



IUSS

Scuola Universitaria Superiore Pavia

L'esametro latino nelle nuove prospettive linguistico-computazionali.

Progetto per il corso "Macchine Parlanti: Spiegazioni Linguistiche e Complessità"

A cura di:

- Alberto Anelli
- Antonio Maresca
- Alberto Pio Tatulli

METRICA E POESIA

Ars: ante enim carmen ortum est quam observatio carminis



La metrica, all'interno dello studio del latino, si configura esclusivamente come indagine a posteriori, scoperta razionale delle leggi insite nel linguaggio dei poeti.

Ars presuppone l'opera, non la crea.

Fine primario della metrica: essere uno strumento per l'analisi della creazione poetica.

Ritmo accentativo e ritmo quantitativo

Ritmo: ῥυθμός

Valore fonemático della quantità.

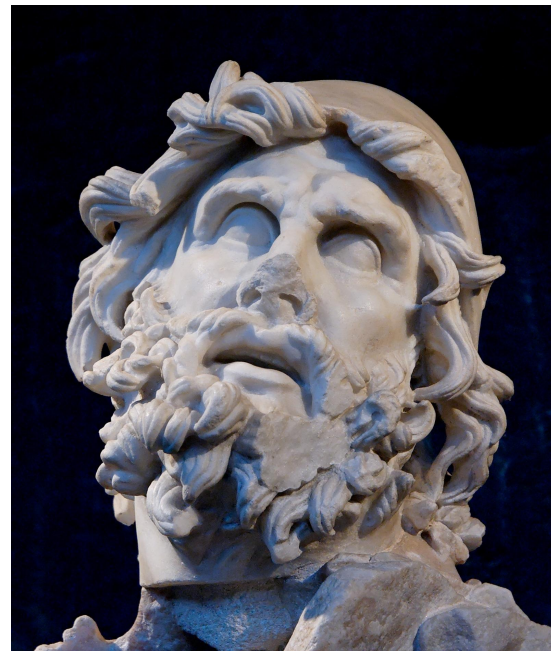
Saturnio → assunzione degli schemi metrici greci

Misura del ritmo → metro → piede (dattilo, spondeo, trocheo...)

caratteristica del metro: articolazione in due movimenti: **ARSI** e **TESI**.

ICTUS: (“pollicis ictus” Orazio; “pedum et digitorum ictus”) → colpo, meccanica percussione

Inversione dei termini: (Aftonio, Atilio Fortunaziano...): “elatio vocis” e “depositio vocis”



ESAMETRO LATINO



Sei piedi: (sei alternanze *arsis/tesi*): schema teorico mai realizzato

Catalessi (trocheo finale)

Primi quattro piedi: possibilità sia di dattilo che di spondeo



quinto piede: tende a conservare immutato il dattilo

variabilità iniziale-fissità finale.

32 combinazioni possibili.

Cesure: maschile (metro ripartito fra arsi e tesi); femminile (incide sulla tesi); centrale.

Funzioni stilistiche della cesura.

Nuove possibili interpretazioni

PAWŁOWSKI, KRAJEWSKI, EDER (2010) :
minor ritmicità dell'esametro latino rispetto a quello greco, dovuta a diverse cause:

1. maggior frequenza di sillabe lunghe
2. versificazione latina: impianto di quella greca
3. appartenenza al registro scritto e non orale.

PAWŁOWSKI e EDER (2001)
Ipotesi: convivenza di quantità e accento metrico.
Risultati: i testi latini tenderanno a suonare più ritmici recitati con accento metrico, rispetto a quando solo la quantità o sia la quantità che l'accento metrico giocano un ruolo nell'esecuzione orale.



