



UNIVERSITÀ DI PISA



UNIVERSITÀ DI PISA
FILOLOGIA, LETTERATURA E LINGUISTICA

CLaSSES

a corpus for linguistic studies
on non-literary Latin

Giovanna Marotta, Irene De Felice and Francesco Rovai

First LiLa Workshop: Linguistic Resources & NLP Tools for Latin

Tuesday 4th June

Structure of the talk

- ❑ CLaSSES
- ❑ Selection of texts
- ❑ Annotation scheme
- ❑ Search interface
- ❑ Future work

CLaSSES

- ❑ A digital resource which allows users to access **non-literary Latin texts** and to perform quantitative linguistic analysis on them
→ Website: <http://classes-latin-linguistics.fileli.unipi.it/>
- ❑ The texts collected have been tagged with **metadata** and **linguistic information**

CLaSSES

- ❑ A digital resource which allows users to access **non-literary Latin texts** and to perform quantitative linguistic analysis on them

→ Website: <http://classes-latin-linguistics.fileli.unipi.it/>

- ❑ The texts collected have been tagged with **metadata** and **linguistic information**



Users can evaluate the statistical incidence of linguistic phenomena and interpret them in the light of the socio-historical context of the Roman world

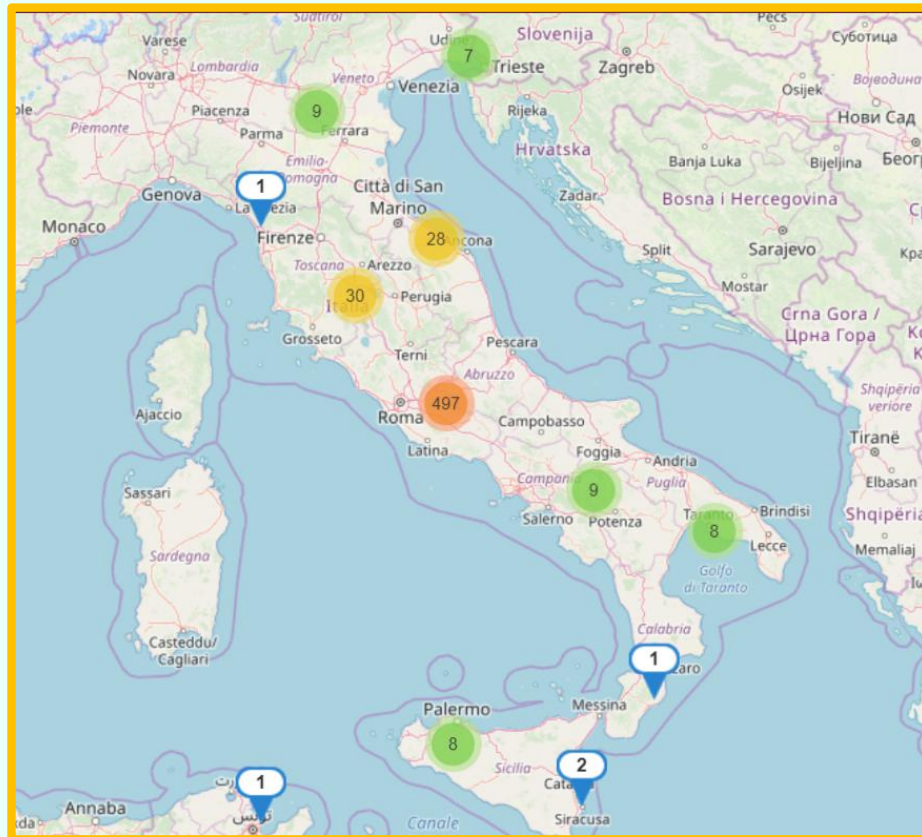
Selection of texts

The corpus aims to be representative of linguistic variation in the Latin language at different levels (Marotta 2015, 2016):

- ❑ **Diatopic level** → variation according to places and regions where Latin was spoken
- ❑ **Diachronic level** → variation among chronological periods
- ❑ **Diaphasic level** → variation among registers and text types (special focus on non-literary Latin)
- ❑ **Diastratic level** → variation according to social classes or social groups

Selection of texts

More than 1600 documents (~26000 tokens)



