



UNIVERSITÀ
CATTOLICA
del Sacro Cuore

eurac
research



UNIVERSITÀ
DI PAVIA

Græcissāre: Ancient Greek Loanwords in the LiLa Knowledge Base of Linguistic Resources for Latin

Greta Franzini^{3, 1}, Federica Zampedri², Marco Passarotti³,
Francesco Mambrini³, Giovanni Moretti³

¹Eurac Research, ²Università di Pavia, ³Università Cattolica Del Sacro Cuore - CIRCSE

CLiC-it 2020
March 1, 2021



This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme - Grant Agreement No. 769994.

Objective. Interoperability (FAIR data).

Method. Linked Data.

Status. Resources connected and upcoming connections:

▶ **Corpora**



▶ **NLP tools**



▶ **Lexica**

- Word Formation Latin: ca. 46,000 lemmas (Classical Latin)
- LatinAffectus: ca. 2,300 entries
- Latin WordNet: ca. 1,000 manually checked entries
- Etymological dictionary of Latin & the other Italic Langs.: ca. 1,400 entries**
- Index Graecorum Vocabulorum in Linguam Latinam: ca. 1,800 entries**
- Latin Vallex 2.0: Valency Lexicon
- Lewis & Short Dictionary

Passarotti M., Mambrini F., Franzini G., Cecchini F. M., Litta E., Moretti G., Ruffolo P., Sprugnoli R. 2020. 'Interlinking through Lemmas. The Lexical Collection of the LiLa Knowledge Base of Linguistic Resources for Latin', in Passarotti M. (ed.), «*Current Approaches in Latin Lemmatization*», *Studi e Saggi Linguistici*, LVIII (1) 2020, pp. 177-212. eISSN: 2281-9142. ISSN: 0085-6827. DOI: 10.4454/ssl.v58i1.277

Index of 1,763 Ancient Greek loanwords in the Latin language.

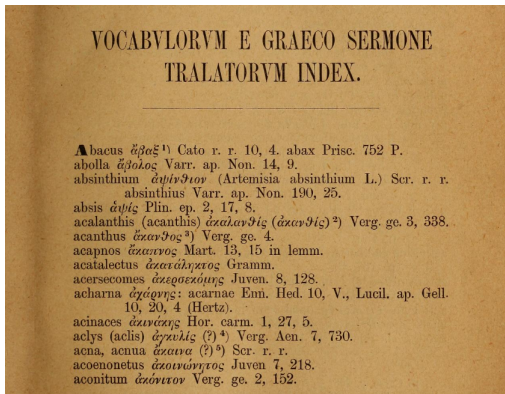


Figure: The *Index*, page 1. Internet Archive.

Optical Character Recognition (OCR)

- ▶ Latin: Manual correction of Internet Archive OCR
- ▶ Ancient Greek: Tesseract (v. 4.1.1) and manual postcorrection

Mapping

- ▶ Latin: Automatic matching and manual disambiguation against the LiLa lemma bank
- ▶ Ancient Greek: Manual matching against LSJ CITE Application

Modelling and interlinking

- ▶ *lemon*, *lime*, and *lemonETY* to model IGVLL lexical entries
- ▶ *Simple Knowledge Organization System (SKOS)* to pair IGVLL and LSJ Ancient Greek lemmas (*exactMatch*, *broadMatch*, *relatedMatch*)
- ▶ *CRM_{inf}* extension (argumentation model) of *CIDOC CRM* and *Open Vocabulary* to represent uncertainty as a confidence value
- ▶ *Dublin CoreTM Metadata Terms* to describe the resource

Example

muscus



Figure: Haeckel's *Muscinae* (1899).
Internet Archive.

μόσχος → *muscus*
“odoriferous plant, moss, musk, ...”

μόσχος

[um:cte2:hmt:lsj:chicago_md:n68835](#) [Link](#)

μόσχος (C), δ.

Α

musk, Aët. 1.131, Alex. Trall. 12; interpol. in Dsc. *Eup.* 1.145 (om. Wellmann). (Borrowed from Pers. *musk*.)

Figure: LSJ entry for μόσχος.

Etymology Link

<http://lila-erc.eu/data/lexicalResources/IGVLL/id/etylink/1120>

rdfs:label	Etymology Link
lemonEty:etyLinkType	borrowing
rdf:type	lemonEty:EtyLink
lemonEty:etySource	< http://lila-erc.eu/data/lexicalResources/IGVLL/id/etymon/1120 > ↳ μύσχος
lemonEty:etyTarget	< http://lila-erc.eu/data/lexicalResources/IGVLL/id/LexicalEntry/muscus > ↳ muscus

INVERSE RELATIONS

is **lemonEty:hasEtyLink** of 1 resource

Figure: Etymology in LiLa.

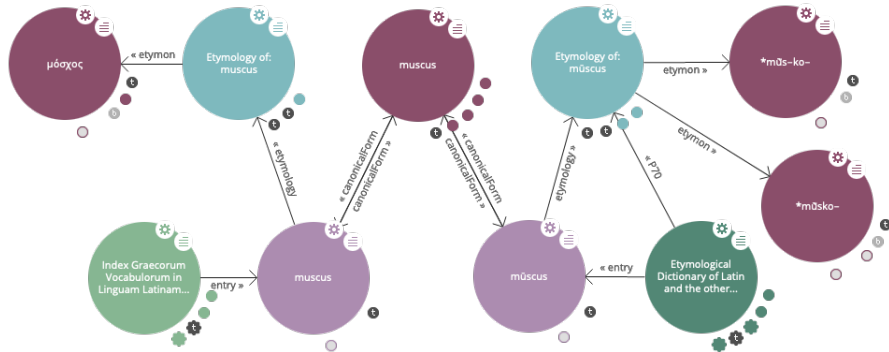


Figure: Etymology in LiLa.

`https://github.com/CIRCSE/index-graecorum-vocabulorum`

Thanks!

Get in touch



LiLa: Linking Latin

Università Cattolica del Sacro Cuore - CIRCSE Research Centre

Eurac Research



info@lila-erc.eu, greta.franzini@eurac.edu



<https://github.com/CIRCSE>



<https://lila-erc.eu>



@ERC_LiLa, @GretaFranzini



Largo Gemelli 1, 20123 Milan, Italy



This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme - Grant Agreement No. 769994.