POSSIBILI APPLICAZIONI PRATICHE

Una macchina che sia in grado non solo di scandire esametri, ma anche di riconoscere lo stile di un autore, può avere diverse applicazioni in campo filologico:

1. Utilizzo da parte di poeti latini contemporanei
2. Attribuzione di frammenti spuri ad un autore piuttosto che a un altro
3. Riconoscimento di parti metriche all'interno di testi prosastici
4. Ecdotica: ricostruzione del testo partendo dalla "Fortuna"



Un possibile algoritmo?

L'algoritmo si prefiggerà l'obiettivo di convertire in dati e numeri l'impronta stilistica dell'autore, in modo da facilitare operazioni di carattere quantitativo e statistico che possano avere eventualmente risvolti pratici, come l'attribuzione di un testo ad un autore mediante l'affinità dei valori individuati.



Tools attualmente reperibili: Pedecerto



Ármă vîrúmquē | cǎnó, | Trôlǎē | quī | primūs āb ōris

DDSS

Ítālǐām, | fátó | prǒfŭgús, | Lāvínǐaque vénit

DSDS



lītŏrǎ, mŭltum_ŭlle_ēt | tērrís | iāctātūs ēt ālto

DSSS

ví sŭpērŭm | sǎevǎē | mēmōrēm | lŭnŏnīs ōb íram;

DSDS

mŭltā quōque_ēt | bēllŏ | pǎssūs, | dŭm | cōndērēt ūrbem,

DSSS

ínferrētquē | dēŏs | Lǎtíŏ, | gēnŭs | ūndē Lǎtínŭm,

SDDD

Álbāníquē | patrēs, | ātque_āltǎē | mŏēnǐā Rŏmǎē.

DSSS

Mŭsǎ, mīhí | cāusās | mēmōrǎ, | quŏ | nŭmīnē lǎēso,

DSDS

quídvē dŏlēns, | rēgínǎ | dēŭm | tŏt | vŏlvērē cāsus

DSDS

ínsígnēm | pīētátē | vîrŭm, | tŏt | ādířē lǎbŏres

SDDD

ímpŭlērít. | Tǎntǎēne_ānīmís | cǎeléstībŭs írǎē?

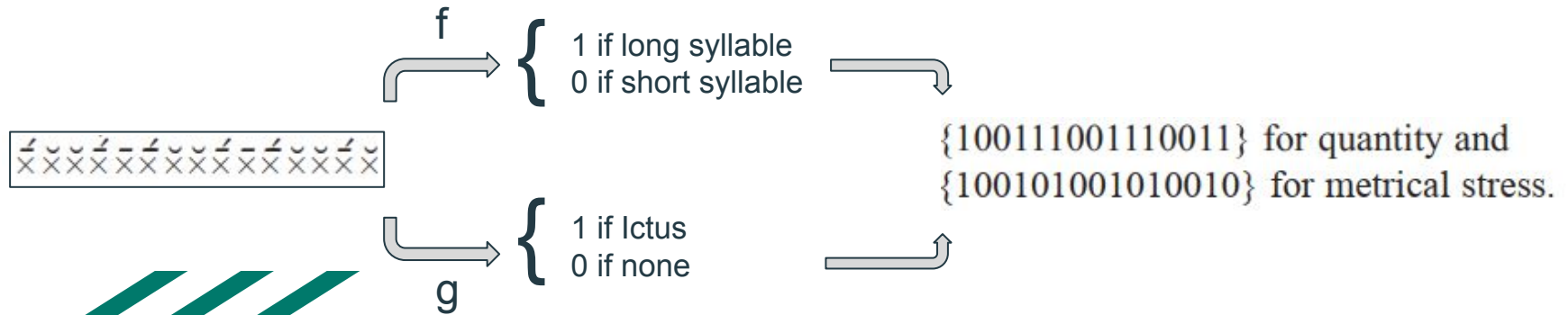
DSDS

```
<line name="1" meter="H" pattern="DDSS">
  <word sy="1A1b" wb="CF">Arma</word>
  <word sy="1c2A2b" wb="CF">                </word>
  <word sy="2c3A" wb="CM">cano,</word>
  <word sy="3T4A" wb="CM">Troiae</word>
  <word sy="4T" wb="DI">qui</word>
  <word sy="5A5b" wb="CF">primus</word>
  <word sy="5c" wb="DI">ab</word>
  <word sy="6A6X">oris</word>
</line>
<line name="2" meter="H" pattern="DSDS">
  <word sy="1A1b1c2A" wb="CM">Italian,</word>
  <word sy="2T3A" wb="CM">fato</word>
  <word sy="3b3c4A" wb="CM">profugus,</word>
  <word sy="4T5A5b5c" wb="DI">Laviniaque</word>
  <word sy="6A6X">venit</word>
</line>
```