1244 inscriptions from
Rome and Italy

4th c. BC – 1st c. AD

Selection of texts

More than 1600 documents (~26000 tokens)

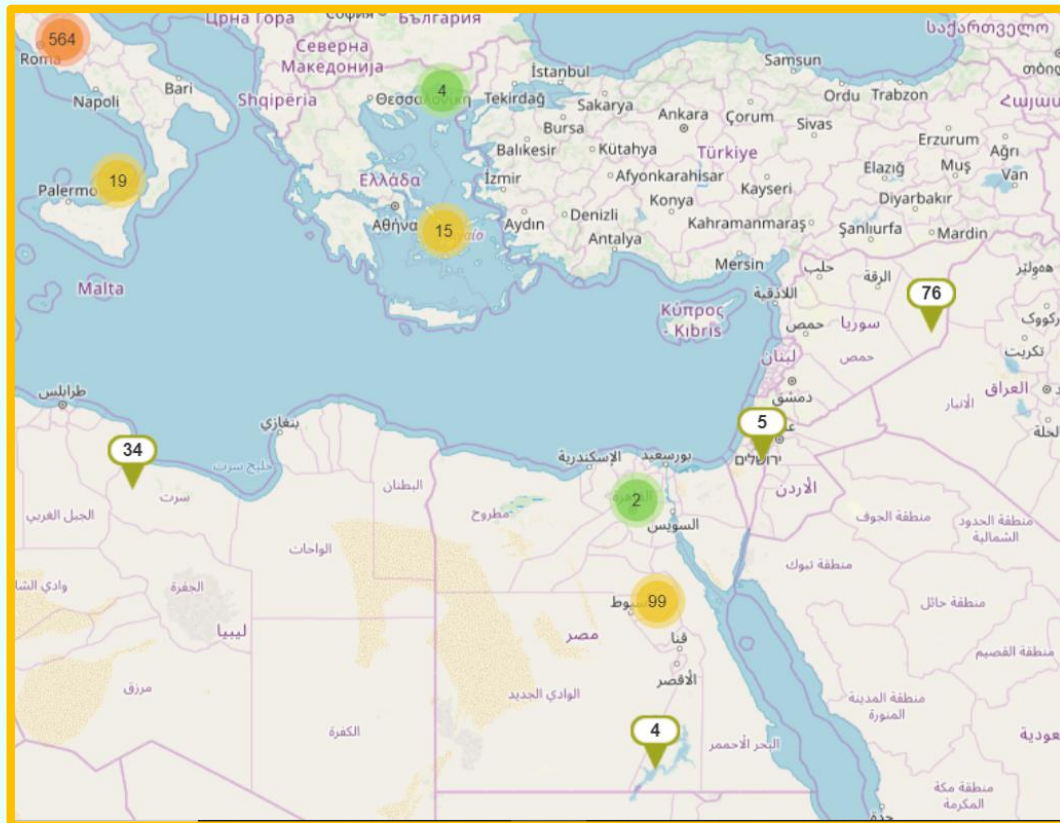


200 tablets from
Roman Britain
(Vindolanda)

1st c. – 3rd c. AD

Selection of texts

More than 1600 documents (~26000 tokens)



219 letters from
Egypt
and the **Eastern**
Mediterranean

1st c. BC – 6th c. AD

Selection of texts

Sources of non-literary texts:

❑ Digital resources, e.g.:

Corpus Inscriptionum Latinarum online

Epigraphische Datenbank Clauss-Slaby

Epigraphische Datenbank Heidelberg

EDR - Epigraphic Database Roma

Vindolanda Tablets Online

Selection of texts

Sources of non-literary texts:

□ Printed books, e.g.:

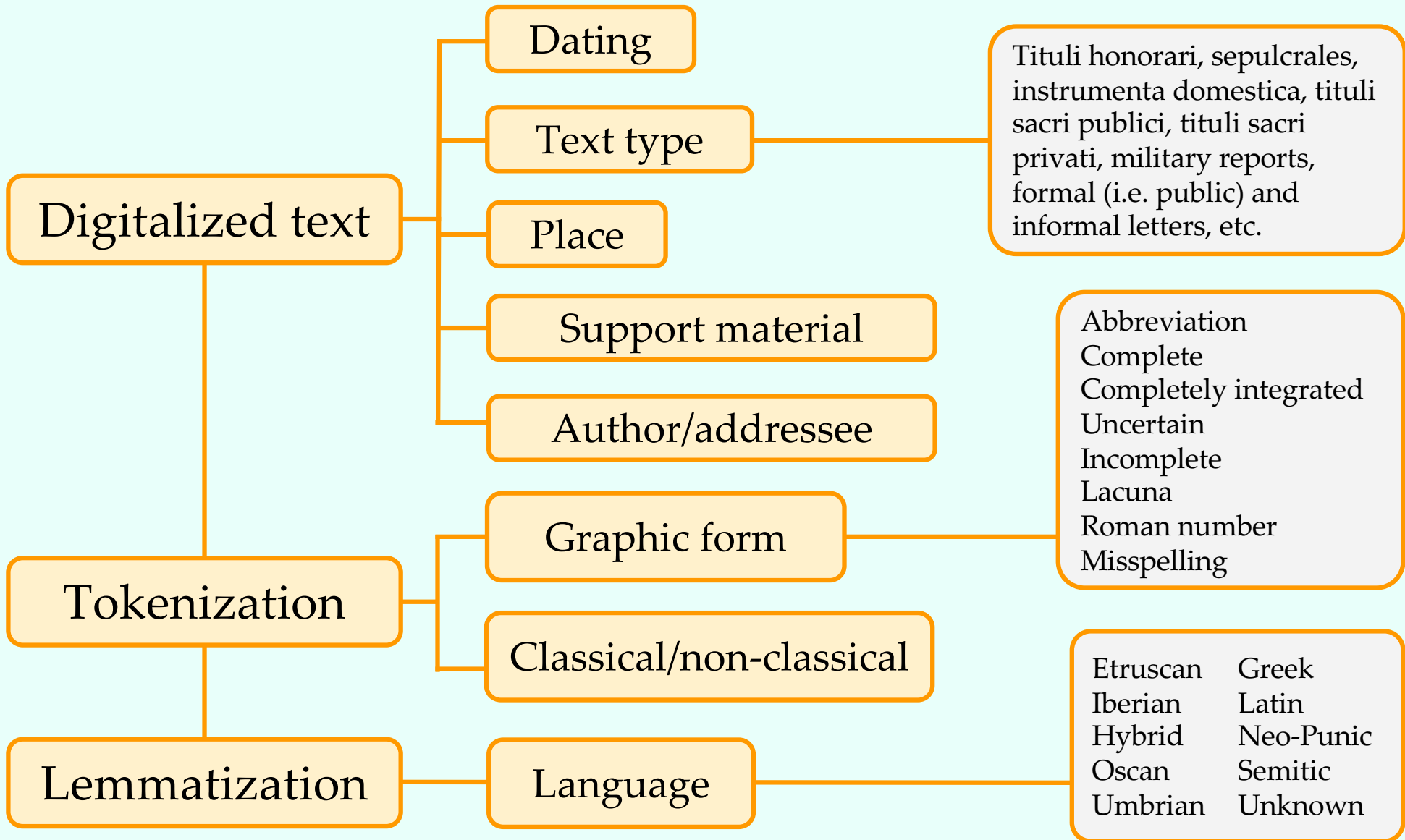
- Bowman, A.K. & Thomas, J.D. (1983), *Vindolanda: The Latin Writing-Tablets*, Society for the Promotion of Roman Studies, London.
- Bowman, A.K. & Thomas, J.D. (1994-2003), *The Vindolanda Writing Tablets: Tabulae Vindolandenses. Voll. 2-3*, British Museum Press, London.
- Cugusi, P. (1992), *Corpus Epistularum Latinarum Papyris Ostracis Tabulis servatarum*, 3 voll., Textus. Firenze, Gonnelli.
- Gordon, J. & Gordon, A. (1958), *Album of dated Latin inscriptions. Vol. 1*, Berkeley-Los Angeles, University of California Press.
- Lommatzsch, E. (1918, ed.), *CIL I² 2,1: Inscriptiones vetustissimae*, De Gruyter, Berlin.
- Wachter, R. (1987), *Altlateinische Inschriften*, Lang, Bern.
- Warmington, E.H. (1940), *Remains of Old Latin. Vol. 4: Archaic inscriptions*, Harvard University Press-Heinemann, Cambridge MA-London.

