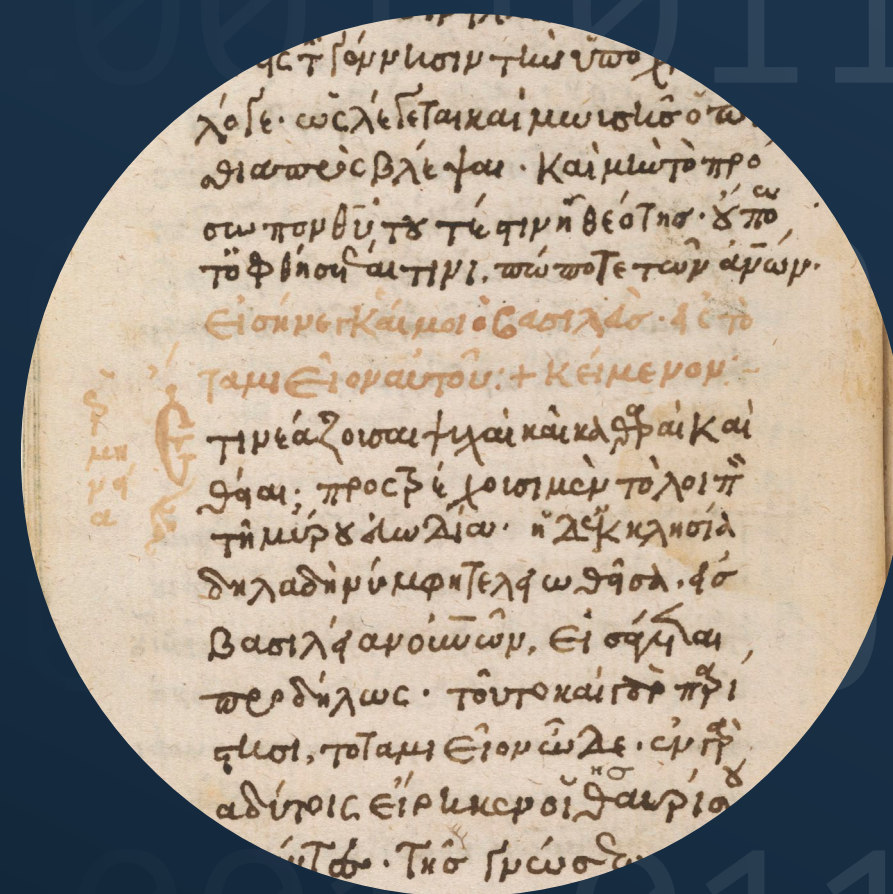


Logion

A no-code language-modeling tool for
premodern Greek philology



A textual critic engaged upon his business is not at all like Newton investigating the motions of the planets: he is much more like a dog hunting for fleas. If a dog hunted for fleas on mathematical principles, basing his researches on statistics of area and population, he would never catch a flea except by accident. They require to be treated as individuals; and every problem which presents itself to the textual critic must be regarded as possibly unique.

– A.E. Housman, “On the Application of Thought to Textual Criticism,” 1921

About Logion

The Princeton Logion Project aims to develop a language modeling tool that helps preserve and elucidate the global archive premodern Greek texts.



Project Goals



Accessible software to assist philologists



**Open-access online collection of
Greek (and Latin?) texts**



PERSEUS DIGITAL LIBRARY

GREGORY R. CRANE, EDITOR-IN-CHIEF
TUFTS UNIVERSITY



THESAURUS LINGVAE GRAECAE ®

A Digital Library of Greek Literature

Project Director: Maria Pantelia

The Software

Objective: LM tool to meaningfully assist in philological tasks by providing plausible, concrete suggestions

Requirements:

- α. no-code**
- β. local device (i.e. no HPC)**
- γ. cross-platform**
- δ. reproducible results**
- ε. philologically sound results**

Gap-filling

πάντα - καὶ οὐδὲν μένει



πάντα [χωρεῖ] καὶ οὐδὲν μένει

Gap-filling

πάντα χ- καὶ οὐδέν μένει



πάντα χ---- καὶ οὐδέν μένει



πάντα χ[ωρεῖ] καὶ οὐδέν μένει



Preliminary results for gap-filling Byzantine Greek

Model	Transmitted			Emendation		
	@1	@10	MRR	@1	@10	MRR
Tiresias	0.0842	0.3366	0.1574	0.1832	0.4010	0.2538
Logion BERT	0.0311	0.1244	0.0593	0.1140	0.2280	0.1477
SyllaMoBert	0.0259	0.0363	0.0292	0.0052	0.0155	0.0104
GreBERTa	0.0000	0.0518	0.0140	0.0466	0.0984	0.0625
Greek Char-BERT	0.0050	0.0792	0.0301	0.0198	0.0693	0.0336