La nostra (più semplice) scommessa: n-grams

“An n-gram is a collection of n successive items in a text document that may include words, numbers, symbols, syntagms, punctuation, ...”

Per un'analisi stilometrica più vicina ai metodi quantitativi della statistica e della programmazione, possiamo intendere gli n-grams come elementi di uno spazio vettoriale, e dunque trattare ciascun verso come un vettore. Ciò legittima ulteriormente l'approccio adottato.



La conversione in stringhe binarie (vettori) dà senso all'introduzione di concetti tipicamente statistici quali la media, la varianza, la covarianza, e consente l'introduzione delle funzioni di autocorrelazione (**ACF**) e di autocorrelazione parziale (**PACF**), che “macroscopicamente” riflettono la sequenzialità dei vari sintagmi all'interno del verso. Inoltre, trattando questo insieme di stringhe come una serie temporale, possiamo applicare i mezzi computazionali del modello ARIMA (**AutoRegressive Integrated Moving Average**), già ampiamente utilizzato in studi di questo genere.

A time series $\{x_1, x_2, \dots, x_n\}$ is defined as a series of realisations of a random variable. The independent variable t , which traditionally stands for real time, **is replaced in textual research by syntagmatic time**, which corresponds to the sequential order of linguistic units. The notion of an instant on the axis of real time thus finds its counterpart in the notion of a **position in the linear arrangement of text**. The interval between two realisations of a series at t_i and t_j is referred to as the lag and marked as $k = t_j - t_i$.

	ACF	PACF
AR(p)	dies out	truncates at lag $p + 1$
MA(q)	truncates at lag $q + 1$	dies out
ARMA(p, q)	dies out	dies out

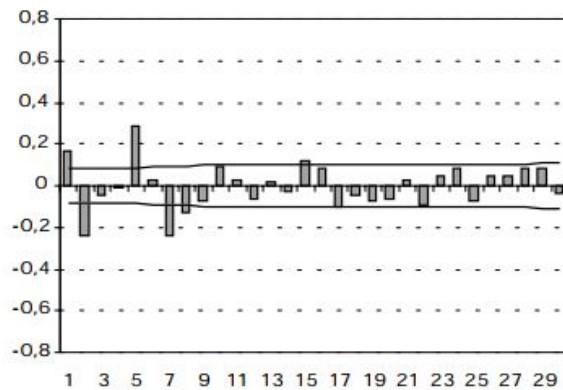


Fig. 1. ACF for the quantity series.

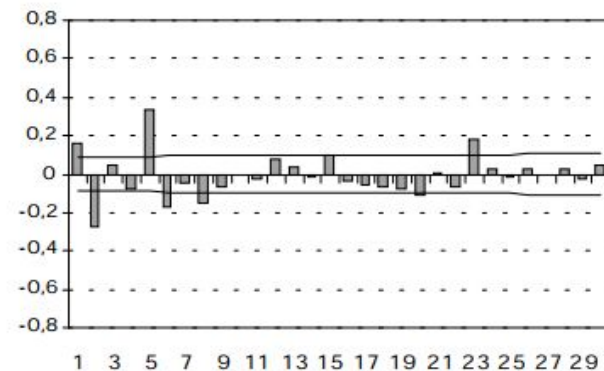


Fig. 3. PACF for the quantity series.