Selection of texts



[D]IOVEI IUNONEI MINERVAI
FALESCE QUEI IN SARDINIA SUNT
DONUM DEDERUNT MAGISTREIS
LLATRIUS KFC SALV[E]NA VOLTAI F
COIRAVERONT (CIL I² 364)

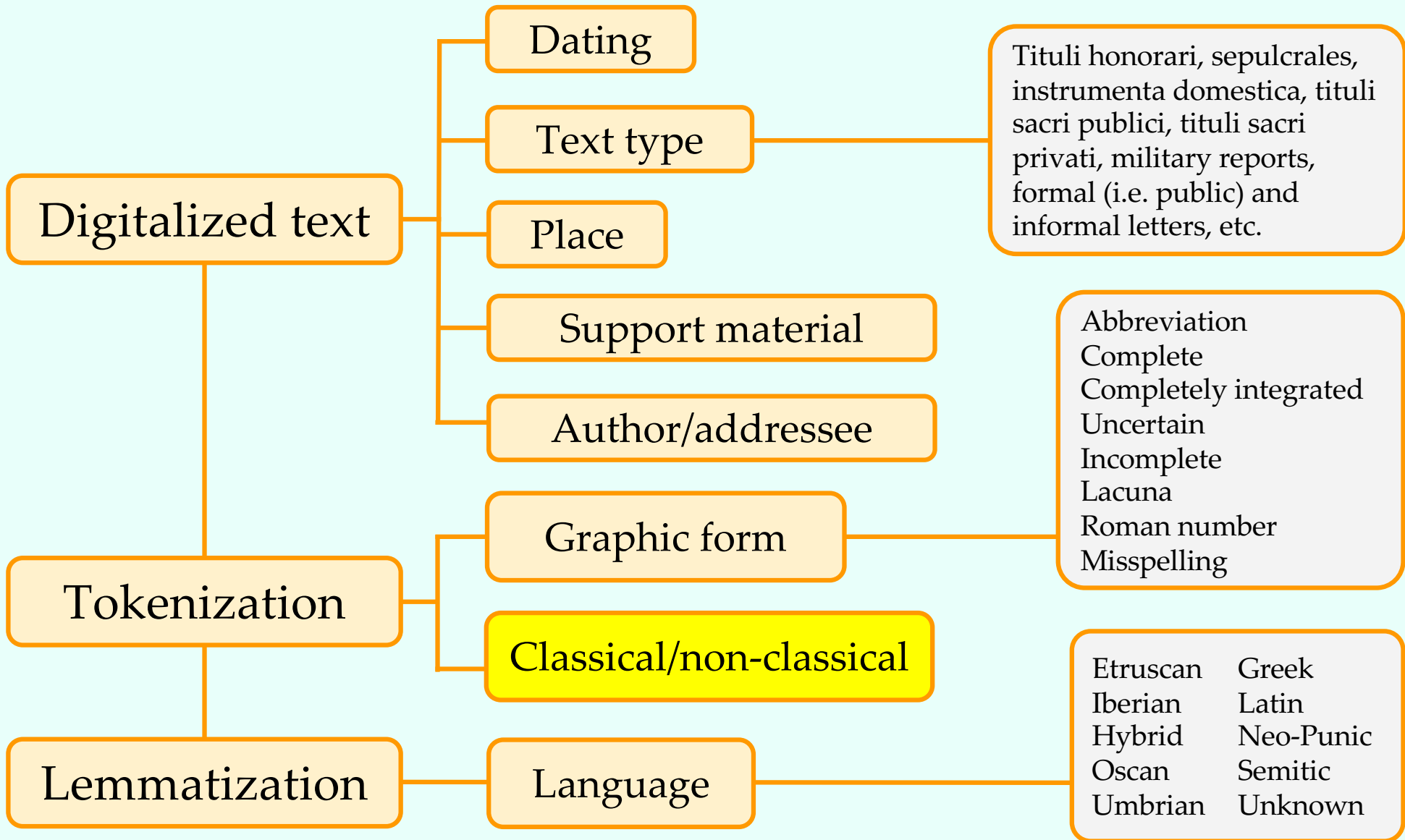
Annotation scheme (De Felice, Donati and Marotta 2015)



