Francesco Stella con Emmanuela Carbé (Univ. of Siena) From ALIM to the Eurasian Latin Archive: The computing analysis of Latin texts through multilingual archives

> Milan, LiLa Workshop 2019, 3-4 June

Centro Studi Comparati UNISI (www.centroideugsu.unisi.it)



DECLAM II





II Centro

Ricerca

Didattica

Pubblicazioni

Semicerchio

Progetti

Biblioteca

News

Il Centro



Porte de Séoul, Dessin de Taylor, d'après une photographie

Il Centro Studi Comparati del Dipartimento di Filologia e critica delle letterature antiche e moderne nasce nel 1994 per iniziativa del medievista e comparatista coreano I Deug-Su, direttore del Centro dal 1994 al 2003.

Promuove ricerche e pubblicazioni di cultura coreana e letteratura comparata, di medievistica e di filologia digitale.

L'archivio del Centro, nelle stanze 414 e 415 del DFCLAM, ospita in deposito provvisorio, per la didattica e la consultazione di studenti e ricercatori, circa 4000 fra estratti, riviste e libri in originale o in fotocopia e quasi 1000 fotografie da manoscritti e cd-rom di argomenti classici e medievali.

Le ricerche del Centro hanno ottenuto finanziamenti da enti extraeuropei (Korea Foundation, National University of Seoul), dalla Commissione Europea (2 progetti Ariane, 1 programma TMR, 1 Creative Culture), dal Ministero italiano

della ricerca e dell'Università (3 PRIN come coordinatori nazionali, 2 come unità locali, 1 progetto Interlink con 4 università estere), dal CNR (2 contributi CNR

Digital projects: from 2007 on

II Centro	Ricerca	Didattica	Pubblicazioni	Semicerchio	Progetti			
	Cultura coreana							
Filologia di	Cultura medioevale							
i ilologia ul	Comparatistica							
For the state of A to the state of the	Filologia digitale		Lexicon					
Fra le attività più innovativ	re del Cellulo di		ALIM	orazione informatica.				
ALIM, Archivio della lati edizioni critiche o stampe			Corpus Rhythmorum Musicum	digitale di testi italiani nazionali. ALIM è consu				
analizzabile da motori di ri			Lo studium aretino del '200	su progetto di Franceso	co Stella) e dal			
primo lemmatizzatore ope	n access di latin	o medievale	Eurasian Latin Archive					
Frutto di un progetto euro edizione critica digitale del sia giunta la musica; è onl criticamente ricostruiti, le corredati da una schedatu	le poesie ritmici ine ad accesso l trascrizioni di tu	he latine mus ibero e in ediz itti i testimoni	zione cartacea con CD-RC i (140) e le riproduzioni f	OM (Sismel 2007): offre i otografiche di ogni mano	opee di cui ci testi scritto (80),			

Some financed projects

2003, 2007, 2009: **DIGIMED and GALILEO** projects with Salamanca, Berlin, Paris ENCh, London KC, Burgos etc. (MIUR)

2007 CRM (Corpus Rhythmorum Musicum- digital:

Latin songs with music); free online, cd-rom with pictures publ. by SISMEL (PRIN, European Commission TMR)

2008: Studium, a Virtual Exhibition on Medieval University in Arezzo (UNISI-Comune-Regione-Banca Etruria)

2012 ff. ALIM Archivio della Latinità Italiana del Medioevo alim unisi.it (medieval Latin texts and documents from Italy in encoded critical edition) open access, open source, including eds. principes (+ Univ. of VE,VR, NA, PA) Funds: PRIN

2014 Sofware Lexicon www.lexicon.unisi.it (L. Tessarolo)

2016 Semi-automatic Lemmatizer of Medieval Latin (S. Arrigoni)

2015-17 European Project **DEMM** (Dig.Ed. of Med. Mss.) **Erasmus+**

2019-2020 *Eurasian Latin Archive* (Carbé-Giannelli)

Funds: Regione Toscana

Linguistic analysis in the CRM (2007 book+cdrom, 2011 online corrected version, 2016 new contents)



CORPUS RHYTHMORUM MUSICUM (SAEC. IV-IX)

I

SONGS FROM NON-LITURGICAL SOURCES CANTI DI TRADIZIONE NON LITURGICA

I. Lyrics / Canzoni

Digital edition coordinated by FRANCESCO STELLA
Musical edition by SAM BARRETT
Digital reproductions of the manuscripts
Audio recordings by choirs Laus cordis and Palma choralis
directed by GIACOMO BAROFFIO and KIM EUN JU
Sofware by LUIGI TESSAROLO



© 2007 SISMEL - EDIZIONI DEL GALLUZZO





CRM-Web: <u>www.corimu.unisi.it</u>



CORPUS RHYTHMORUM MUSICUM (saec. IV-IX) - I



DIGIMED FILOLOGIA DIGITALE DEI TESTI MEDIOLATINI



Università degli studi di Siena ad Arezzo

Università degli studi di Bergamo



CORPUS RHYTHMORUM MUSICUM (saec. IV-IX) - I >

Introductions

Texts

Manuscripts

Search

Concordances

Statistics

Tables

Help



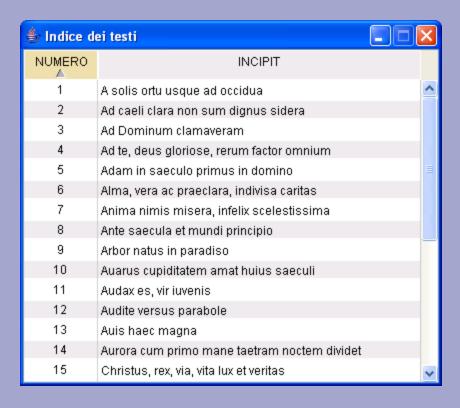
CORPUS RHYTHMORUM MUSICUM (SAEC, IV-IX)

SONGS FROM NON-LITURGICAL SOURCES CANTI DI TRADIZIONE NON LITURGICA

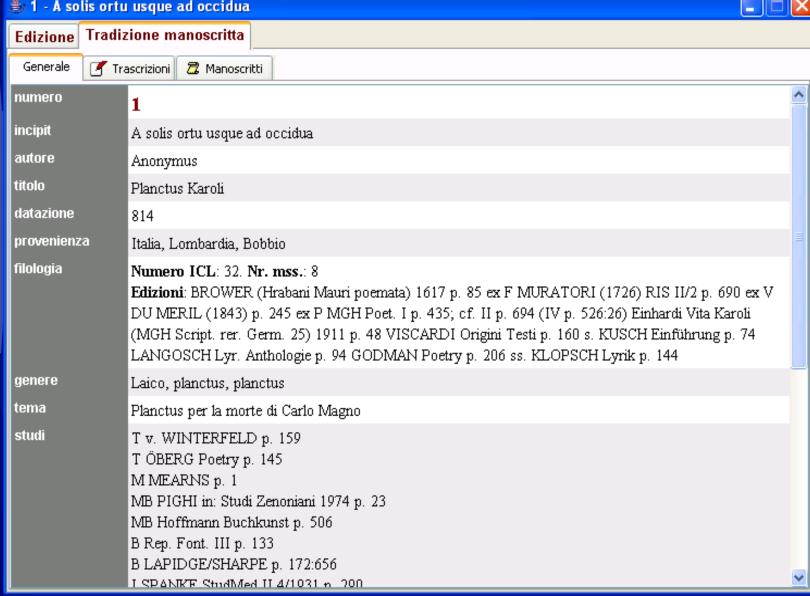


Public version (CRM): Index

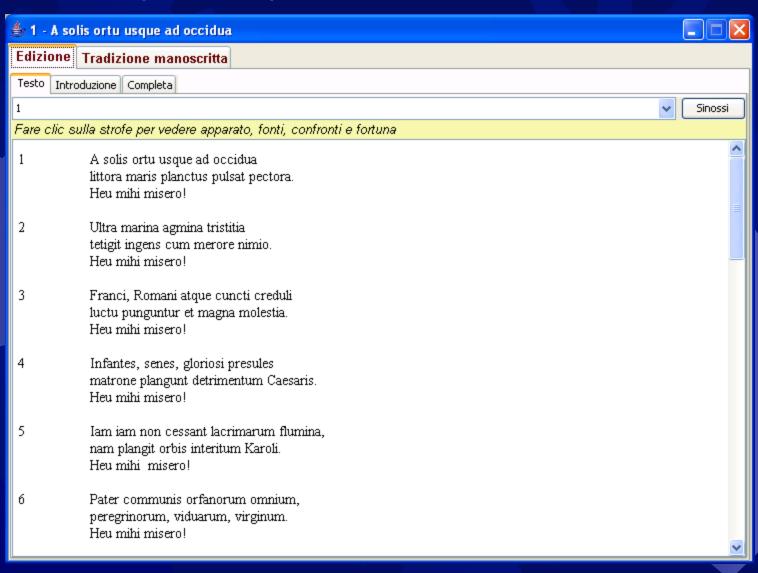




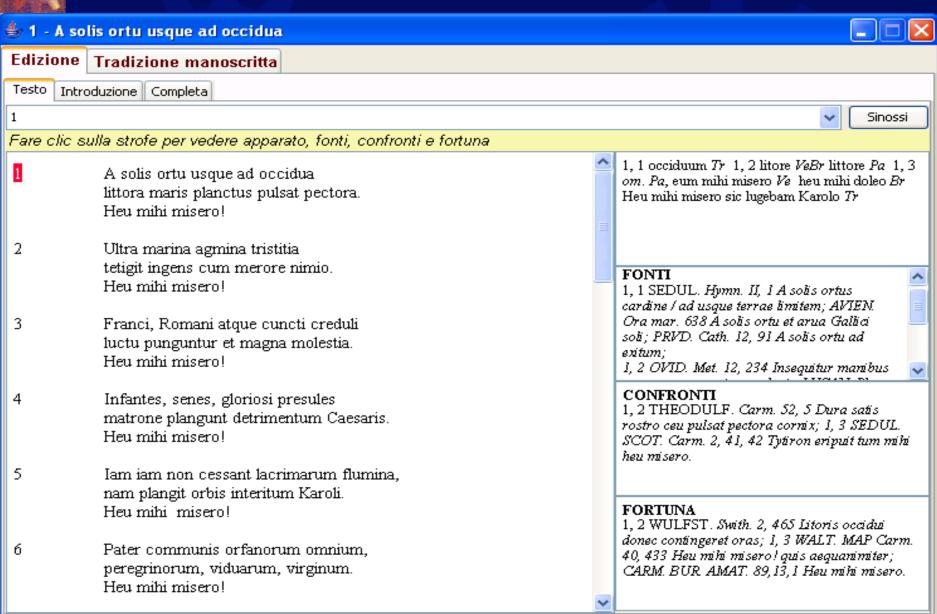
General infos on the text 🏂 1 - A solis ortu usque ad occidua Tradizione manoscritta Edizione Generale Trascrizioni 🛮 🎏 Manoscritti numero 1 incipit A solis ortu usque ad occidua autore Anonymus titolo Planctus Karoli datazione 814 provenienza Italia, Lombardia, Bobbio



Edition (critical)



Clickable apparatus



Synoptic edition

ut refriceretur lingua eius in flammis ardentihus.

🏂 19 - Homo quidam erat dives valde in pecuniis d) Bu, Fu, Fu1, Fu2, Kas, Lo, Wi (1-3, 5, 7-8) a) Br Pa (1-8, 10-11, 9, 12-13) b) Lei (1-7, 10-11, 13) c) Ve (1-5, 7-8, 10, 9, 11) Homo quidam erat dives valde in pecuniis, Homo quidam erat dives valde in pecuniis, Homo quidam erat dives valde in pecuniis, Homo quidam erat dives valde in deliciis. purpura et bysso induebatur, epulans splendide: purpura et bysso induebatur, epulabat splendide purpura e bysso induebatur, epulabat splendide: purpura et bysso induebatur, epulans splendide: caducam vitam diligendo amisit perpetuam. caducam vitam diligendo amisit perpetuam. caducam vitam diligendo amisit perpetuam. caducam vitam diligendo amisit perpetua. Lazarus quidam mendicus, circumdatus ulcera, Lazarus guidam mendicus, circumdatus vulnera, Lazarus quidam mendicus, circumdatus ulcera, Lazarus quidam mendicus, circumdatus ulcere, ad ianuam divitis iacebat cum dolore nimio; ad ianuam eius divitis iacebat cum dolore nimio; ad ianuam divitis iacebat cum dolore nimio; ad ianuam divitis <iacebat> cum dolore nimio; cupiebat saturari de micis mense divitis. Postulabat: nemo dabat, nemo miserebitur. Postulabat: nemo dabat, nemo miserebatur; Postulabat: nemo dabat, nemo miserebatur; Postulabat: nemo dabat, nemo miseretur; Tantum canes veniebant, consolabant pauperem, Tantum canes veniebant, consolabant pauperem. Tantum canes veniebant, consolabant pauperem. Tantum canes veniebant, consolantes pauperem, ulcera eius lingebant, curabantur vulnera. ulcera eius lingebant, curabantur vulnera. ulceram eius lingebant, curabantur vulnera. ulcera eius lingebant, curabantur vulnera. Tempus pauperis advenit, migravit a saeculo; Tempus pauperis advenit, migravit a saeculo; Tempus pauperi advenit, migravit a saeculo; caruit praesentem vitam, mutavit in melius: caruit praesentem vitam, mutavit in melius: caruit praesentem vitam, mutavit in melius: angeli eius portabant animam in requiem. angeli eius portabant animam in requiem. angeli dei portabant animam in requiem. Mortuus est autem dives, in infernum ducitur: Mortuus est autem dives, in infernum ducitur: Mortuus est autem dives, in infernum ducitur: Mortuus est autem dives, in infernum misericordiam non fecit, eam non merebitur. misericordiam non fecit, eo non merebitur. misericordiam non fecit, eam non consequitur. misericordiam non fecit, eam non consequitur. Pro epulas poenas recepit, cruciatur anxiae. pro epulas poenas recepit, cruciatur anxiae. pro micis panis, quas negavit, in infernum ducitur. pro micis panis, quas negavit, in infernum ducitur. In tormentis cum adesset, elevavit oculos, In tormentum cum hesit, elevavit oculos, Habraham a longe cernit, in sinu eius Lazarus; Abraham a longe cernit, in sinu eius Lazarum; ut recognovit, quem dispexit, in nullo ut recognovit, quem dispexit, nulla consolabitur. Patrem vocat Habraam, ut mittat Lazarum, Patrem vocat Abraham, ut admittat Lazarum, Patrem vocat Abraham, ut emittat Lazarum, Patrem vocat Habraham, ut emittat Lazarum, digiti extremi sui aquae guttam tribuat, digiti extrema sui aquae guttam tribuat, digiti extremi sui guttam aque tribuat, digiti extremi sui guttam aquae petiit,

ut refrigeretur linguam eius in flammis amentihus

ut refrigeretur lingua eius in flammis amentihus

ut refrigeret linguam eius in flammis ardentihus.

