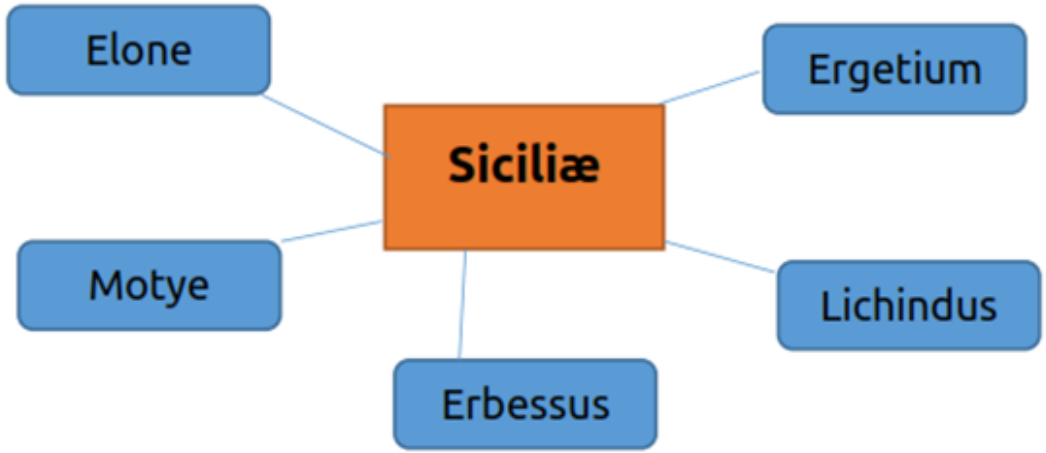
Relations between Entities

Phrase	First NE	Relation	Second NE
Apollinis **ex** Callisto	Apollinis	ex	Callisto
Phœnissas **ex** Cypro	Phœnissas	ex	Cypro
Ergetium**, urbs** Siciliæ	Ergetium	, urbs	Siciliæ
Lichindus**, urbs** Siciliæ	Lichindus	, urbs	Siciliæ
Erbessus**, urbs** Siciliæ	Erbessus	, urbs	Siciliæ
Motye**, urbs** Siciliæ	Motye	, urbs	Siciliæ
Elone**, urbs** Siciliæ	Elone	, urbs	Siciliæ
Hellanicus **libro primo** Deucalioniæ	Hellanicus	libro primo	Deucalioniæ
Hellanico **libro primo** Lesbicorum	Hellanico	libro primo	Lesbicorum
Hellanico **libro primo** Troicorum	Hellanico	libro primo	Troicorum
Marsyas **libro primo** Macedonicorum	Marsyas	libro primo	Macedonicorum
Clidemus **libro primo** Protogoniæ	Clidemus	libro primo	Protogoniæ
Clidemus **libro primo** Atthidis	Clidemus	libro primo	Atthidis

Table 9.11 Relation between NEs

Relations between Entities

Ergetium, **urbs** Siciliæ	Ergetium	, urbs	Siciliæ
Lichindus, **urbs** Siciliæ	Lichindus	, urbs	Siciliæ
Erbessus, **urbs** Siciliæ	Erbessus	, urbs	Siciliæ
Motye, **urbs** Siciliæ	Motye	, urbs	Siciliæ
Elone, **urbs** Siciliæ	Elone	, urbs	Siciliæ



Named-Entities Classification

PLEIADES is a gazetteer containing information about ancient world places and their locations and names.

http://pleiades.stoa.org/

Named-Entities Classification

A Dictionary of Greek and Roman Antiquities (1890)

William Smith, LLD, William Wayte, G. E. Marindin, Ed

http://www.perseus.tufts.edu/hopper/text?doc=Perseus%3atext%3a1999.04.0063

GeoNames

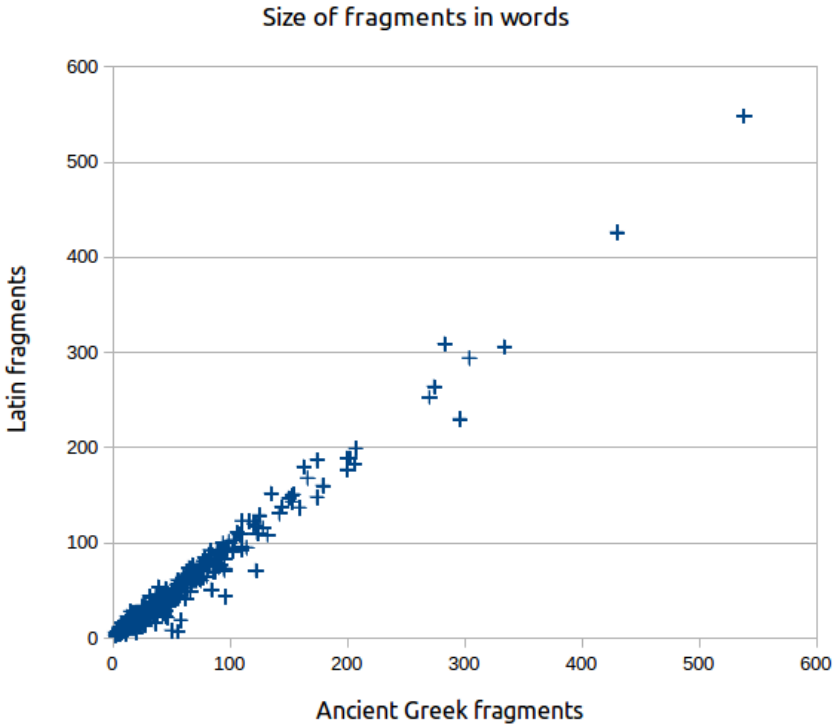
The GeoNames geographical database covers all countries and contains over eight million placenames that are available for download free of charge.