Annotation scheme (De Felice, Donati and Marotta 2015)

B	C	D	E	F	G	H	I	J	K	L	M
token	token_id	lang.	graphic form	lemma	cl./non-cl.	classic form	text type	dating	author	addressee	place
Victorl	BTT-180-19	Latin	Complete	Victor	non-classical	Victori	Numerata	104-120	Unkn. Caius II	-	Vindolanda
frumenti	BTT-180-2	Latin	Complete	frumentum	classical		Numerata	104-120	Unkn. Caius II	-	Vindolanda
lusu	BTT-180-20	Latin	Complete	iubeo	non-classical	iussu	Numerata	104-120	Unkn. Caius II	-	Vindolanda
Spectati	BTT-180-21	Latin	Complete	Spectatus	classical		Numerata	104-120	Unkn. Caius II	-	Vindolanda
comodati	BTT-180-22	Latin	Complete	comodatus	non-classical	commodati	Numerata	104-120	Unkn. Caius II	-	Vindolanda
m	BTT-180-23	Latin	Abbreviation	modius	-		Numerata	104-120	Unkn. Caius II	-	Vindolanda
xxvi	BTT-180-24	Latin	Roman number	-	classical		Numerata	104-120	Unkn. Caius II	-	Vindolanda
in	BTT-180-25	Latin	Complete	in	classical		Numerata	104-120	Unkn. Caius II	-	Vindolanda
follibus	BTT-180-26	Latin	Complete	follis	classical		Numerata	104-120	Unkn. Caius II	-	Vindolanda
tribus	BTT-180-27	Latin	Complete	tres	classical		Numerata	104-120	Unkn. Caius II	-	Vindolanda
patrl	BTT-180-28	Latin	Complete	pater	non-classical	patri	Numerata	104-120	Unkn. Caius II	-	Vindolanda
m	BTT-180-29	Latin	Abbreviation	modius	-		Numerata	104-120	Unkn. Caius II	-	Vindolanda
em[ensi]	BTT-180-3	Latin	Uncomplete	emetior	classical		Numerata	104-120	Unkn. Caius II	-	Vindolanda
xix	BTT-180-30	Latin	Roman number	-	classical		Numerata	104-120	Unkn. Caius II	-	Vindolanda

Annotation scheme (De Felice, Donati and Marotta 2015)



Classical / Non-classical

❑ Forms that do not clearly belong to the Classical literary language are annotated as **non-classical** (Marotta 2015, 2016)

❑ For each non-classical form, an **additional level** of linguistic annotation is added, comprising:

1. Corresponding classical form

dede

classical *dedit*

lemma *do*

Menervai

classical *Minervae*

lemma *Minerva*

regus

classical *regis*

lemma *rex*

2. Type of variation involved

Annotation of non-classical variants

For all non-classical forms, the type of variation is annotated, regarding:

- ❑ Vowel alternations (*Minervai* for *Minervae*)
- ❑ Consonant processes (*dede* for *dedit*)
- ❑ Morphophonological alternations – if the alternation in the vowel or consonant system occurs in a morphological position (as in *Minervai*, *dede*)

Annotation of non-classical variants

Main phenomena of vowel alternations:

- ❑ lowering/rising
- ❑ length (*I longa, apex, gemination*)
- ❑ syncope/epenthesis
- ❑ monophthongization

Annotation of non-classical variants

Main consonant processes

- ❑ assimilation/dissimilation
- ❑ length (gemination/degemination)
- ❑ deletion (word-final or nasal consonant)/insertion
- ❑ voice alternations
- ❑ deaspiration

Classical / Non-classical

Consonantism

Vocalism

final -s deletion

lowering $u > o$

Cornelio

Cornelius

-o nom. masc. II decl.

Lemma

Morphohonology

Search interface





CLASSES

Corpus for Latin Sociolinguistic Studies on Epigraphic textS



















Search Rome and Italy

Advanced search

Search

Reset filters  Visualize map   Export data 

Showing 1-50 of 11,804 items.