Clickable stemma codicum

🍰 1 - A solis ortu usque ad occidua

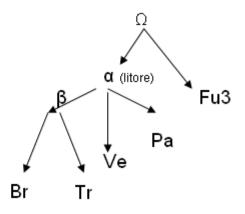
Edizione Tradizione manoscritta

Testo Introduzione Completa

altri casi di errore evidente).

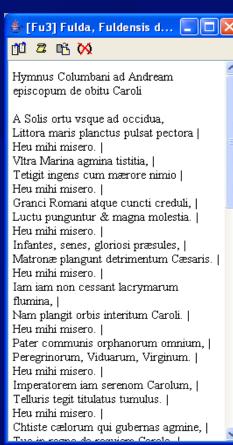
La forma del ritornello accomuna Ve e Tr., ma possiamo ritenere che sia in lezione giusta. Non ci sono errori che accomunano Fu3 Ve Pa, che hanno invece ognuno evidenti differenze sia n lacune (Pa e Fu3, non coincidenti) sia di vere e proprie redazioni (Pa), sia di errori singolari (ne ha 7, 1 acmina; 13, 1 perpensa; 14, 1 gleber terre;).

Una ricostruzione grafica dei rapporti fra i manoscritti come emergono dal testo che abbiamo p

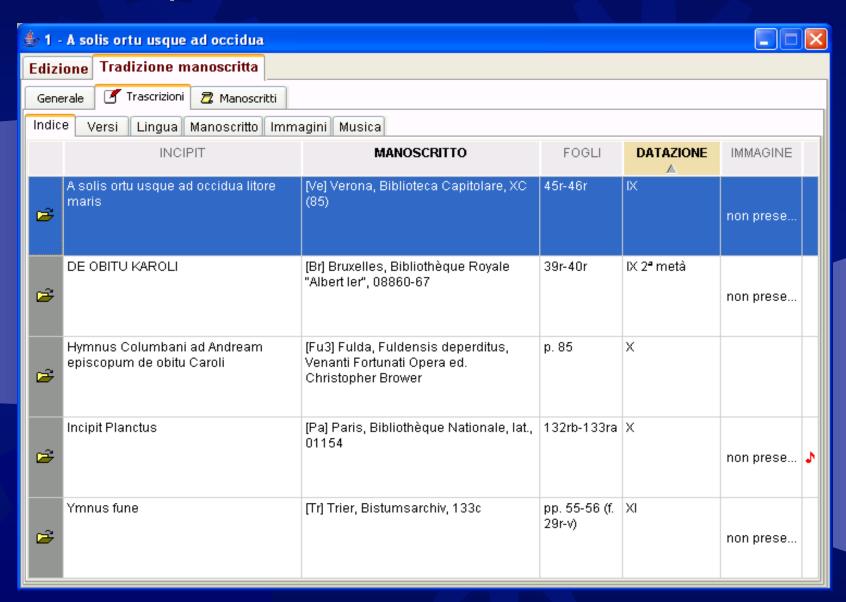


ELEMENTI METRICI E LINGUISTICI

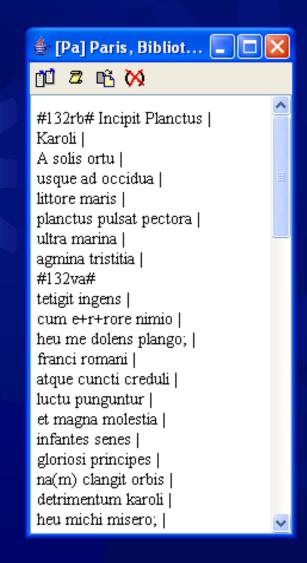
I versi sono dodicisillabi (5p+7pp, 12i.7) con cesura dopo la quinta sillaba ad andamento preval (A solis, infantes, Iam iam, Hoc poscunt, con assonanze irregolari, Ve tibi, Nox mihi, Quae cuncti gli altri incipit di strofa sono invece apparantemente trocaici) e assonanze irregolari (in a- la str. 1 e 8 e la 16; in &ndashu la 6, la 9 e la 18; in &ndasho la 11, in &ndashe la 19), Diastole a 5, 2 interitu