Homer, Iliad, I.i-vii

μῆνιν ἄειδε θεὰ Πηληϊάδεω Ἀχιλῆος
οὐλομένην, ἣ μυρὶ Ἀχαιοῖς ἄλγε' ἔθηκε,
πολλὰς δ' ἰφθίμους ψυχὰς Ἄϊδι προΐαψεν
ἡρώων, αὐτοὺς δὲ [—] τεῦχε κύνεσσιν
οἴωνοῖσί τε πᾶσι, Διὸς δ' ἐτελείετο βουλή,
ἐξ οὗ δὴ τὰ πρῶτα διαστήτην ἐρίσαντε
Ἄτρεΐδης τε ἄναξ ἀνδρῶν καὶ δῖος Ἀχιλλεύς

Parse letters

Process string at character level

-classify as vowel (V) or consonant (C)

Store character info in **_PhoneticUnit** classs

λόγος



kind: V
text: ó
base: o
is_diphthong: False
has_iota_sub: False
has_macron: False
has_breve: False
is_double: False
word_initial: False

_PhoneticUnit

kind (str) -- vowel (V) or consonant (C)

text (str) -- original character w/ diacritics

base (str) -- character sans diacritics

is_diphthong (bool) -- True/False

has_iota_sub (bool) -- True/False

has_macron (bool) -- long?

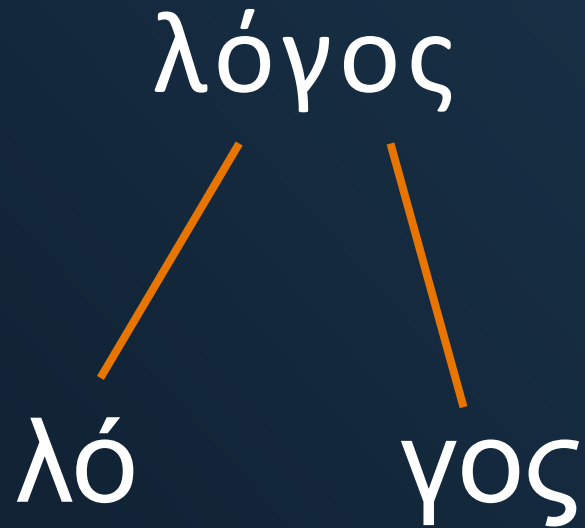
has_breve (bool) -- short?

is_double (bool) -- double consonant (ζ, ξ, ψ)?

word_initial (bool) -- 1st letter of word?

Build syllables

Function divides string into syllables
- ignores space + punct



Syllable anatomy

onset -- optional initial consonant

nucleus -- mandatory vowel/diphthong

coda -- optional end consonant

Rules

- one vowel/diphthong per syllable
- lone C starts syllable
- 2 Cs -> 1st closes prev. syllable
 - ἔκτος
- plosive+liquid/nasal -> both start syllable
 - πατρός
- double consonant opens+closes syllable
 - ῥίψε

Measure (+synizesis)



_assign_syllable_weight

For each syllable:

- > Last syllable of line? ----yes?----> X
- > Closed syllable? ----yes?----> L
 - > exception: σ before plosive ----> X
- > Muta cum liquida ----> L or X
- > Long vowel followed by vowel? ----> X
- > Short vowel followed by λ , ρ , μ , ν , or σ --
--> X
- > Short vowel followed by digamma list ----
> X
- > Intrinsic vowel length ----> L,S,X

Check against template

L	L		L	L		L	L		L	L		L	X
L	S	S	L	S	S	L	S	S	L	S	S	L	X

L	X	X	L	S	X	L	X	S	X	S	X	L	X	X	L	X
ἄνδρα μοι ἔννεπε,			μοῦσα, πολύτροπον,			ὃς μάλα πολλὰ										

L X S L X L X S L S X X S S L X
πλάγχθη, ἐπεὶ Τροίης ἱερὸν πτολίεθρον ἔπερσεν

Error correction

Μετὰ δὲ τὴν βασιλείαν , οὕπῳ ἡλίου τὸν ἐνιαύσιον κύκλον
περιοδεύσαντος ὁμοῦ δὲ τίκεται παιδίον τῷ βασιλεῖ , καὶ
τοῦ βασιλείου εὐθὺς καταξιοῦται ὀνόματος · τὸ δὲ ἄλλῳ ὡς
πρὸ τῆς βασιλείας γεγεννημένῳ , ὃ τε θαυμάσιος Μιχαὴλ
καὶ ὁ μετὰ τοῦτον Ἀνδρόνικος ἰδιώτα ἦστην

τε

Error correction

Μετὰ δὲ τὴν βασιλείαν , οὕπῳ ἡλίου τὸν ἐνιαύσιον κύκλον
περιοδεύσαντος ὁμοῦ **[MASK]** τίκτεται παιδίον τῷ
βασίλει , καὶ τοῦ βασιλείου εὐθὺς καταξιοῦται ὀνόματος .
τὸ δὲ ἄλλῳ ὥς πρὸ τῆς βασιλείας γεγεννημένῳ , ὃ τε
θαυμάσιος Μιχαὴλ καὶ ὁ μετὰ τοῦτον Ἀνδρόνικος ἰδιώτα
ἦστην