Token ID	Form	Lemma	Language	Graphic Form	Classical/Non-classical	Text Type	Place	Dating from	Dating to	Text	More info
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>		
CIL-I2-2490a-1	Gabinio	Gabinus	Latin	Complete	Non-classical	Instr. Dom.	-	200 BC	150 BC	 	
CIL-I2-23-1	Atilies	Atilius	Latin	Complete	Non-classical	Instr. Dom.	Roma (RM) / Roma	300 BC	200 BC	 	
CIL-I2-23-2	Saranes	Saranus	Latin	Complete	Non-classical	Instr. Dom.	Roma (RM) / Roma	300 BC	200 BC	 	
CIL-I2-23-3	C	Gaius	Latin	Abbreviation	-	Instr. Dom.	Roma (RM) / Roma	300 BC	200 BC	 	
CIL-I2-23-4	M	Marcus	Latin	Abbreviation	-	Instr. Dom.	Roma (RM) / Roma	300 BC	200 BC	 	
CIL-I2-23-5	f	filius	Latin	Abbreviation	-	Instr. Dom.	Roma (RM) / Roma	300 BC	200 BC	 	

Search interface



















Search Rome and Italy

Advanced search

Search

Reset filters  Visualize map   Export data 

Showing 1-50 of 11,804 items.

Token ID	Form	Lemma	Language	Graphic Form	Classical/Non-classical	Text Type	Place	Dating from	Dating to	Text	More info
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>		
CIL-I2-2490a-1	Gabinio	Gabinus	Latin	Complete	Non-classical	Instr. Dom.	-	200 BC	150 BC	 	
CIL-I2-23-1	Atilies	Atilius	Latin	Complete	Non-classical	Instr. Dom.	Roma (RM) / Roma	300 BC	200 BC	 	
CIL-I2-23-2	Saranes	Saranus	Latin	Complete	Non-classical	Instr. Dom.	Roma (RM) / Roma	300 BC	200 BC	 	
CIL-I2-23-3	C	Gaius	Latin	Abbreviation	-	Instr. Dom.	Roma (RM) / Roma	300 BC	200 BC	 	
CIL-I2-23-4	M	Marcus	Latin	Abbreviation	-	Instr. Dom.	Roma (RM) / Roma	300 BC	200 BC	 	
CIL-I2-23-5	f	filius	Latin	Abbreviation	-	Instr. Dom.	Roma (RM) / Roma	300 BC	200 BC	 	

Search interface

Atilies Back to index

Token ID	CIL-I2-23-1
Form	Atilies
Language	Latin
Graphic Form	Complete
Lemma	Atilius
Classical/Non-classical	Non-classical
Classical form	Atilii
Text Type	Instrumentum domesticum
Place	Roma (RM) / Roma
Dating	300-200 BC
Context	Atilies Saranes C M f
Inscription	Atilies Saranes C M f

Linguistic phenomena

Vowels	Consonants	Morphophonology
No linguistic phenomena	No linguistic phenomena	NOUN - 2nd declension - Plural: Nominative -e(i)s

Future work

Extension of
CLaSSES: section
Roman Britain

with 573 tablets
(~6500 tokens)