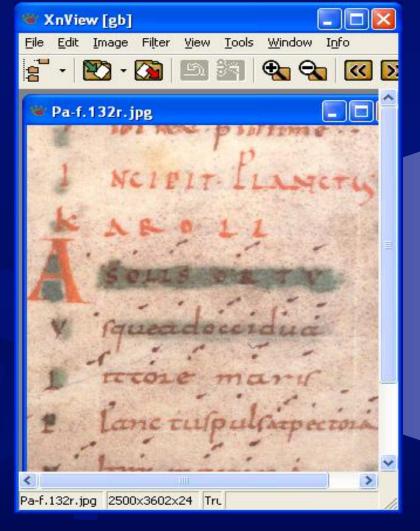


Manuscripts

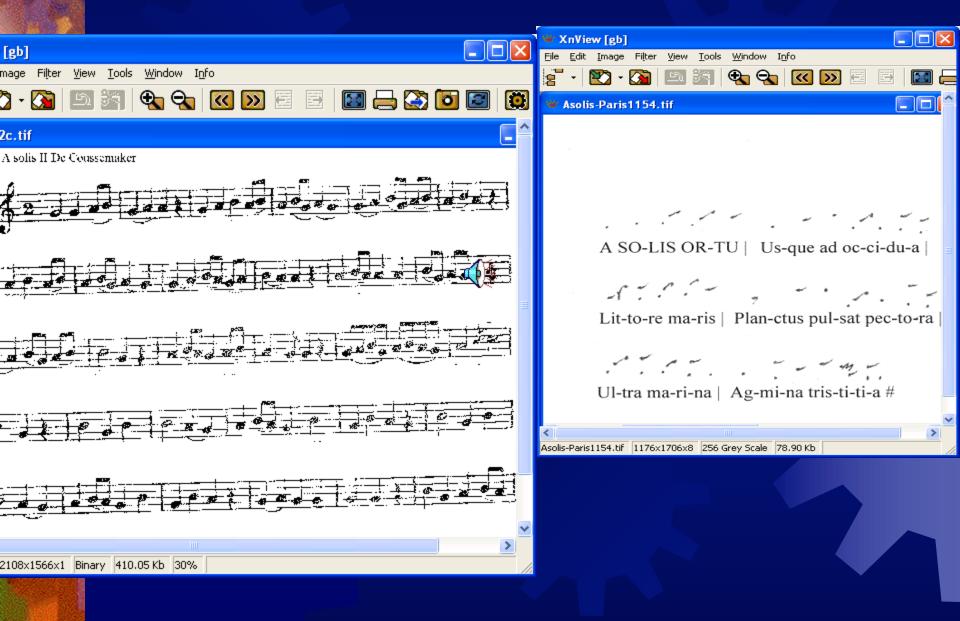


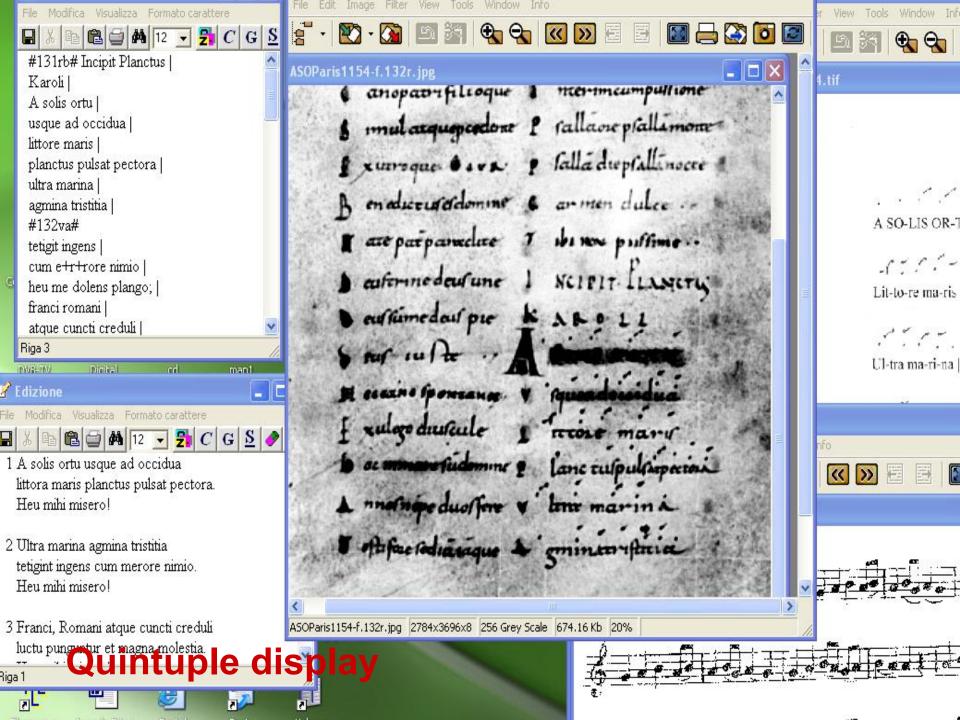
Mss images and transcriptions



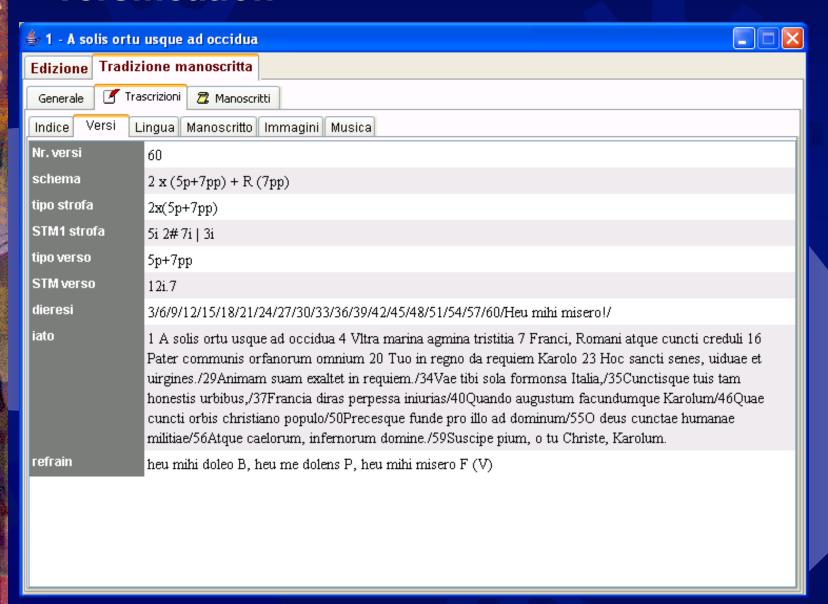


Music's transcripts and records

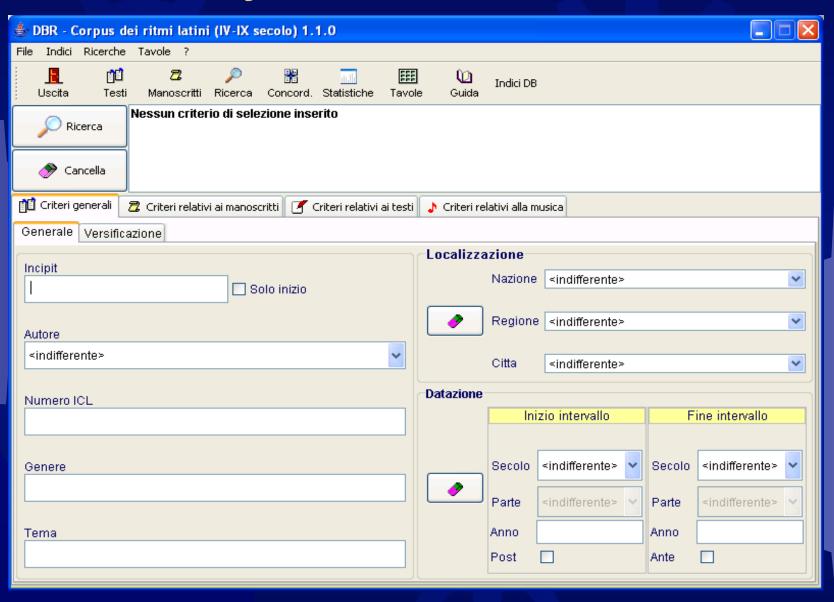




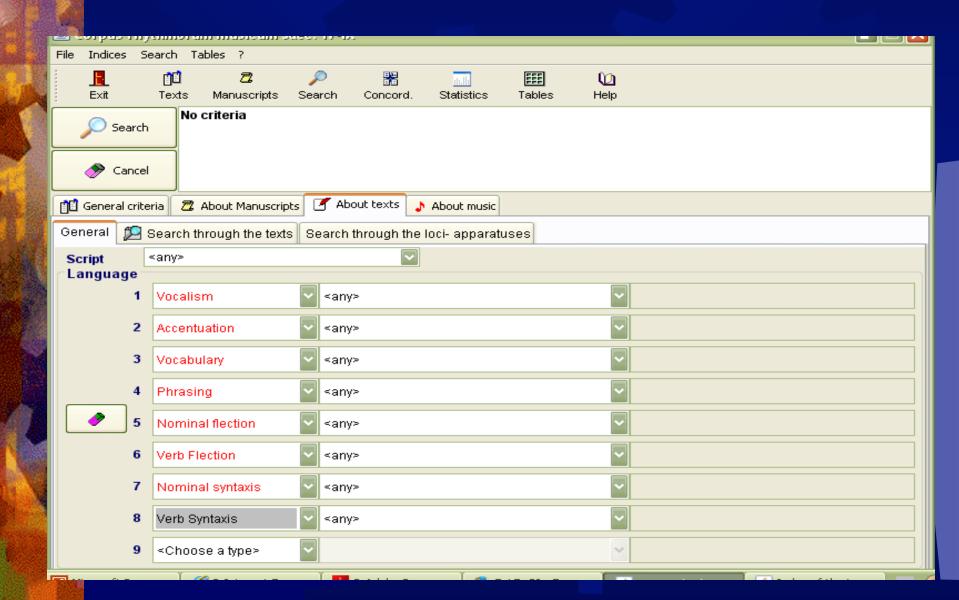
Versification



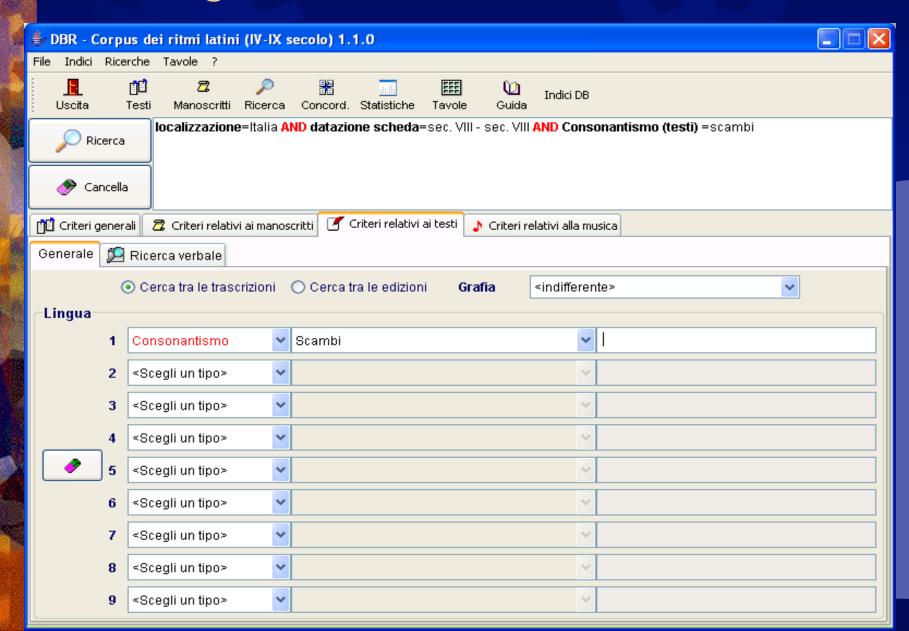
Every field cross-searchable



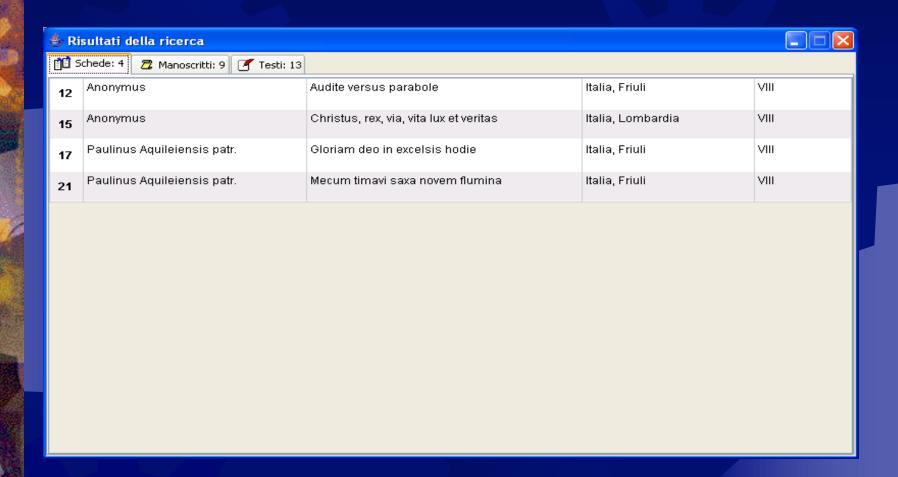
Search display



Searchsample: Italy+viii century+consonant exchanges



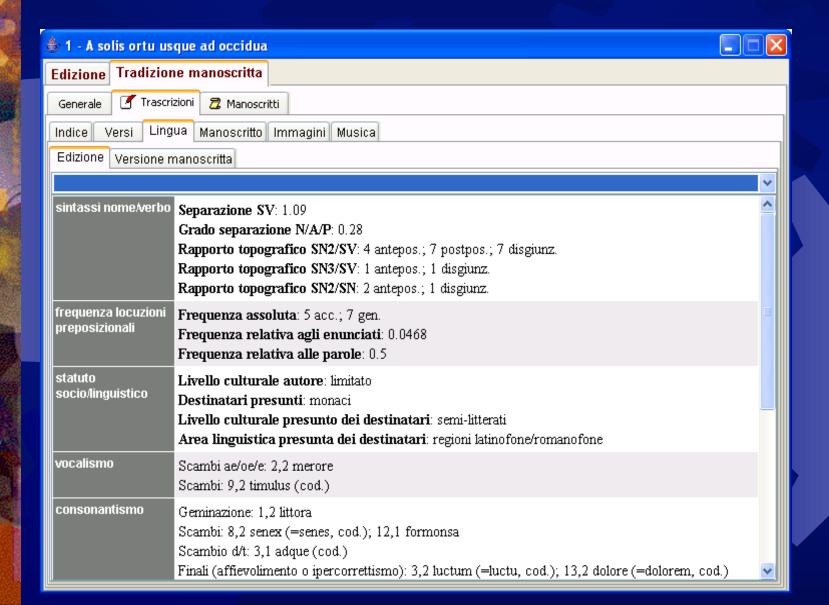
Search results



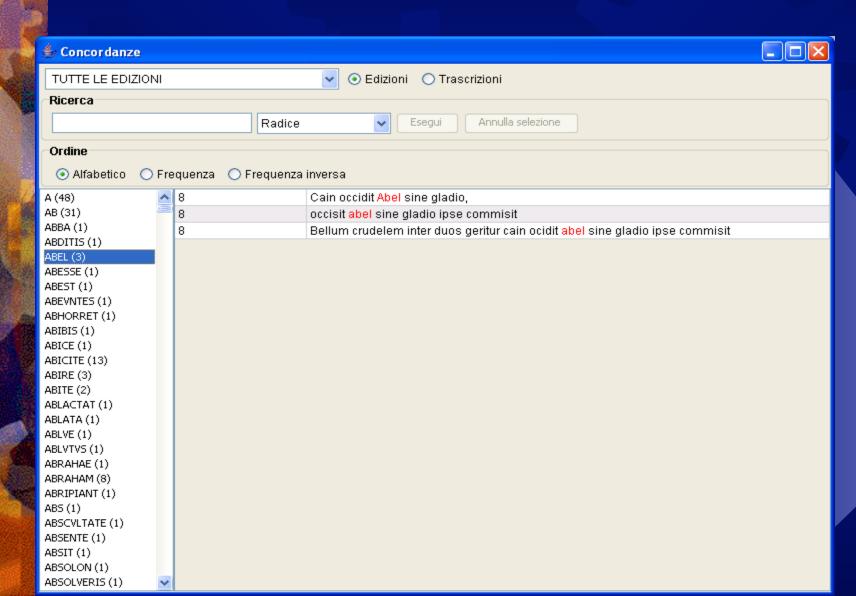
Clickable results

∌ Risul	tati della ricerca			X
Carlo Sche	de: 4 💆 Manoscritti: 9 📝 Testi: 13			
12	Audite versus parabule quod quon non puero con	[Va1] Città del Vaticano, Biblioteca Apostolica Vaticana, Pal. lat., 493 (sez. III)	106v	^
12	De puero interfecto	[CI] Apre il testo rand, Bibliothèque Municipale, 240 (189)	109rb-rc	
12	Audite uersus parabole de quon	[Ve] Verona, Biblioteca Capitolare, XC (85)	22r-23r	
15	DE FIDE ET CARITATE.	[Br] Bruxelles, Bibliothèque Royale "Albert ler", 08860-67	67v-69v	
15	VERSUS	[Pa] Paris, Bibliothèque Nationale, lat., 01154	113vb-116ra	
17	Gloriam deo in excelsis hodie caelestis primum ce	[Ro] Roma, Biblioteca Angelica, 123 (B. III. 18)	187v-190r	∄
17	Gloriam deo in excelsis hodie celestis primum cecinit	[Na] Napoli, Biblioteca Nazionale, IV. G. 68	92v	
17	DE NATIVITATE DOMINI	[Ve2] Verona, Biblioteca Capitolare, LXXVIII (83)	62v-63r	
17	GLORIA Deo	[Fu3] Fulda, Fuldensis deperditus, Venanti Fortunati Opera ed. Christopher Brower		
17	De natiuitate domini	[Br] Bruxelles, Bibliothèque Royale "Albert ler", 08860-67	48v-52r	
17	Gloriam deo in excelsis hodie celestis primum	[Ve] Verona, Biblioteca Capitolare, XC (85)	3v	
17	INCIPIT VERSVS	[Pa] Paris, Bibliothèque Nationale, lat., 01154	123ra-127va	~

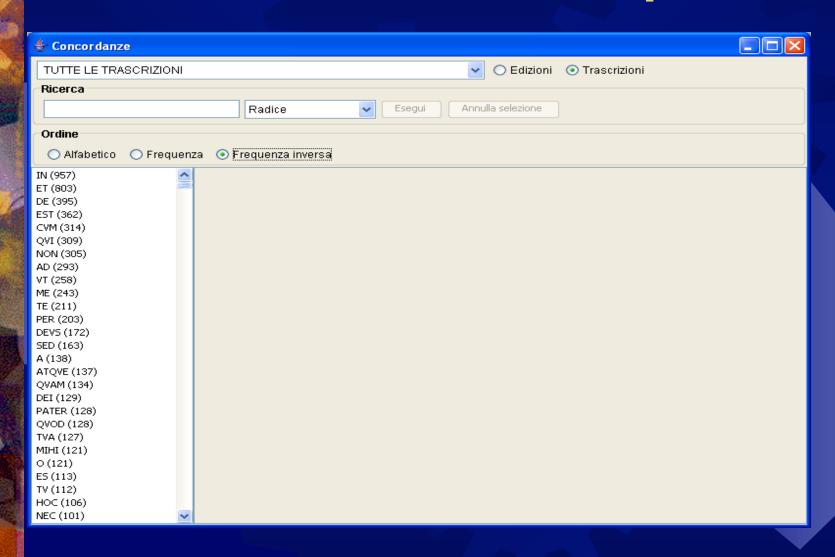
Synopsis of the linguistic data



Concordances of the forms (editions or transcriptions)