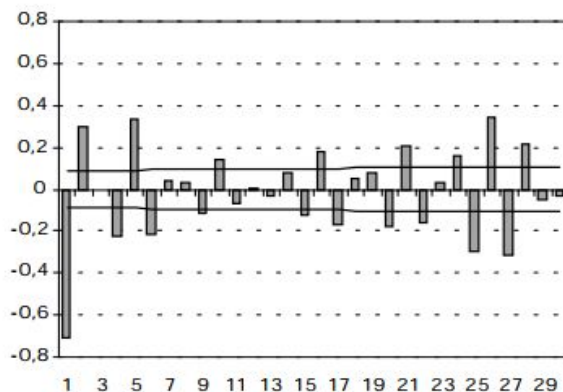


Fig. 2. ACF for the stress series.

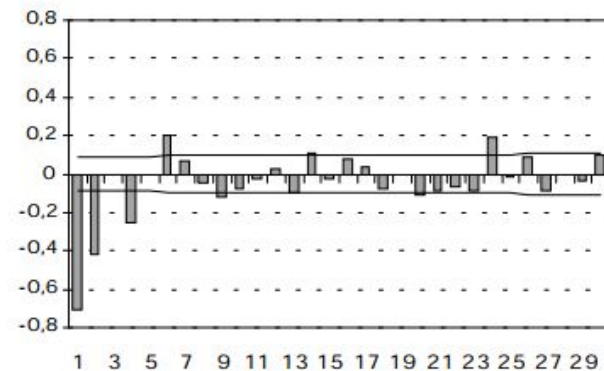
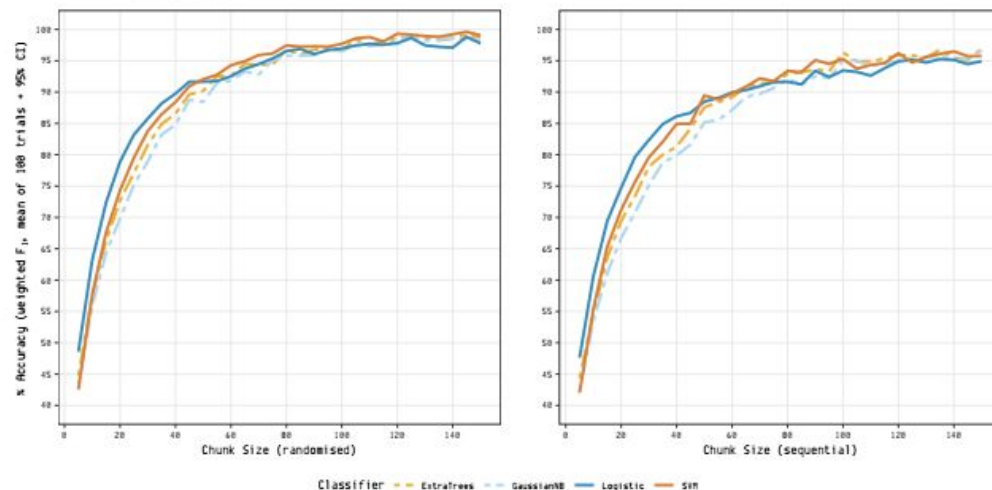


Fig. 4. PACF for the stress series.

Una via alternativa: “generic chunking”

Dopo che ogni verso è stato convertito in un opportuno vettore, questo può in teoria essere considerato già un dato importante per i valori statistici considerati. Tuttavia, risulta più efficace raggruppare i versi in “chunks” di lunghezza fissata. L’obiettivo è di valutare i valori medi delle occorrenze di ciascuna successione delle possibili 32 per un esametro e ottenere dei dati relativi allo stile compositivo “generico” di un dato autore. E’ inoltre utile anche randomizzare i versi nei vari chunks (per normalizzare le variazioni stilistiche dovute, ad esempio, ad una scena particolare e localizzata nel testo).