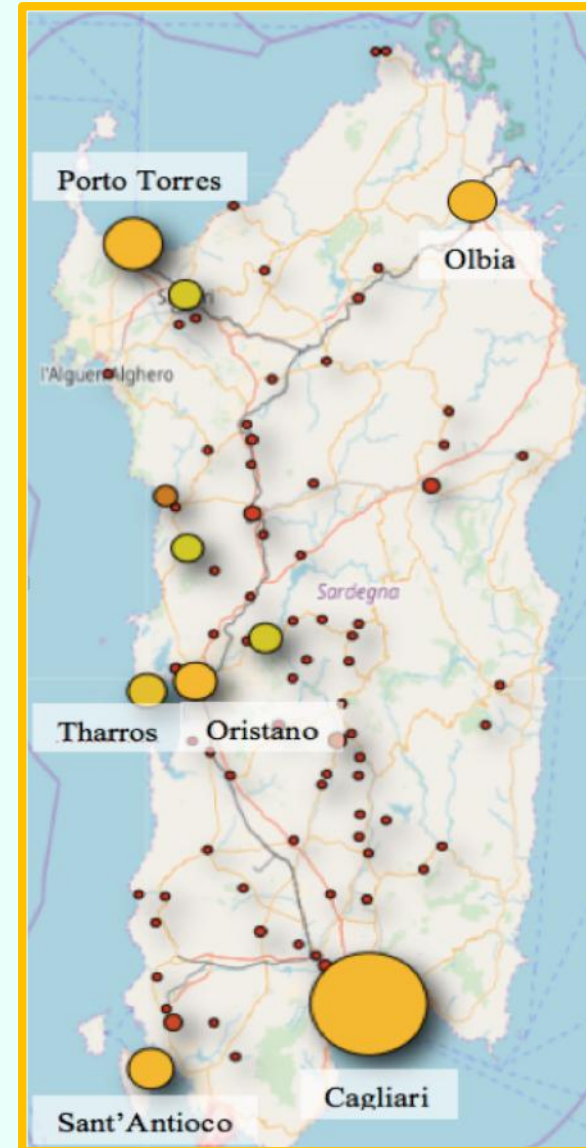


Future work

Creation of a
new section: **Sardinia**

- 1158 inscriptions
- ~14000 tokens

PhD thesis by Lucia Tamponi
on epigraphic Latin from Sardinia



THANK YOU!

Special thanks go to people who collaborated in this project:

Serena Barchi, Francesca Cotugno, Margherita Donati and Lucia Tamponi

First LiLa Workshop: Linguistic Resources & NLP Tools for Latin

Tuesday 4th June

References

De Felice, I., Donati, M., & Marotta, G. (2014). *CLaSSES: A new digital resource for Latin epigraphy*, in R. Basili, A. Lenci, & B. Magnini (eds.), *Proceedings of the First Italian Conference on Computational Linguistics CLiC-it 2014 and the Fourth International Workshop EVALITA 2014 (Pisa, 9-11/12/2014)*, Pisa University Press, Pisa, pp. 132-137.

De Felice, I., Donati, M., & Marotta, G. (2015), *CLaSSES: A new digital resource for Latin epigraphy*, in «Italian Journal of Computational Linguistics», 1, 1, pp. 119-130.

Donati, M. (2015), *Variazione e tipologia testuale nel corpus epigrafico CLaSSES I*, in «Studi e Saggi Linguistici», 53, 2, pp. 21-38.

Donati, M., Rovai, F. & Marotta, G. (2018), *Prospettive sociolinguistiche sul latino: un corpus per l'analisi dei testi epigrafici*, in A. Garcia Leal & C. E. Prieto Entrialgo (eds.), *Latin Vulgaire - Latin Tardif XI*, Georg Olms Verlag Ag, pp. 631-643.

Marotta, G. (2015), *Talking stones. Phonology in Latin inscriptions?*, in «Studi e Saggi Linguistici», 53, 2, pp. 39-63.

Marotta, G. (2016), *Sociolinguistica storica ed epigrafia latina. Il corpus CLaSSES I*, in «Linguarum Varietas», 5, pp. 149-159.

Marotta, G. & Tamponi, L. (2019), *Omission of final -s in Latin inscriptions: Time and Space*, «Transactions of the Philological Society», 117(1), pp. 79-95.

References

Tamponi, L. (2017), *Sull'alternanza vocalica <o> ~ <u> nelle epigrafi latine di epoca arcaica*, in Marotta, G. & Strik Lievers, F. (2017, a cura di), *Strutture linguistiche e dati empirici in diacronia e sincronia*, Pisa, Pisa University Press, pp. 111-132.

Tamponi, L. (2018), *I nominativi plurali in -e(i)s, -is della seconda declinazione nelle epigrafi latine di età arcaica: un caso di interferenza osca?*, in Costamagna, L., Di Domenico, E., Marcaccio, A., Scaglione, S. & Turchetta, B. (2018, a cura di), *Mutamento Linguistico e Biodiversità. Atti del XLI convegno annuale della Società Italiana di Glottologia*, Biblioteca della Società Italiana di Glottologia 41, Il Calamo, pp. 281-286.