Direct and reversal frequencies



Interactive table of the linguistic data

\$	Statistiche																	X
N°	INCIPIT	Sep. SV	Sep. N/A/P	SN2/SV antep.	SN2/SV postp.	SN2/SV disg.	SN3/SV antep.	SN3/SV postp.	SN3/SV disg.	SN2/SN antep.	SN2/SN postp.	SN2/SN disg.	Loc. prep. acc.	Loc. prep. gen.	Loc. prep. abl.	Loc. prep. /enunc.	Loc. prep. /par.	
5	Adam in saeculo primus in domino	3,829	3,829	2	1	0	0	0	0	0	0	0	0	7	0	0,233	0,583	۸
21	Mecum timavi saxa novem flumina	3,319	0,519	48	28	36	4	3	7	•21	5	22	12	11	0	0,085	0,305	
20	Hug dulce nomen, Hug propago n	1,669	2,440	6	1	6	0	1	0	0	2	2	1	8	0	0,615	0,345	
16	Fuit domini dilectus languens a B	1,549	1,299	24	22	16	8	0	0	5	13	5	16	31	0	0,026	0,282	
26	Quique cupitis audire	1,500	1,269	4	7	5	3	5	0	1	2	0	7	21	0	0,075	0,635	ı
17	Gloriam deo in excelsis hodie	1,500	2,720	20	23	22	7	4	2	10	10	9	8	40	0	0,066	0,310	
6	Alma, vera ac praeclara, indivisa c	1,440	1,679	16	9	16	3	5	1	10	13	7	12	9	0	0,046	0,114	Ξ
13	Auis haec magna	1,360	1,049	6	2	2	0	0	0	4	6	6	6	5	0	0,015	0,151	
2	Ad caeli clara non sum dignus sid	1,246	0,936	17	12	15	1	5	1	7	10	14	9	15	0	0,059	0,291	
4	Ad te, deus gloriose, rerum factor	1,190	0,699	14	9	13	0	4	2	5	4	7	10	11	0	0,051	0,307	
22	O mi custos, o mi heros, mi pater	1,169	1,399	48	28	36	4	3	7	215	0	22	25	47	0	0,050	0,206	
25	Qui signati estis Christo	1,116	0,300	11	3	10	1	4	1	2	3	3	7	15	0	0,077	0,372	
1	A solis ortu usque ad occidua	1,090	0,280	4	7	7	2	0	1	1	0	1	5	7	0	0,046	0,500	
24	Qui de morte estis redempti	1,049	1,000	17	4	10	1	11	0	3	4	3	4	17	0	0,063	0,300	~

Criteria: based on specific scholarly requirements for medieval Latin

- 1) Calculation of the ratio '**Total number of prepositions in a single text / Governed cases**'. This indicates how many prepositions govern the accusative, how many the ablative and how many the genitive.
- 2) Calculation of the ratio '**Total number of prepositions** *I* sentences (verbs) and words'. When prepositions don't govern the case expected according to classical grammatical rules, the real case that occurs has been considered (e.g. *cum duas filias* is preposition with accusative).
- 3) Ratio **Prepositions/Verbs**: verbs are considered, besides the finite moods, also joint participles and ablative absolutes, gerunds and gerundives with final value; participles, infinite and substantive gerunds or gerunds with evident adjectival value are not considered. An auxiliary verb + infinite has been counted as independent from the governing verb only in the case of an objective clause or final value.

Syntax

- 1) **Subject/Verb Separation**: shows the ratio 'Total words between subject and its verb / Total number of the expressed verbs'. Participles and indefinite moods are generally considered only when they have verbal function. If there are more subjects, the calculation starts from the last one; if there are more verbs (e.g. modal verbs + infinite), the last one has been considered. The *-que* participle is not counted as an independent word.
- 2) **Noun/Adjective/Participle Separation**. The data have been obtained by dividing the number of words that separate nouns from their adjectives (or participles) by the tTotal number of nouns involved. Where there are more adjectives or participles, thea average number is considered. The *-que* participle is not counted as an independent word as well.
- 3) **SN2** (direct object) / Verb Ratio. The position of the direct objects (only nouns, not their adjectives or participles) in relation to the verb of the sentence: how many times the accusative is placed before, how many times it is postponed and how many times detached (that is separated by more than one word). The inner accusatives, the motion accusatives andor the accusatives that are subjects of an infinitive sentence are not counted.
- 4) **SN3** (dative/ablative cases) / **Verb** ratio: the position of the non-direct complements without preposition, with the exception of connected adjectives or participles, in relation to the governing verb: how many times it is situated before or after and how many times separated from the verb by one or more words. Ablative absolutes are not considered.
- 5) **SN4 (genitive)** / **Nouns** ratio: the position of the genitive (only the substantive) in relation to its noun: how many times before, how many times after, and how many times separated by one or more words.



Literary sources

Filtri



Ricerca

Ricerca avanzata	sul testo			6
Parola 1	1 T Paro	la 2	Cerca per prossimità 0	
Numero di testi:	360			
<i>Liber de temp</i> Albertus Milio		ibus. Cont	inuatio Regina et Cronica Imperatorum	4
Type: Fonti Letter Genre: Storiograf				
			sanctum Petrum, scilicet in Kalendas Iullii usque ad Kalendas Ianuari ții, et stetit per menses sex. Et domnus Gibertus de Arcellis de	ı
Annales Pisa Bernardus Ma				
Type: Fonti Letter Genre: Storiograf				
			generationes X. A Noe usque ad Abraam fuerunt anni MDCCCIII, nni MXXIIII, generationes XIIII. A David rex usque ad transmigratione	m Bab
Annales Sano Anonimo	ctae Trinitatis	Veronens	ies	
Type: Fonti Letter	rarie			

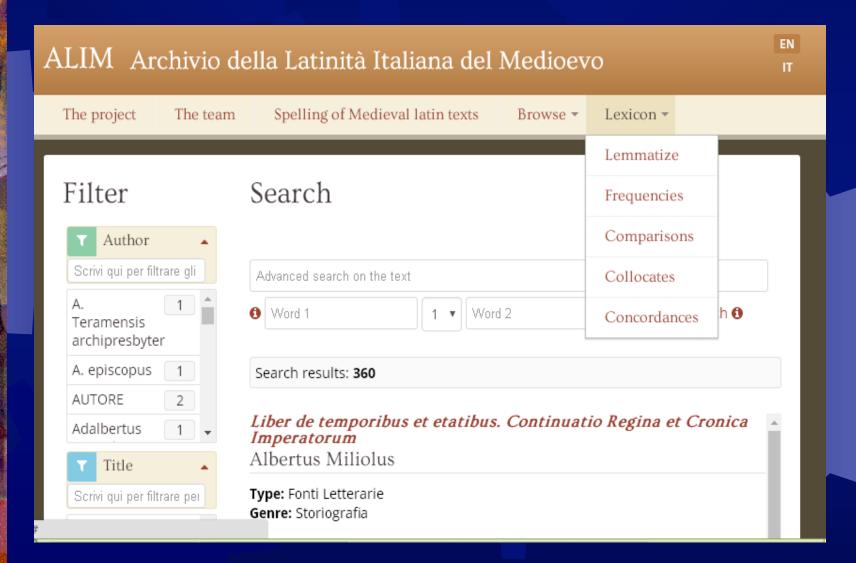
Documentary sources



Lexicon in ALIM2

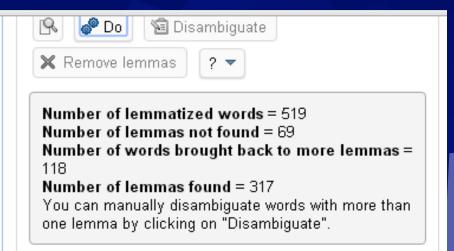


Embedded Lexicon



New: Prosodic Lemmatizer

Title	Lemmas	
Albertus Miliolus - Liber de temporibus et etatibus. Continuatio Regina et Cronica Imperatorum		
Bernardus Marango - <i>Annales Pisani</i>		
Anonimo - Annales Sanctae Trinitatis Veronenses	~	
Anonimo - <i>Annales Veronenses antiqui</i>	~	
Anonimo - <i>Annales breves Veronenses</i>	~	
Anonimo - <i>Annales veteres Veronenses</i>		
Raimundus Lullus - <i>Ars brevi</i> s		
Benzo Albensis - Ad Heinricum IV. imperatorem libri VII		
Anonimo - Carmen de Ambrosio grammatico	~	
Anonimo - Chronicon Novaliciense		
Laurentius Valla - De falso credita et ementita Constantini donatione		
Iohannes de Virgilio - De figuris secundum Iohannem de Virgilio		
Anonimo - <i>De more medicorum</i>		
Iohannes de Virgilio - <i>Diaffonus</i>		
O-vive 0time (4.070t- 4.000)		



Title	Lemma
pre1	Lemma non trovato
in	ĭn
nomine	nōmen
domini	dőmĭnűs
dicam	dīcă dīcŏ
uos	vōs
audite	audĭŏ
uestro	vester
bono	bönüm bönüs
proferam	prōfĕrŏ

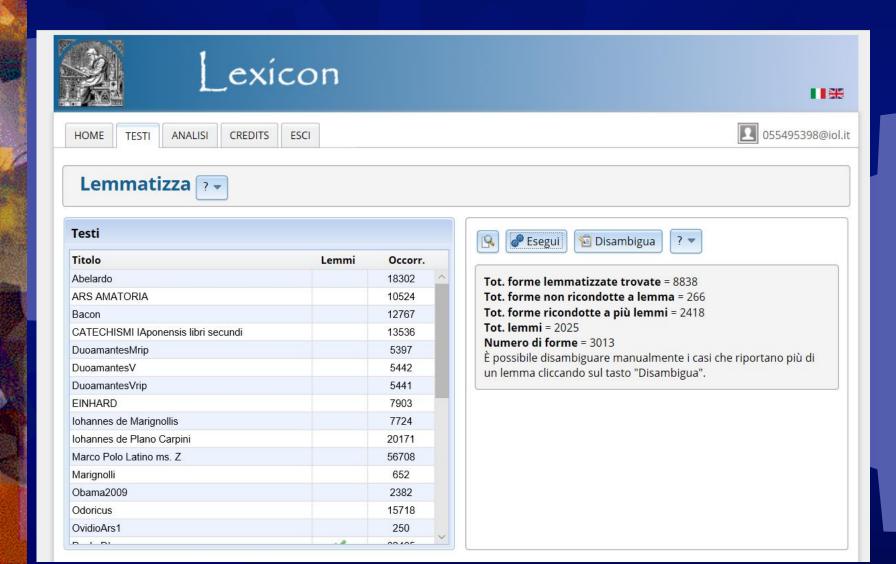
Disambiguation tool

rmen de Ambrosio grammatico (alim-188.txt)

>]1. In nomine Domini dicam(=dĭcă_00|dīcŏ_19), vos audite, Vestro bono(=bŏnŭm_04|bŏnŭs_11) proferam <vobi scite(=sciscŏ_19|scĭŏ_21|scītē_29); Primo detis veniam(=vĕnĭă_00|vĕnĭŏ_21) lingue(=) non ŏlītŭs_11|pŏlĭŏ_21|pŏlītē_29), Gerarde(=) cum sotiis, doctor redimite(=rĕdĭmŏ_19|rĕdĭmĭŏ_21). [</PRE>] >]2. Duo vobis referam pro(=prō_29|prōh_29) vestri(=vōs_-2|vester_13) honore(=hŏnor_05|hŏnōrŭs_11): Primo loquar nre dolore. De vestro dulcissimo defuncto nastore. Et secundo(=sĕcundŭs_11|sĕcundŏ_17) hreviter de



Paulus Diaconus, Carmina



Odoricus (15k words, 1min38 sec, 7% errors, mostly allographies)



exicon



HOME TESTI

ANALISI

CREDITS

ESCI

055495398@iol.

Lemmatizza ? -



Titolo	Lemmi	Occorr.	
Andrew Control	Lemmi		
Abelardo		18302	-
ARS AMATORIA		10524	_
Bacon		12767	
CATECHISMI IAponensis libri secundi		13536	
DuoamantesMrip		5397	
DuoamantesV		5442	
DuoamantesVrip		5441	1
EINHARD		7903	
Iohannes de Marignollis		7724	
Iohannes de Plano Carpini		20171	
Marco Polo Latino ms. Z		56708	
Marignolli		652	
Obama2009		2382	
Odoricus	4	15718	
OvidioArs1		250	
B . B.		00.405	



Tot. forme lemmatizzate trovate = 14805

Tot. forme non ricondotte a lemma = 915

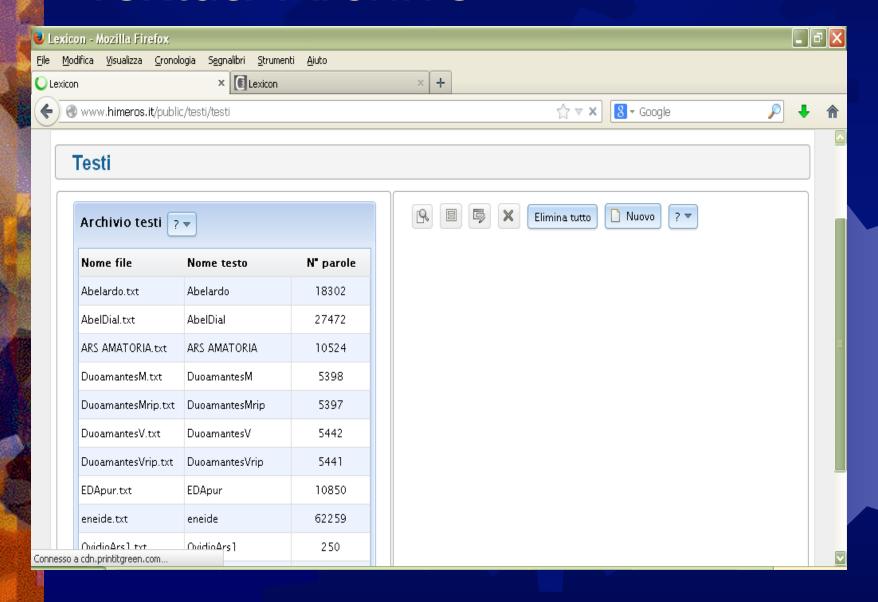
Tot. forme ricondotte a più lemmi = 3579

Tot. lemmi = 1739

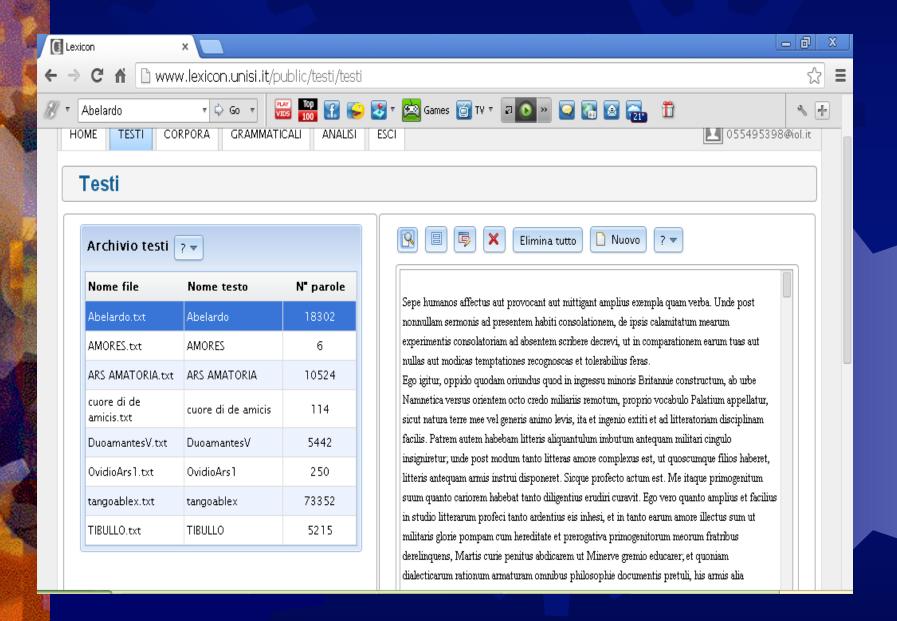
Numero di forme = 3692

È possibile disambiguare manualmente i casi che riportano più di un lemma cliccando sul tasto "Disambigua".