Figure 2: Classification Accuracy by Chunk Size: Randomised vs Sequential Chunks



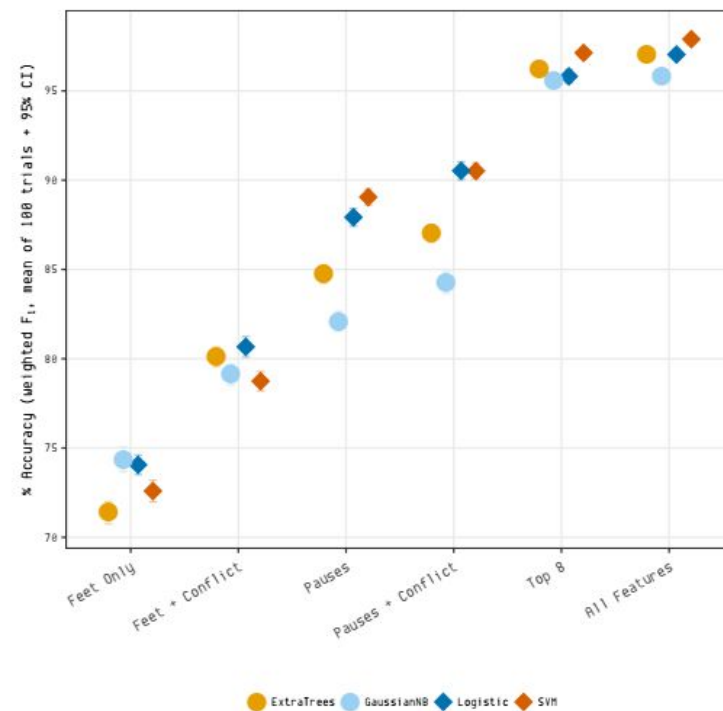