γὰρ	.35
τε	.3
στην	.22
ταισ	.15

Error correction

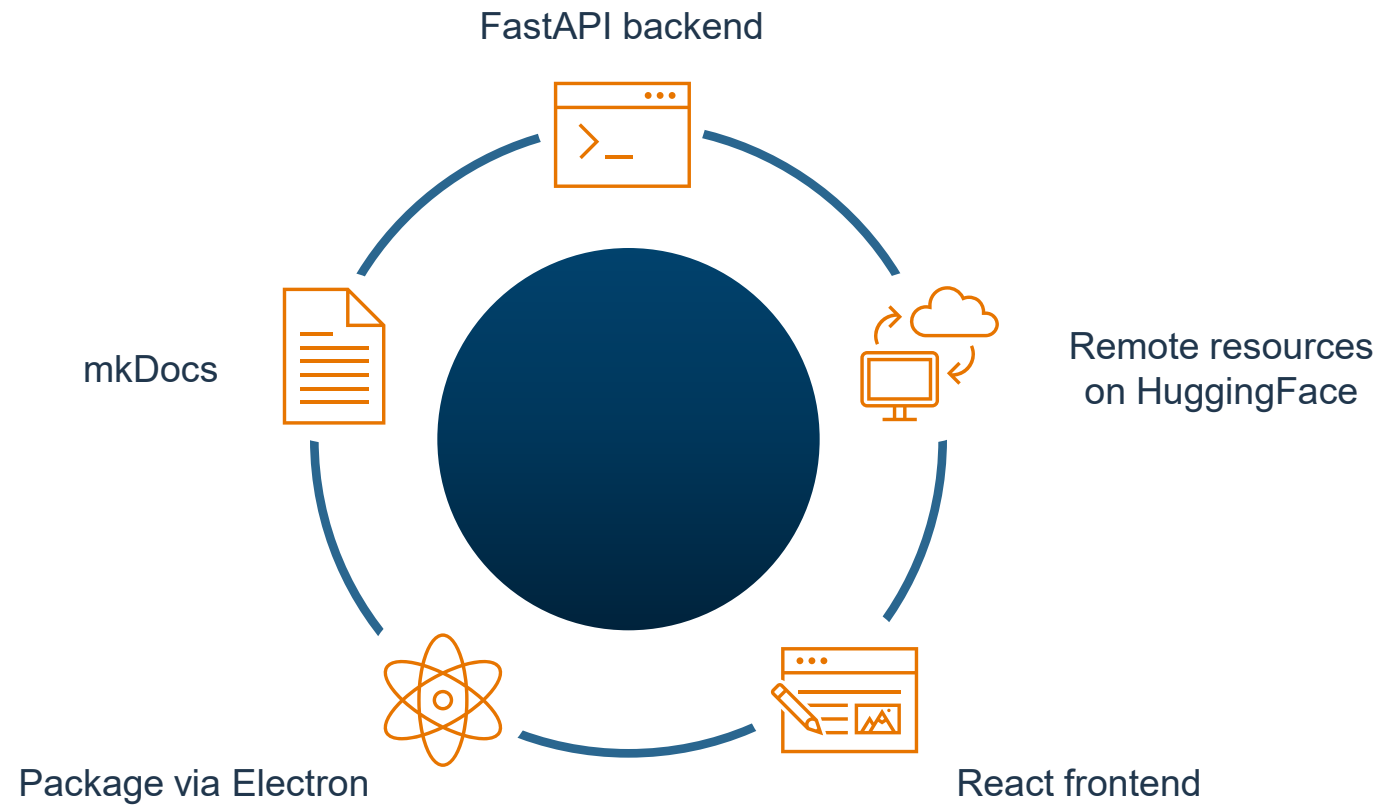
Μετὰ δὲ τὴν βασιλείαν , οὕπω ἡλίου τὸν ἐνιαύσιον κύκλον
περιοδεύσαντος ὁμοῦ **[MASK]** τίκτεται παιδίον τῷ
βασιλεῖ , καὶ τοῦ βασιλείου εὐθὺς καταξιοῦται ὀνόματος ·
τὸ δὲ ἄλλω ὡς πρὸ τῆς βασιλείας γεγεννημένω , ὃ τε
θαυμάσιος Μιχαὴλ καὶ ὁ μετὰ τοῦτον Ἀνδρόνικος ιδιώτα
ἦστην

LD = 1

{ τε

.3

Logion App



App Versions



Desktop App

Downloadable app that runs on local hardware. MacOS and Windows compatible.

MMF user testing



Open OnDemand

Select users register as Princeton affiliates for HPC access. Faster. Useful for large-scale philological tasks. Interested in deployment with other institutions that use OOD.

MMF user testing



CLI

CLI for model training/fine-tuning, text processing, and large-scale error report generation. For more technically savvy users interested in LM experimentation and training own models.

In development

Logion Philological Tasks



Gap-filling prediction

Encoder model



Error-word detection

Encoder model / Discriminator



Fragmentary text completion

Decoder model



Machine translation

Seq2Seq model

Future developments

- **Multi-word gap-filling**

- **Translation**

GH: <https://github.com/princeton-logion/>

HF: <https://huggingface.co/princeton-logion>

Demo