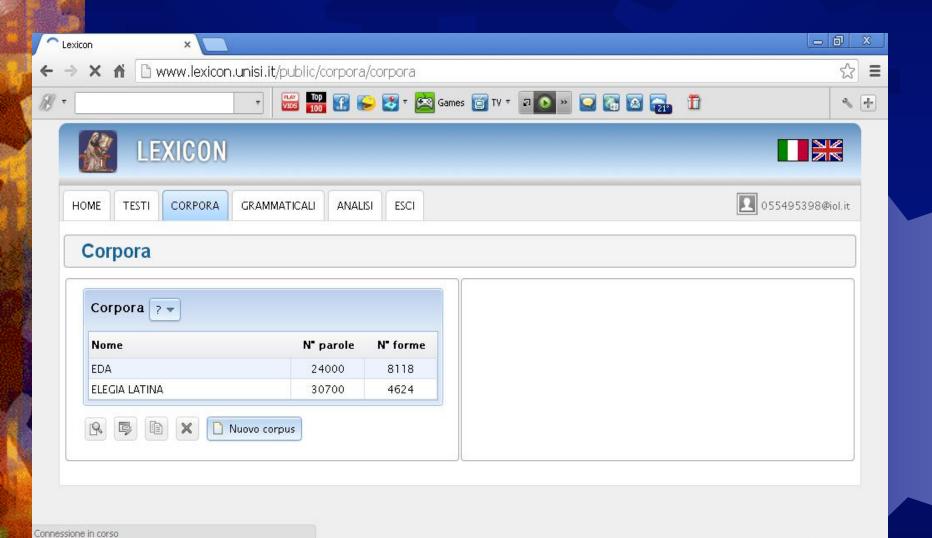
Textual Archive



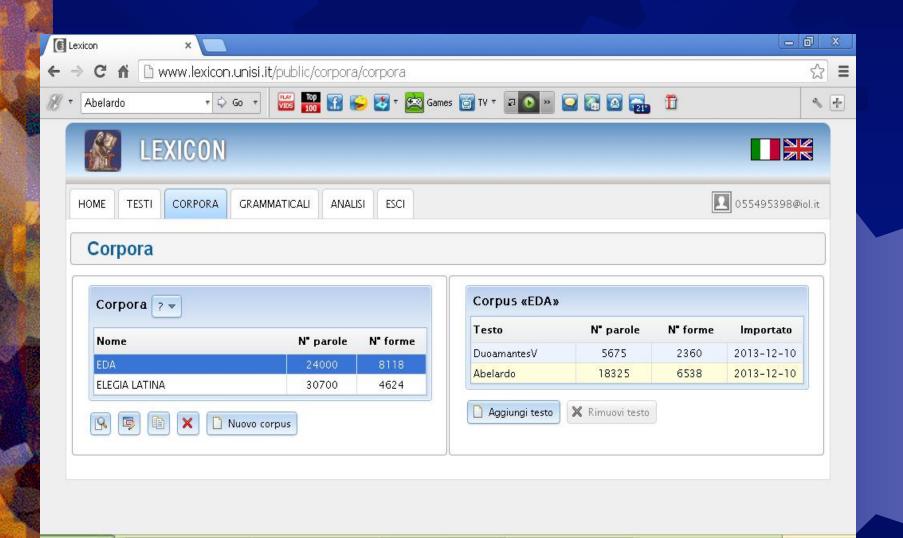
Showing the text



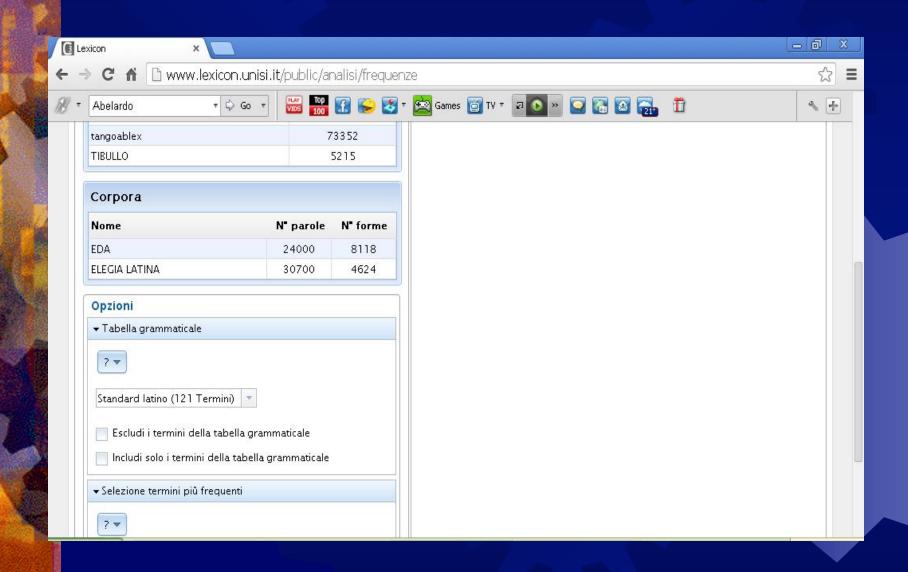
Corpora



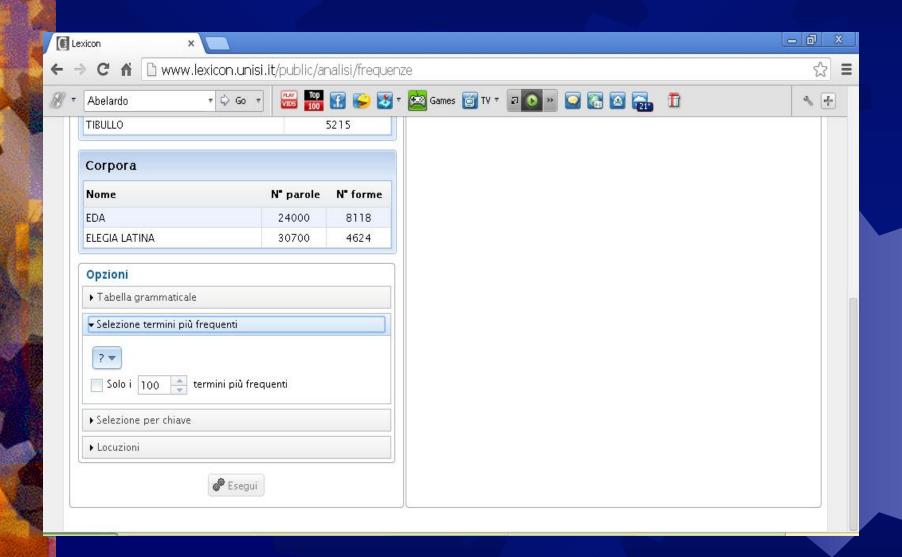
Opening the corpus



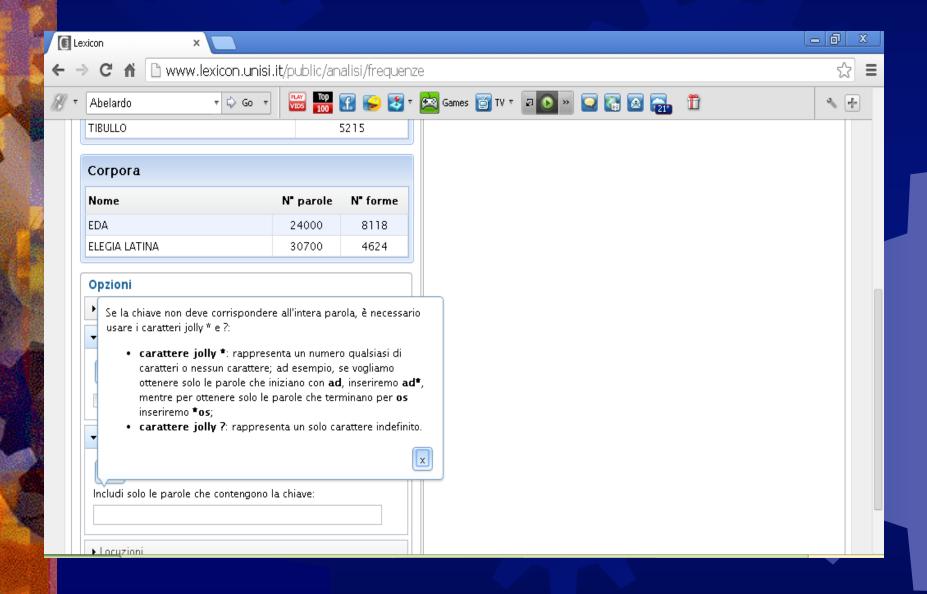
Grammatical Tables



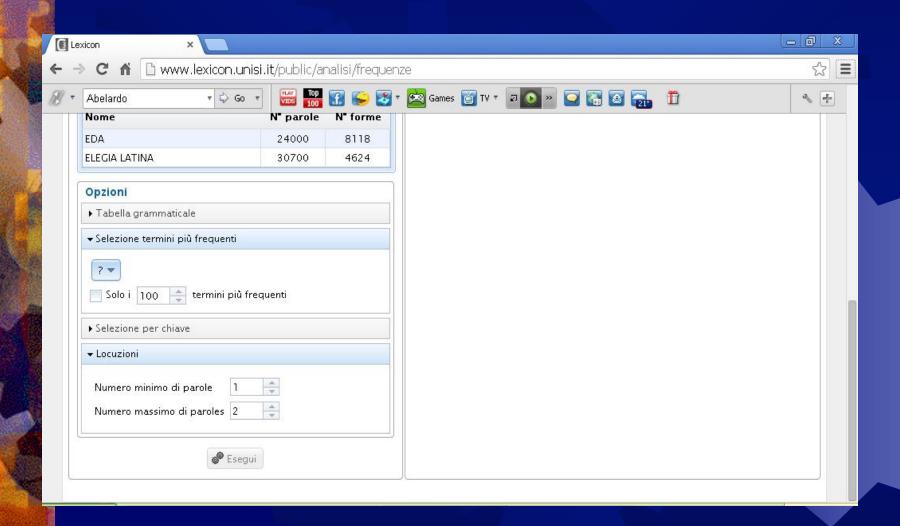
Selecting the number of words



Selecting the key



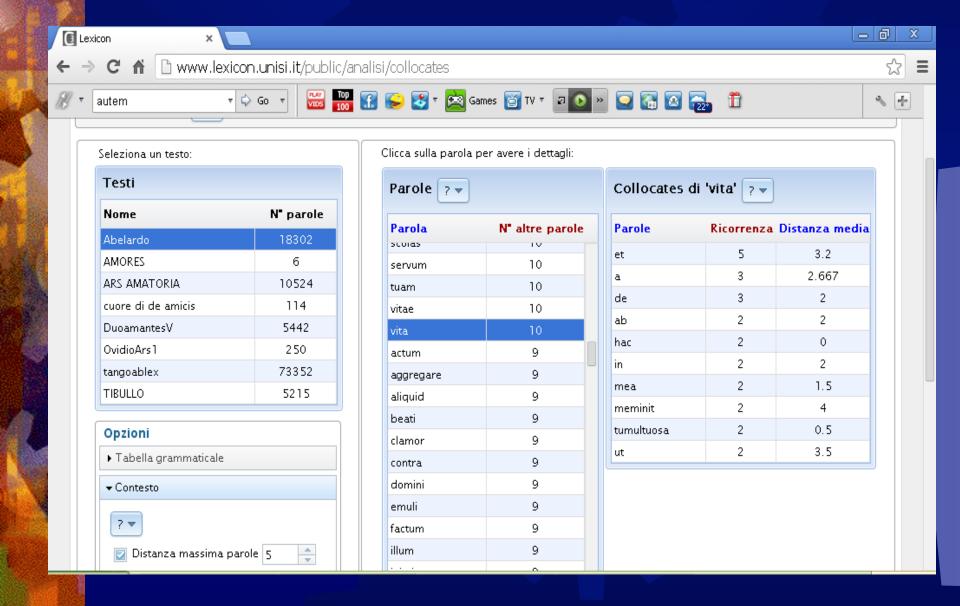
Clusters (= word-ngrams)