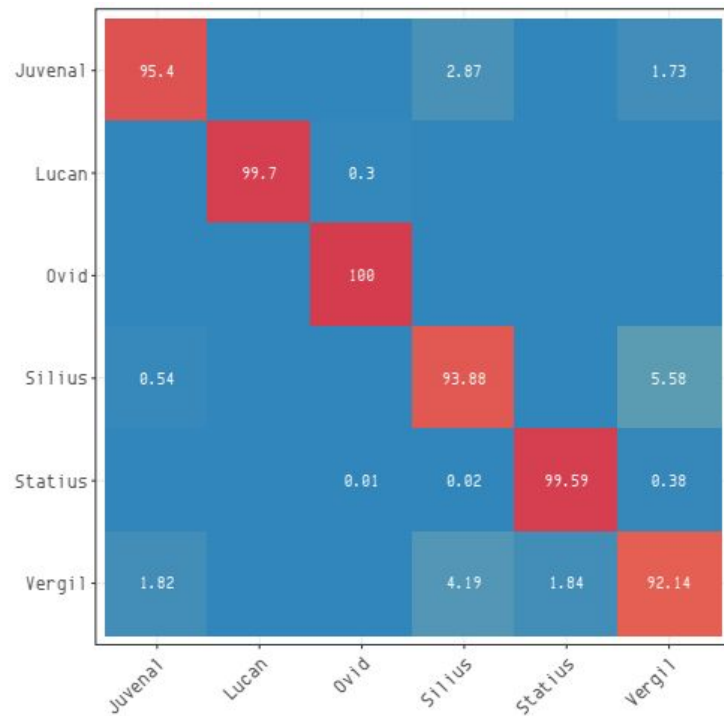
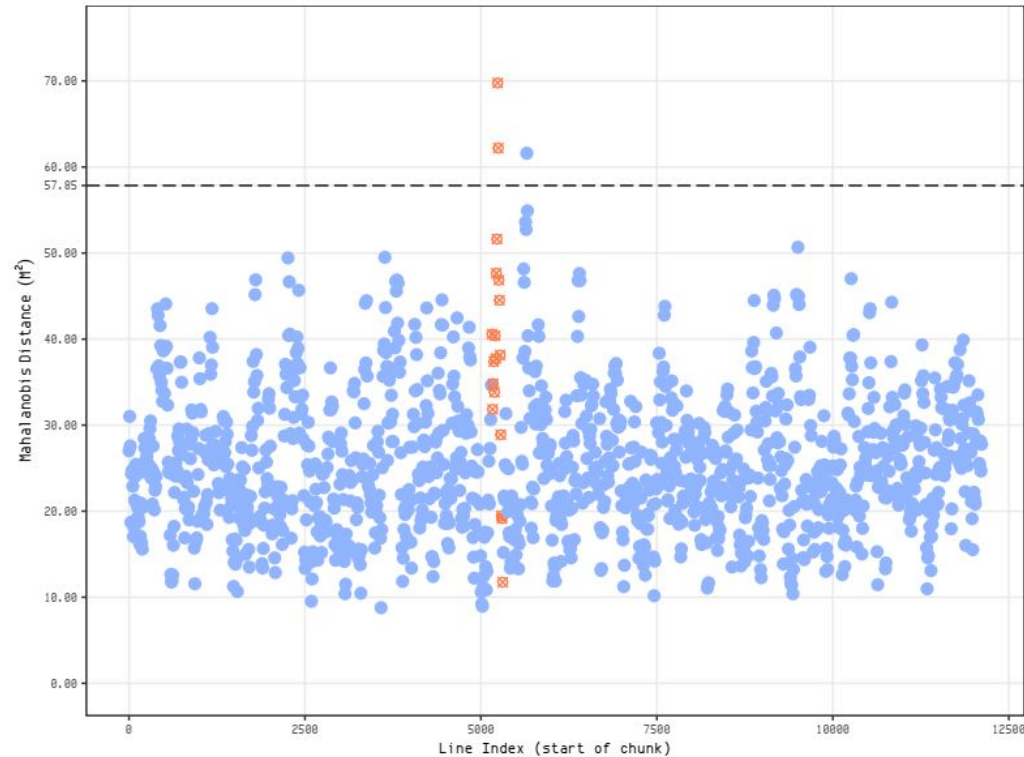
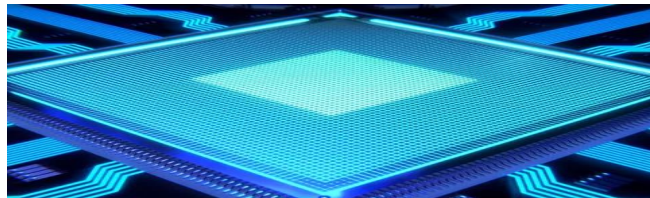