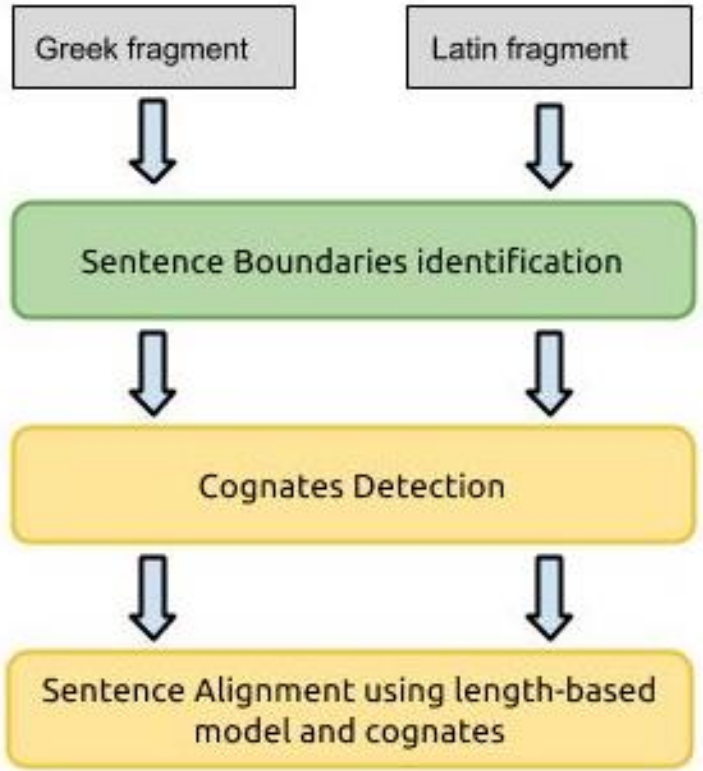
http://www.geonames.org/

Word Alignment

Word Alignment

1. Sentence Boundary Identification
2. Sentence Alignment





Word Alignment model

Μακτώριον, πόλις Σικελίας· Φίλιστος πρώτῳ, ἣν ἔκτισε μόνην. Τὸ ἐθνικὸν, Μακτωρῖνος.

Mactorium, urbs Siciliæ, de qua Philistus libro primo; quam condidit solam. Gentile, Mactorinus.

1. **Entities Alignment** *(Transliteration & Modified Levenshtein distance)*

	Μακτώριον	πόλις	Σικελίας	Φίλιστος	πρώτῳ	ἣν	ἔκτισε	μόνην	Τὸ	ἐθνικὸν	Μακτωρῖνος
Mactorium											
urbs											
Siciliæ											
de											
qua											
Philistus											
libro											
primo											
quam											
condidit											
solam											
Gentile											
Mactorinus											

2. Dynamic Lexicon pairs *(Jaccard coefficient)*

	Μακτώριον	πόλις	Σικελίας	Φίλιστος	πρώτῳ	ἦν	ἔκτισε	μόνην	Τὸ	ἐθνικὸν	Μακτωρῖνος
urbs	0	0.72	0	0	0	0	0	0	0	0	0

2. Dynamic Lexicon pairs *(Jaccard coefficient)*

	Μακτώριον	πόλις	Σικελίας	Φίλιστος	πρώτῳ	ἦν	ἔκτισε	μόνην	Τὸ	ἐθνικὸν	Μακτωρῖνος
Mactorium											
urbs											
Siciliæ											
de											
qua											
Philistus											
libro											
primo											
quam											
condidit											
solam											
Gentile											
Mactorinus											

3. Co-occurrence Significance

Biemann and Quasthoff, 2005

Given two words A (Greek), B (Latin), each occurring a, b times in sentences (respectively), and k times together. We calculate the significance sig(A, B) of their occurrence (trans-co-occurrence) as follows:

$$sig(A,B) = \frac{x - k \log x + \log k!}{\log n}$$

$$with \quad x = \frac{ab}{n} \quad , n: number\ of\ sentences$$

For instance, (**πόλις**) occurs 55 times in the Greek fragments, and (**urbs**) occurs 57 times in the Latin fragments, and they occur 45 times

Sig (πόλις, urbs) = 19,36

	libro	primo	quam	condidit	solam	**Gentile**	Mactorinus
ἐθνικὸν	1.58	0.15	0.14	0.38	0.63	**6.39**	0.61

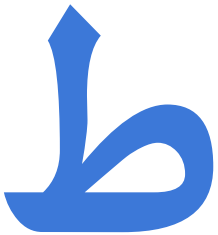
3. Co-occurrence Significance

Biemann and Quasthoff, 2005

	Μακτώριον	πόλις	Σικελίας	Φίλιστος	πρώτῳ	ἣν	ἔκτισε	μόνην	Τὸ	ἐθνικὸν	Μακτωρῖνος
Mactorium											
urbs											
Siciliæ											
de											
qua											
Philistus											
libro							0.15			1.58	
primo							0.18			0.15	
quam							1.04			0.14	
condidit							**2.85**			0.38	
solam							0.71			0.63	
Gentile							0.38			**6.39**	
Mactorinus							0.71			0.61	

Morphological Disambiguator

Demo




THANKS!