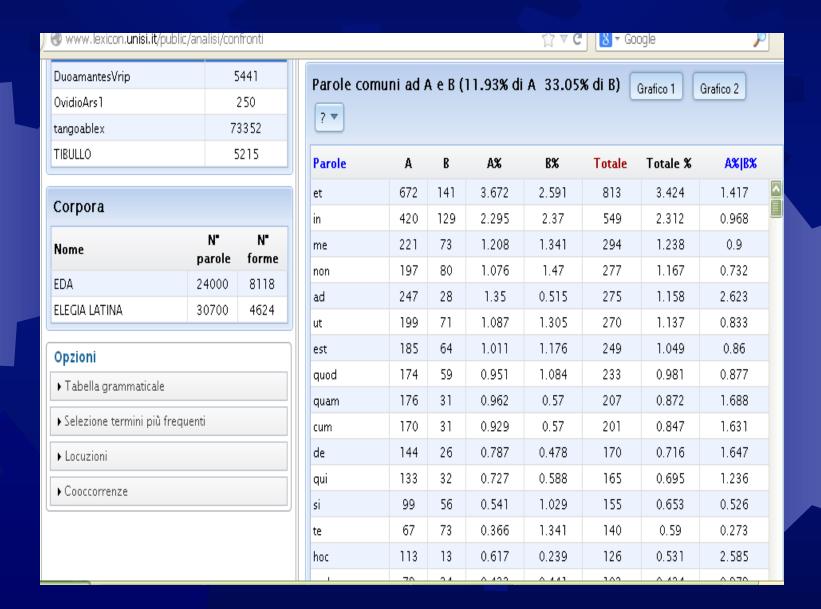
Reversal frequencies

Parola	Ricorrenza	Frequenza %
et	813	3.388
in	549	2.288
me	294	1.225
non	277	1.154
ad	275	1.146
ut	270	1.125
est	249	1.038
quod	233	0.971
quam	207	0.863
cum	201	0.838
de	170	0.708
qui	165	0.688
si	155	0.646
te	140	0.583
hoc	126	0.525
mihi	112	0.467
sed	103	0.429
autem	102	0.425
nec	99	0.413

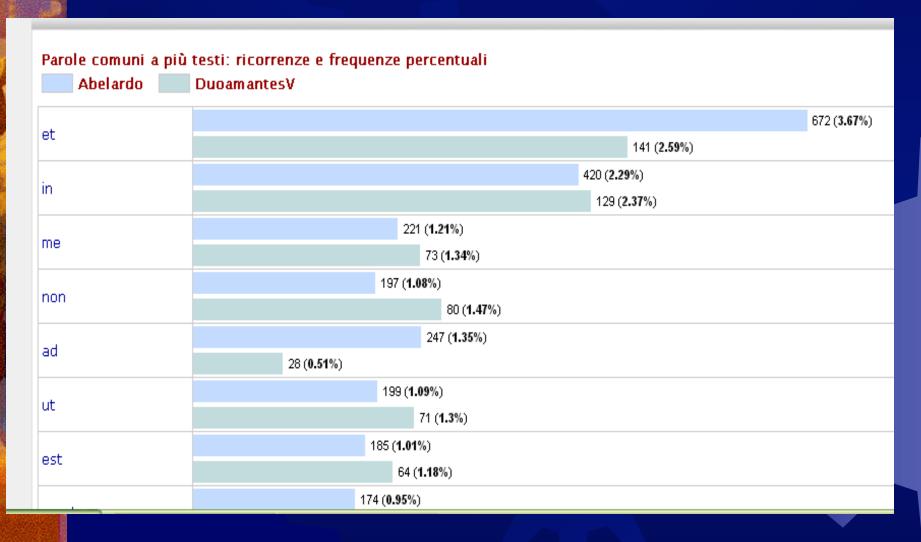
Collocates



Comparison typologies: AxB



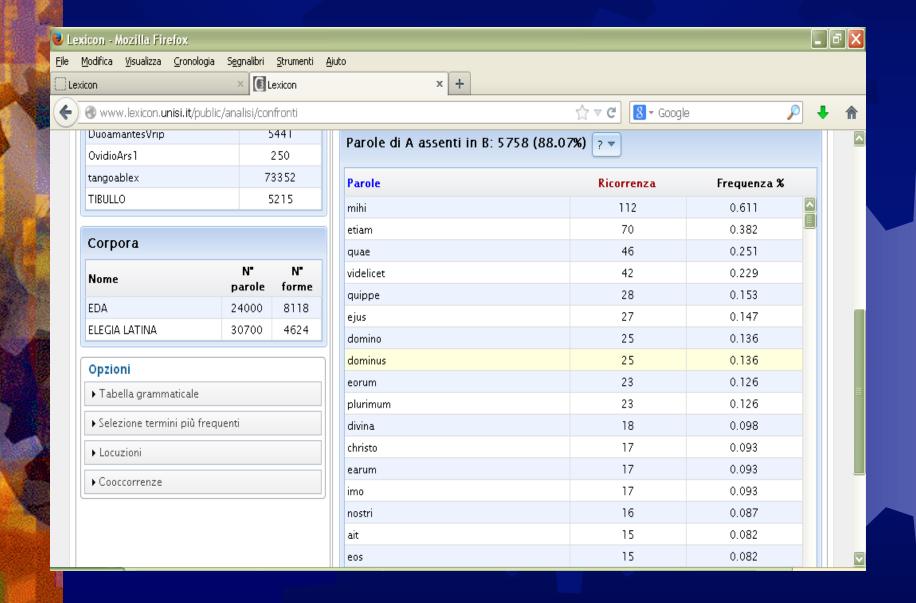
AxB Graph1



AxB Graph 2



A - **B**



B - A

раванансэтр	2111
OvidioArs1	250
tangoablex	73352
TIBULLO	5215

Corpora

Nome	N" parole	N" forme
EDA	5675	2360
ELEGIA LATINA	30700	4624

Opzioni

▶ Tabella grammaticale

▶ Selezione termini più frequenti

▶ Locuzioni

▶ Cooccorrenze

Parole di B assenti in A: 1514 (66.03%)

?	♥

Parole	Ricorrenza	Frequenza %
michi	39	0.704
nichil	26	0.469
dulcissima	20	0.361
lux	7	0.126
velim	7	0.126
eciam	6	0.108
gaudio	6	0.108
mi	6	0.108
nimis	6	0.108
alterum	5	0.09
amiciciam	5	0.09
dulce	5	0.09
dulcis	5	0.09
factis	5	0.09
gaudium	5	0.09
gracia	5	0.09
nil	5	0.09

caduto

Many projects and tools, poor literature

We don't need just archives and tools, we need scholarly purposes and hypothesis to check: knowledge-led developments

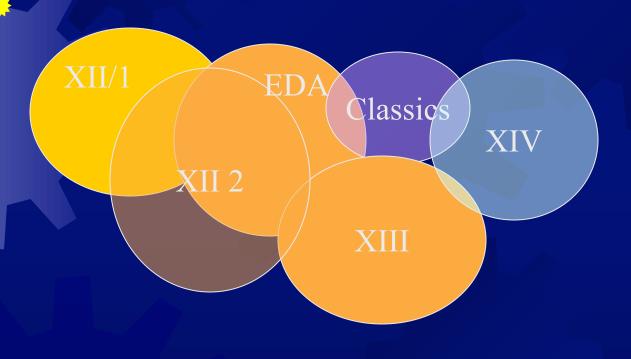
Non-lemmatized corpora: «Blind reading»

Sinclair 2004, p. 192: 'In corpus-driven linguistics you do not use pre-tagged text, but you process the raw text directly and then the patterns of this uncontaminated text are able to be observed.'

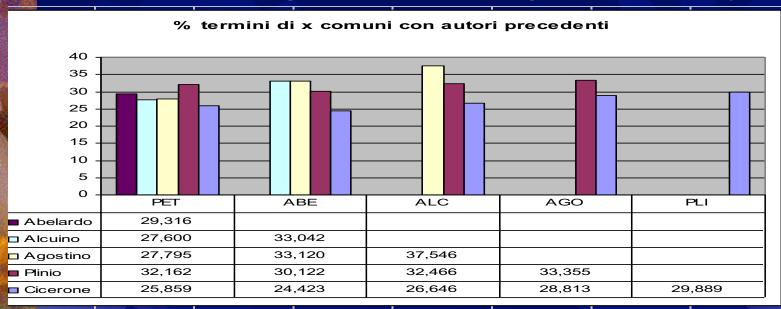
E.g.: Literary applications

- To get statistical data on authors, genres, epoques, linguistic trends and standard
- To search for the author of anoymous or disputed texts.
 - LATIN:
- The Corpus Caesarianum (Luon-Mellet 2003)
- The case of Karolus Magnus et Leo Papa
- Thomas Aquinas
- Epistolae duorum amantium
- Hildegard of Bingen's secretary: Guibert of Gembloux
- Gregory the Great Commentaries

Intertextual overlap (for dating the texts)



Extension: the genre vocabulary. The letters (2012)





Function words

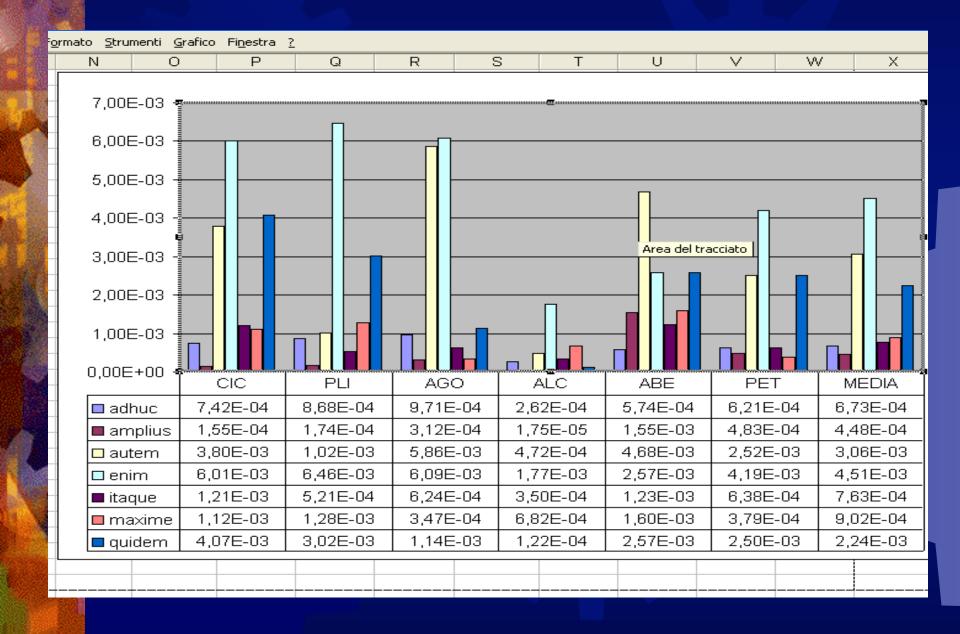
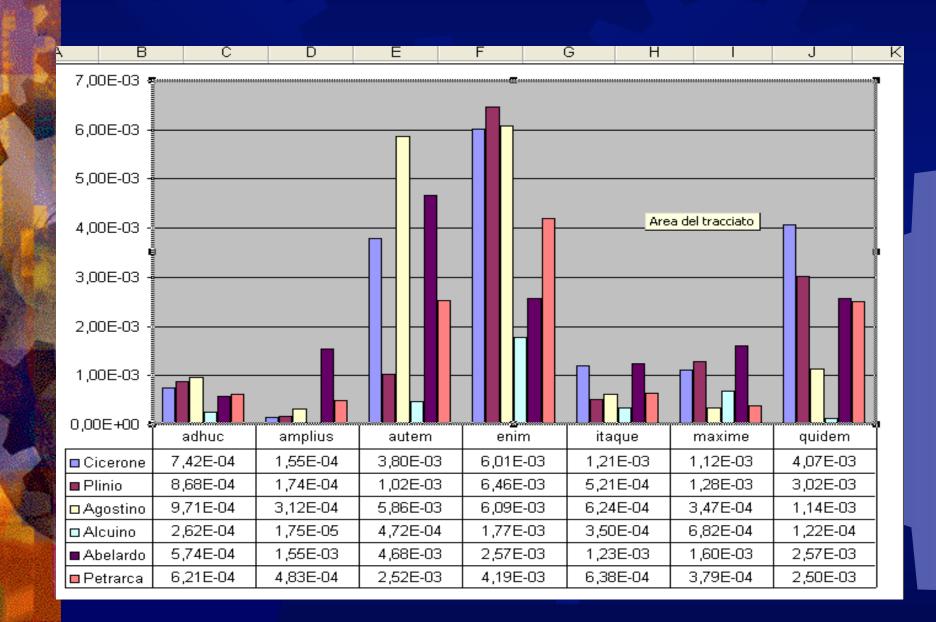
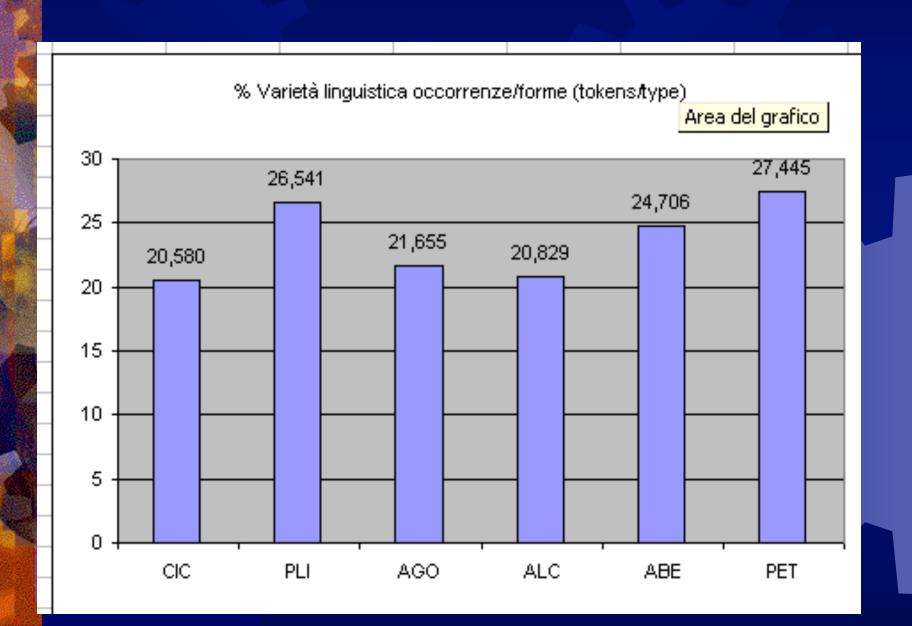