Figure 6: The M^2 distance through the *Punica*, taking a rolling window of 81-line chunks. The distance of the *Additamentum* is shown as a dashed horizontal line. Chunks that overlap the *Additamentum* are drawn crossed.



Parola d'ordine: risparmio



Come visto a lezione un aspetto primario di questo tipo di ricerche é il consumo computazionale per cui si adotta spesso un approccio economico a costo di una fase dispendiosa di programmazione.

Un tool di questo tipo è: “mqdq-parser”. Il MQDQ è un corpus di poemi, memorizzati in XML che permette di associare ad ogni elemento vari tag, di facile accesso, il che semplifica l'estrazione e la catalogazione dei testi.

Il codice è open-source e disponibile su github.



bnagy / mqdq-parser

Public

MQDQ-parser

Il codice è scritto in Python3, un linguaggio ad alto livello, interpretato, di facile composizione e comprensione a discapito di un più lungo tempo di elaborazione e un maggiore utilizzo delle risorse, che però per i requisiti del progetto risulta trascurabile.

É interessante evidenziare l'utilizzo di alcune librerie che ricorrono comunemente in diversi codici analizzati volti a scopi simili. Queste sono:

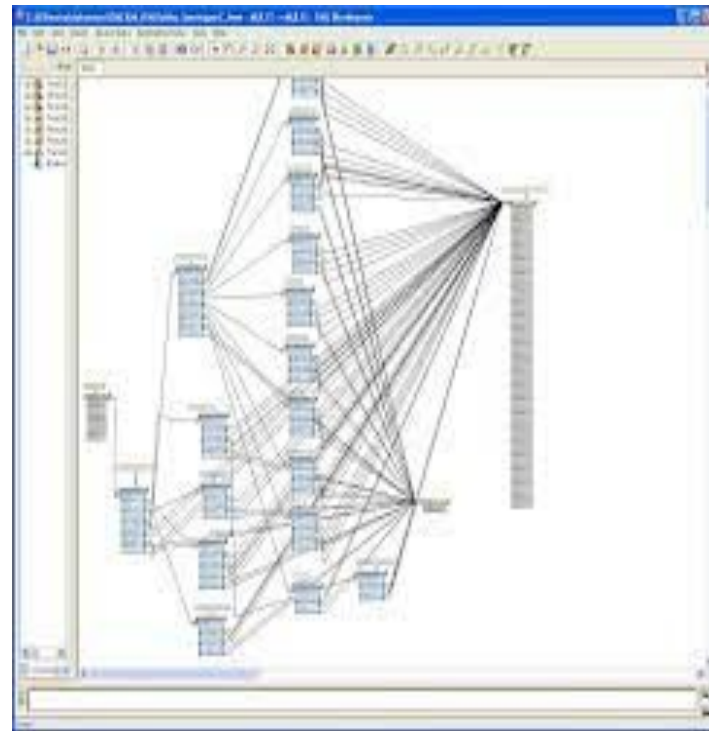
1. beautifulsoup4,
2. numpy,
3. pandas,
4. scipy,
5. seaborn.

```
18     install_requires=[
19         "setuptools",
20         "beautifulsoup4",
21         "numpy",
22         "scipy",
23         "pandas",
24         "seaborn",
25         "dominate",
26     ],
27 )
```

Chiron

Un altro esempio, sicuramente più avanzato, è “Chiron”, che utilizza il concetto di chain of layers aumentando precisione e scalabilità con un significativo risparmio computazionale.

Il concetto si basa sulle somiglianze fra linguaggi e grammatiche, per cui è non è raro che nell’analisi di un testo greco il programma passi attraverso layers del sistema latino ecc...



Cosa abbiamo notato e come potrebbe essere applicata l'AI.



Data la natura del problema il codice presentato, come mostrano anche i notebook, è sicuramente funzionale, ma comunque produce talvolta errori.

Ciò che l'allenamento di un algoritmo di ML supervisionato, che a nostro avviso ben si presta per lo scopo, migliorerebbe è:

1. una personalizzazione del modello per ratio di composizione
2. una maggiore scalabilità ad altre forme metriche
3. un'eventuale evoluzione per la verifica e/o generazione (in un'istanza successiva) di una generazione "à la manière de".

BIBLIOGRAFIA

A Multilanguage, Modular Framework for Metrical Analysis: It Patterns and Theorical Issues, Daniele Fusi, 2015

Adam Pawlowski & Maciej Eder (2001) Quantity or Stress? Sequential Analysis of Latin Prosody, Journal of Quantitative Linguistics, 8:1, 81-97, DOI: 10.1076 jqul.8.1.81.4093

Propedeutica al latino universitario, A. Traina, 1998.

Carmen et Standard Error: Computational Methods in Stylometry for Classical Latin Poetry, Benjamin C. Nagy, 2021

TIME SERIES MODELLING IN THE ANALYSIS OF HOMERIC VERSE, ADAM PAWŁOWSKI, MAREK KRAJEWSKI, MACIEJ EDER

www.Pedecerto.eu