Table by word



Linguistic variation by author



Questions

- Does it exist a linguistic characterization of a literary genre?
- It is the genre's or the author's language that dominates?
- Which is the diachronical variance rate of the Latin vocabulary?
- Which is the variation rate of an author's vocabulary across different literary genres?
- Which is the threshold value of a statistical deviation?
- Some checks to find an «authorship threshold» through clusters statistics

Attest. Non shared 3-grams cluster = different authors?

Sigla	Tipo	Nome		
А	Testo	Abelardo		
В	Testo	DuoamantesV		
Mostra A - B	Mostra B - A	Mostra A x B	Cooccorrenze ?	
Locuzioni d	i A assenti in	B: 173 (100	0%)	
Locuzioni			Ricorrenza	Frequenza %
ut supra memin	i		5	0.027
et ob hoc			4	0.022
ad quam quide	m		3	0.016
ad quas quiden	n		3	0.016
illud est beati			3	0.016
lamentabantur 1	flentes dominum		3	0.016
mihi super hoc			3	0.016
mulieres sedent	tes ad		3	0.016
non est facile			3	0.016
non multo auter	m		3	0.016
post paucos die	es		3	0.016
sicut scriptum e	st		3	0.016

Non shared 3 grams-clusters: same author!

6: 1	т.			
Sigla	Tipo	Nome		
А	Testo	Abelardo		
В	Testo	AbelDial		
Mostra A - B	Mostra B - A	Mostra A x B		
Locuzioni	7 usseller ii		Ricorrenza	Frequenza %
ut supra memin	i		5	0.027
et ob hoc			4	0.022
ad quam quide	m		3	0.016
ad quas quiden	ı		3	0.016
illud est beati			3	0.016
lamentabantur f	lentes dominum		3	0.016
mihi super hoc			3	0.016
mulieres sedent	es ad		3	0.016
non est facile			3	0.016
non multo auten	n		3	0.016

shared 2-3 clusters: same author (4-12%)?

Sigla	Tip	00	Nome	7			
А	Tes	to	DuoamantesM				
В	Tes	to	Eloisa				
Mostra A - B	Mostra omuni a	a B - A Mostra A x B Cooccorrenze ? ▼ ad A e B (12.5% di A 4.62% di B) Grafico 1 Grafico 2					
Locuzioni	A	В	A%	В%	Totale	Totale %	A% B%
et in	6	18	0.111	0.185	24	0.159	0.601
in te	8	15	0.148	0.154	23	0.152	0.961
non est	6	11	0.111	0.113	17	0.112	0.983
in me	3	10	0.056	0.103	13	0.086	0.541
est et	3	9	0.056	0.093	12	0.079	0.601
a te	6	5	0.111	0.051	11	0.073	2.162
non enim	2	7	0.037	0.072	9	0.06	0.515
qui non	2	6	0.037	0.062	8	0.053	0.601
et non	3	4	0.056	0.041	7	0.046	1.351
et si	2	5	0.037	0.051	7	0.046	0.721
in omni	4	3	0.074	0.031	7	0.046	2.402
quod in	2	5	0.037	0.051	7	0.046	0.721

shared 2-3 clusters: same author (9-22%)!

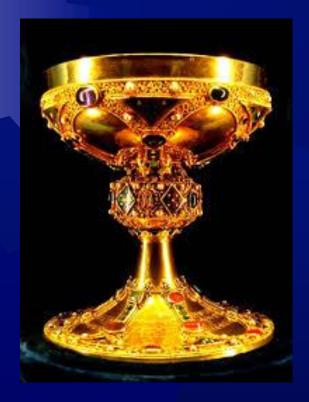
А	Tes	sto	Abelardo					
В	Tes	sto	AbelDial					
Mostra A - B	Mostra	Mostra B - A Mostra A x B Cooccorrenze ? ▼						
Locuzioni co	muni a	ad A e	B (22.879	6 di A 9.6	4% di B)	Grafico 1	Grafico 2	
Locuzioni	A	В	A%	В%	Totale	Totale %	A% B%	
id est	13	30	0.071	0.109	43	0.094	0.65	_
unde et	7	30	0.038	0.109	37	0.081	0.35	
sicut et	5	31	0.027	0.113	36	0.079	0.242	
et cum	11	21	0.06	0.076	32	0.07	0.786	
et in	14	18	0.076	0.066	32	0.07	1.167	
et de	11	16	0.06	0.058	27	0.059	1.032	
hoc est	3	24	0.016	0.087	27	0.059	0.188	
est et	9	17	0.049	0.062	26	0.057	0.795	
in hoc	14	12	0.076	0.044	26	0.057	1.751	
potius quam	10	15	0.055	0.055	25	0.055	1.001	
quod si	11	14	0.06	0.051	25	0.055	1.179	
magis quam	11	13	0.06	0.047	24	0.052	1.27	
scriptum est	12	12	0.066	0.044	24	0.052	1.501	

Shared clusters in different authors (Peter of Blois-Richard of S. Victor)

Sigla	Ti	90	Nome	Nome				
А	Tes	sto	PBloisChar					
В	Tes	sto	RiccDe4grad	dibus				
Mostra A - B	Mostra			Mostra A x B Cooccorrenze ? ▼				
Locuzioni co	omuni a	ad A e	B (5.81%	di A 13.3	5% di B)	Grafico 1	Grafico 2	
Locuzioni	Α	В	А%	B%	Totale	Totale %	A% B%	
et in	38	16	0.155	0.222	54	0.17	0.699	
in hoc	12	38	0.049	0.526	50	0.158	0.093	
est in	30	6	0.122	0.083	36	0.113	1.473	
non est	22	5	0.09	0.069	27	0.085	1.296	
non potest	3	24	0.012	0.332	27	0.085	0.037	
est et	19	7	0.078	0.097	26	0.082	0.799	
anima mea	12	10	0.049	0.139	22	0.069	0.353	
i cor	15	4	0.061	0.055	19	0.06	1.104	
est quod	13	5	0.053	0.069	18	0.057	0.766	
in se	16	2	0.065	0.028	18	0.057	2.356	
non solum	14	4	0.057	0.055	18	0.057	1.031	
toto corde	11	7	0.045	0.097	18	0.057	0.463	

and the Holy Graal of the Latin «stylometry»?

 The p-value (significance value) of the quantity of function words and percentage of 3-clusters, which identifies an author, a genre, an epoque in the Latin language.



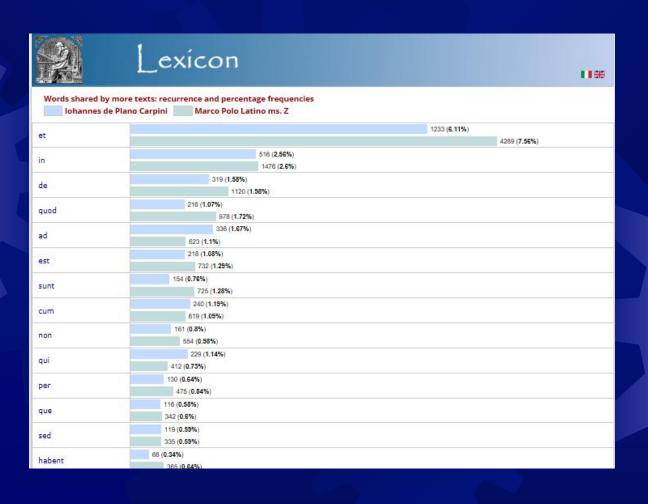
p=60FW p=9-12%3CI

Some trials of comparations by Lexicon: Historia Mongalorum and Marco Polo



nclude in th				om the com			
Sigla	Tip	00	Nome				
A	Te		lohannes de	Plano Carpii	ni		
В	Te	esto Marco Polo Latino ms. Z					
Display A -		lay B - /			of B) Cha	? ▼ ar 1 Char	2
Words	A	В	Α%	В%	Total	Total %	A% B%
et	1233	4289	6.113	7.563	5522	7.183	0.808
in	516	1476	2.558	2.603	1992	2.591	0.983
de	319	1120	1.581	1.975	1439	1.872	0.801
quod	216	978	1.071	1.725	1194	1.553	0.621
ad	336	623	1.666	1.099	959	1.247	1.516
est	218	732	1.081	1.291	950	1.236	0.837
sunt	154	725	0.763	1.278	879	1.143	0.597
cum	240	619	1.19	1.092	859	1.117	1.09
non	161	554	0.798	0.977	715	0.93	0.817
qui	229	412	1.135	0.727	641	0.834	1.563
per	130	475	0.644	0.838	605	0.787	0.769
que	116	342	0.575	0.603	458	0.596	0.954
sed	119	335	0.59	0.591	454	0.591	0.999
habent	68	365	0.337	0.644	433	0.563	0.524
etiam	180	175	0.892	0.309	355	0.462	2.892
quia	79	255	0.392	0.45	334	0.434	0.871

Graphics



Words of the *HM* missing in Marco Polo

Words of A absent from B: 2700 (59.2%)							
Words	Occurrence	Frequency %					
bati	34	0.169					
imperator	34	0.169					
immo	22	0.109					
vidimus	20	0.099					
chingiscan	19	0.094					
imperatorem	17	0.084					
ruscie	16	0.079					
tentorium	15	0.074					
ducem	14	0.069					
eidem	14	0.069					
habebamus	13	0.064					
iis	13	0.064					
litteras	13	0.064					
kytaorum	12	0.059					
manducant	11	0.055					
sagittis	11	0.055					
tertio	11	0.055					
diligenter	10	0.05					
iccirco	10	0.05					
invenimus	10	0.05					
mandatum	10	0.05					
modicum	10	0.05					
nichilominus	10	0.05					

in HM

Words of Marco Polo missing

Display A - B Display B - A Display A x B Collocates ? ▼

Words of B absent from A: 7489 (79.68%)

|--|

Words	Occurrence	Frequency %
ydola	113	0.199
insula	100	0.176
versus	100	0.176
regno	83	0.146
mercimoniis	74	0.13
qualiter	67	0.118
naves	66	0.116
mercimonia	55	0.097
postmodum	53	0.093
	47	0.083
itaque	47	0.083
argon	46	0.081
dietis	45	0.079
noveritis	44	0.078
caydu	43	0.076
artibus	42	0.074
disceditur	40	0.071
itur	40	0.071
quemadmodum	40	0.071
acmat	39	0.069
miliaria	39	0.069
nogay	39	0.069
nascuntur	38	0.067

Odoricus against Historia Mongalorum

Words	Occurrence	Frequency %
bati	34	0.169
duces	33	0.164
ibidem	31	0.154
tartaros	30	0.149
immo	22	0.109
chingiscan	19	0.094
munera	18	0.089
dux	17	0.084
occidunt	17	0.084
principes	16	0.079
ruscie	16	0.079
tartaris	16	0.079
credimus	15	0.074
pape	15	0.074
tentorium	15	0.074
ducem	14	0.069
eidem	14	0.069
ignoramus	14	0.069
nomina	14	0.069
secundo	14	0.069
diu	13	0.064
habebamus	13	0.064
iis	13	0.064
item	13	0.064
litteras	13	0.064

HM against Odoricus

Words of B absent from A: 2363 (65.75%)

|--|

Words	Occurrence	Frequency %
	72	0.458
contrata	43	0.274
fr	29	0.185
huius	28	0.178
hii	23	0.146
recedens	21	0.134
cadi	20	0.127
insula	20	0.127
millia	20	0.127
veni	20	0.127
habetur	19	0.121
ydolum	19	0.121
copia	17	0.108
vidi	17	0.108
iiiior	16	0.102
versus	16	0.102
saraceni	15	0.095
lo	14	0.089
milliaria	14	0.089
capite	13	0.083
istum	13	0.083
regno	13	0.083
contrate	12	0.076
ivi	12	0.076
penitus	12	0.076

Next comparisons

- Medieval missionaries vs Jesuits
- Medieval descriptions of China vs classical ones (Plinius) and modern ones (from Piccolomini to Matteo Ricci and Martino Martini)
- Scientific dictionaries, religious dictionaries, etc.

Some expected results

- Calculation of the overlap rate and exclusion rate of the lexicon concerning East regions between medieval and modern documents
- Extraction of the Latin vocabulary about Chinese,
 Japanese and Korean realia (real entities: objects,
 customs, names, places)
- Comparison about descritpion of material and intellectual or religious phenomena in original language and in Latin/other European languages
- Adopting the SLA research methods of analysis.



Asian Collection in ALIM

NOTI SICULO | en.amm.ums...t/ Conection/asiatic conection/

Asiatic Collection

This is a Collection of Latin texts from XIII to XVII century concerning East Asia, with main reference to China, Japan an Korea, which is aimed to a comparative linguistic analysis, to be carried both among the included texts and towards other Latin texts of different epochs and areas or non-Latin texts on the same ... Continue reading Asiatic Collection →



This is a Collection of Latin texts from XIII to XVII century concerning East Asia, with main reference to China, Japan an Korea, which is aimed to a comparative linguistic analysis, to be carried both among the included texts and towards other Latin texts of different epochs and areas or non-Latin texts on the same topics, inside the project *Eurasian Latin Archive:* data-mining and linguistic (genre, author, topic) thresholds of Latin Language, in cooperation between University of Siena/Regione Toscana and the University of Hanyang (Seoul), the universities of Otsuma and of Chukyo at Nagoya, Japan, waiting for an agreement with the Foreign Studies University of Beijing (China). After being implemented, the archive will be consultable also in the indipendent platform of the ELA project.

The first authors included will be:

- Iohannes de Plano Carpini (XIII sec.), already encoded
- · Odoricus de Portu Naonis (XIII sec.), encoded
- Iohannes de Marignollis (1290-1360), encoded
- Peregrinus de Castello
- · Andreas de Perugia
- Paolinus OFM (XIV sec.)
- Marco Polo Latin version (XIV sec.)
- Matteo Ricci (1552-1610) De Christiana expeditione apud Sinas (version byTrigault) Tetrabiblion Sinense de moribus
- Niccolò Longobardi (1559-1654) Epistola de Statu Siniensis; Recentissima de amplissimo regno

Das-Memo> Eurasian Latin Archive



Progetto

Notizie

Search...

Q

> ITA

EN

DAS-MeMo

Eurasian Latin Archive

Eurasian Latin Archive

Unità di ricerca

Wiki

Fonti e documenti

Documentazione progetto

Risorse utili

Eurasian Latin Archive

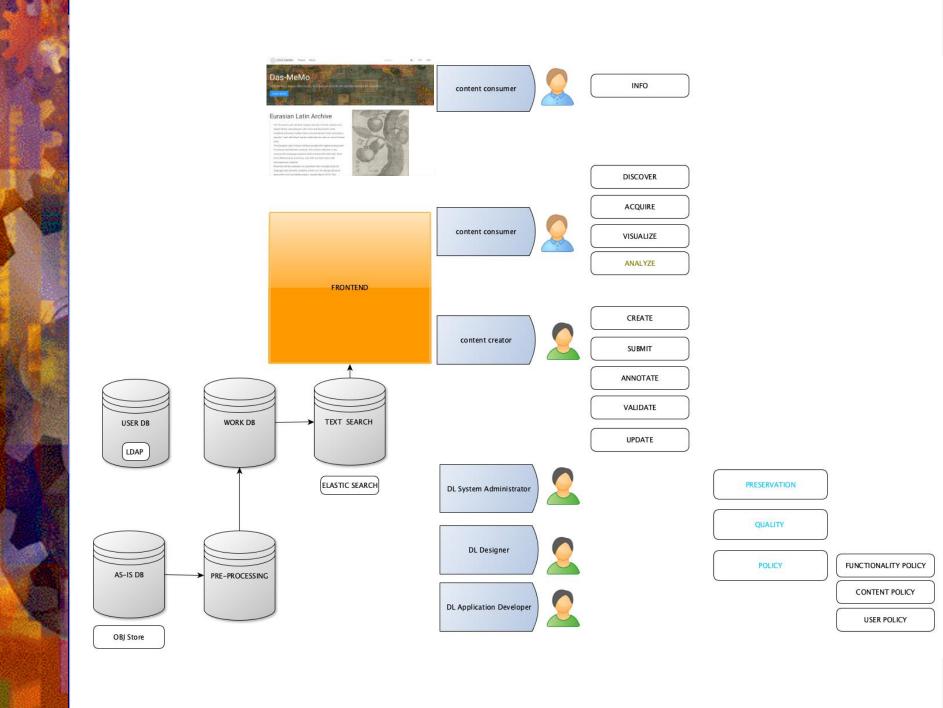


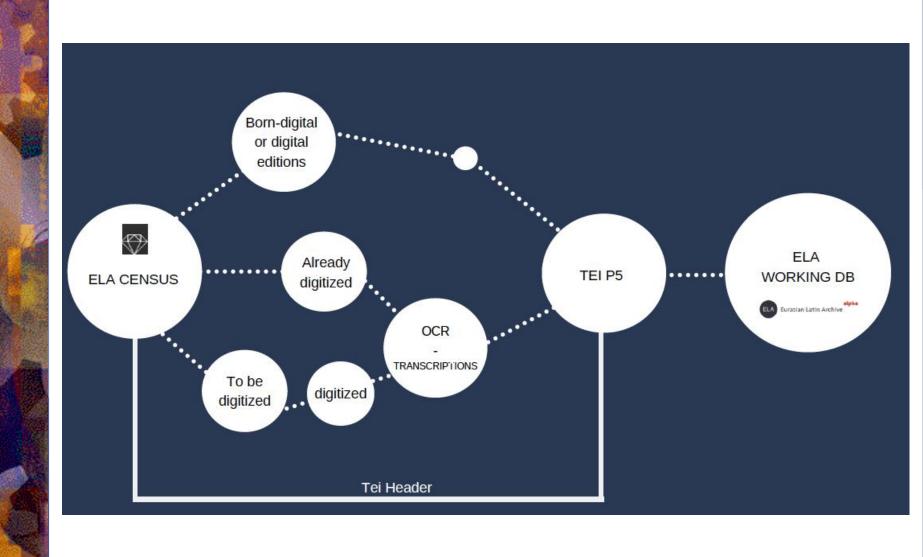
Il progetto "Eurasian Latin Archive" è nato per la creazione di un archivio digitale di testi e documenti latini riguardanti l'Asia Orientale, includendo in particolare una sezione dedicata alla "Via della seta latina", che comprende testi latini e latino-cinesi. La raccolta è finalizzata all'analisi linguistica comparativa sia interna sia con altri testi latini di epoche e aree diverse, sia con testi non latini di argomento omogeneo.

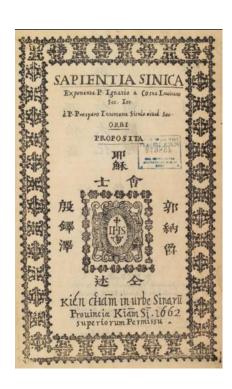
I materiali saranno fruibili attraverso una piattaforma dotata di strumenti per l'analisi linguistica e semantica, che è attualmente in fase di progettazione all'interno del progetto Das-MeMo, avviato a marzo 2018. La piattaforma si ispirerà al progetto Alim, la cui "Collezione Asiatica" rappresenta la fase embrionale dell'Eurasian Latin Archive.

ELA-by E. Carbé and N. Giannelli









restimoniu accipe cu anno atatis sua so Mandavinii agevet in Regno lui, trime seri spatio tanta illicò eo administrante revii in melius mute in causarit ac metu; quo factu est ut finitimi Regni lui causarit ac metu; quo factu est ut finitimi Regni lui cara al Proceres dolo usi ad Rege Regni lu dono miserint puellas, forma cantente gratia, alissa dotibus prostato. Rex munus admisit, et cu regni Primatibus itaillis audiendis indulsite, ut ad trimestre negori publici obeundi, et Regis adeundi uia precluderetur: que con relicia auli gori publici obeundi, et Regis adeundi uia precluderetur: que con relicia auli parvia, et munere publico, quod gerebat, contulit se ad Ragni lui filippia que est peter la guar parvia, et munere publico, quod gerebat, contulit se ad Ragni lui filippia que est peter la guar lui de contuita est redactus, imo in Regno te contuita acdifficultatibus inuita con che uin remuria est redactus, imo in Regno te viculis acdifficultatibus inuita con che uin mon semel est coquisteus; at Confucius per viculis acdifficultatibus inuita con che uintente qua illi innitar ecquis tande mini nocear zelus indefessus propagada toto imperio sana doctrina ad extremi usquita spiritui neo pracipue astuabat, qui patris uix se contines finibus etia ad peregrina regna et maria per ragrada desideriu iniecit. discipulos numeralit 3072, quos inter uirrute et litteris eminabat septuaginta duo; doctrina sua uti et discipulorii quatur gradus, seu ordines il-



Search

ELA Eurasian Latin Archive

Search the ELA archive using different techniques



Document List

The list of texts available on the Platform



Browse

Discover ELA by browsing the archive. Follow Authors, Titles, Time Periods and other topics



Tools

Linguistic and Topic Analysis Tools are under construction



Census

A first census of the documents that will constitute the archive



Help

Tips on how to use the platform and guidelines for documents use and sharing



28

Authors



63

Documents



193608

Words



314

Census

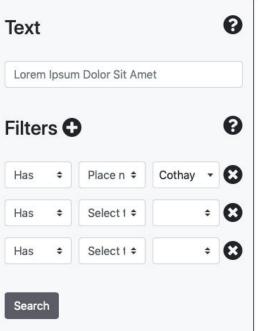
Has

Has

Has

THE PROJECT THE ARCHIVE THE WORKSHOP

DAS-MEMO



Author 1	Title	N Date ↑	
lohannes de Monte Corvino	Epistola (2)	1305	View/Download
Iohannes de Monte Corvino	Epistola (3)	1306	View/Download

Showing 1 to 2 of 2 entries

Previous

Next





Author	↑ Title ↑	Date ↑↓	
Bonaventura Ibañez	Epistola ad Episcopos Tunkinenses, 29 Ian. 1675	1675	View/Download
Bonaventura Ibañez	Brevis Relatio	1668	View/Download
Bonaventura Ibañez	Epistola ad Cardinales S.C.P.F., 5 nov. 1672	1672	View/Download
Bonaventura Ibañez	Epistola ad Episcopos Tunkinenses, 5 febbr. 1674	1674	View/Download
Inácio da Costa, Prospero Intorcetta	Sapientia Sinica	1662	View/Download

Showing 1 to 5 of 5 entries

Previous 1 Nex

Browse by:

- AUTHOR
- PLACE
- PERSON
- DATE
- DOCUMENT TYPE

• 1501/1600 **ᢒ** • 1601/1700 **ᢒ**

- SOURCE
- GENRE
- GEOGRAPHIC REFERENCE
- LANGUAGE



Publisher: ELA Date: 1326

Language: Latin

License: CC BY-NC-SA

Epistola

Andreas de Perusio 12..-1332

in: Sinica Franciscana, I: Itinera et relationes fratrum Minorum saeculi XIII et XIV, a cura di A. Van Den Wyngaert, Firenze, Quaracchi, 1929, 373-377

page: 373

Fr. Andreas de Perusio de Ordine Minorum fratrum divina permissione vocatus Episcopus, reverendo patri fratri Guardiano Perusini Conventus salutem et pacem in Domino sempiternam. Et infra sequitur.

1. Nara propter immensam terrarum mariumque distantiam, inter me et vos interiectam, vix sperare possum, quod littere ad vos per me transmisse ad manus vestras valeant pervenire. Et infra subsequitur. Noveritis me itaque cum bone memorie fratre Peregrino Coepiscopo, et mee peregrinationis individuo comite, post multos labores et langores, inedias, variaque incomoda, atque pericula in terra pariter et in mari, ubi fuimus rebus omnibus et etiam tunicis et habitibus spo

page: 374

liati, demum, Deo iuvante, ad <u>Cambaliensem civitatem</u>, que sedes est imperii <u>magni Chanis</u>, <u>anno dominice incarnationis millesimo CCC* XVIII</u>, ut credo, pervenisse; ubi secundum mandatum a <u>Sede Apostolica</u> nobis datum, Archiepiscopo consecrato, moram ibi per quinquennium ferme contraximus. Infra quod temporis spatium procuravimus alafa ab Imperatore magnifico pro victu et vestitu octo personarum. Est autem alapha impense, quas Imperator tribuit nunciis magnatorum, oratoribus, bellatoribus et diversarum artitum artificibus et ioculatoribus, pauperibus et diversis diversarum conditionum personis; que impense plurium latinorum Regum introitus expensasque transcendunt.

2. De divitiis, magnificentia et gloria huius magni Imperatoris, de vastitate Imperii, multitudine populorum, numerositate civitatum, et magnitudine earumdem, et de ordinatione imperii, in quo nemo adversus alium ausus est levare gladium, transeo, quia longum foret

