

# The Charition (Χαρίτιον) Mime

Decoding the “Indian Language” through Typology and Entropy

Freja Lindgren, Tiago Tresoldi



UPPSALA  
UNIVERSITET

26<sup>th</sup> International Conference on Historical Linguistics (ICHL26), Heidelberg, 06.09.2023

413

[illegible]

1. *Създаване на нов документ*  
 2. *Отваряне на съществуващ документ*  
 3. *Записване на документ*  
 4. *Изтриване на документ*  
 5. *Копиране на документ*  
 6. *Печатане на документ*  
 7. *Промяна на име на документ*  
 8. *Промяна на тип на документ*  
 9. *Промяна на статус на документ*  
 10. *Промяна на дата на документ*  
 11. *Промяна на място на документ*  
 12. *Промяна на автор на документ*  
 13. *Промяна на версия на документ*  
 14. *Промяна на описание на документ*  
 15. *Промяна на метаданни на документ*  
 16. *Промяна на настройки на документ*  
 17. *Промяна на права на документ*  
 18. *Промяна на история на документ*  
 19. *Промяна на статистика на документ*  
 20. *Промяна на отчети на документ*

[illegible]

# Introduction

- A Greek theatre play discovered in Papyrus Oxyrhynchus 413
- Dated to the 2nd century CE
- Combines elements from the *Odyssey*, Aeschylus' *Suppliants*, and, especially, Euripides' *Iphigenia in Tauris*
- Largely comedic or burlesque, with musical instructions indicating a rich musical backdrop



# Content

- Plot: a Greek party arrives in India to rescue Chariton, a Greek girl held captive, and they manage by intoxicating the locals
- Characters in the play converse in an unknown language
- The language might be
  - gibberish
  - gibberish with some actual linguistic material
  - an amalgamation of (Indian) languages, maybe even a creole
  - an unknown language, approximately transcribed in Greek



# Previous work

- First transcription, translation and notes in Grenfell and Hunt (1903)
  - “The language is no doubt to a large extent of an imaginary nature, but it may include some genuine non-Hellenic elements” (p 43)
- Hultzsch (1904), Sastri (1926) & B. A. Saletore (1936) identify the language as Kannada (Dravidian) based on locale and language
  - They also provided attempted translations
- Shivaprasad Rai (1985) and U. Padmanabha Upadhyaya (1996) posit it as Tulu (Dravidian)
  - They also provided attempted translations





# Issues

- The translations have been doubted based on several reasons:
  - Speculative (Barnett 1926, Rice 1929)
  - Not following a scientific method (Cunningham apud Tsitsiridis 2005)
  - Translating items to contemporary Kannada & Tulu (Barnett 1926, Rice 1929)
- Issues identified by us also includes:
  - inconsistency of translations (same expression, different translations)
  - Translating expressions into complete sentences when it is clearly fragmented
  - indications that are based exclusively on the English translations and little understanding of the literature (e.g., usage of Sotadean metre)
  - Not using regular sound correspondences
  - Reasons to believe we have a musician's copy (and not an entire fragment)

# However!

- We believe they are right about natural language!
  - This presentation is about the work-in-progress towards such claim
- The language looks natural at first glance
- It is very different from the “barbarian” and “gibberish” languages in similar sources (e.g. Aristophanes)
  - It is not “broken Greek”
  - It is not onomatopoeic, repetitive, of lower phonological complex
  - The play does not really work if the audience does not understand the languages, particularly due to the extent of the “barbarian” lines
  - There seems to be a correction of the unknown language
  - Utterances seem to show inflection
  - There are other examples of foreign language in Ancient drama, such as in Plauto’s *Poenulus* (thanks to Marwan Kilani for pointing this out)



.....  
ΒΑΣ. βραθις.  
ΚΟΙ. βραθεις.

It's watery, pour in wine. *Many drums*

F.

Skalmakatabapteiragoumi.

70

Z.

Tougoummi *Drums* Nekelekethro. F. Eitoubelletra  
choupteragoumi.

B.

Ah! *Cymbals* No unpleasantness!

stop! *Drums Cymbals* Ah! *Cymbals* What are you doing?

Z.

Trachountermana.

F.

Boullitikaloumbaï platagoulde. *Cymbals* Bi[

B.

Apuleukasar. *Drums* KING. Chorbonorbothorba[  
toulmionaxizdeSPI. platagoulde. *Cymbals* Bi[  
seosarachis. *Drums* king. [...],orado *Cymbals* satur[

75

KING.

Ouamesaresumpsaradara *Cymbals* ei *Cymbals* ia *Cymbals*  
da[

B.

Martha *Cymbals* marithouma edmaïmaï *Cymbals*  
maïtho[ thamouna martha *Cymbals* marithouma *Drums*  
[.....]tun[

80

KING.

Malpiniakouroukougoubi *Cymbals* karako [...],ra.[

ALL

Aba. KING Zabede *Cymbals* zabiligidoumba.

ALL

Aba



# Contact between southern India and roman Egypt

- Most roman coins found south in India (Hall 2010)
- Ship every three days to India (Strabo 1st century CE)
- Contacts between Egypt and India during the Ptolemaic and Roman Periods (Kilani 2017), including the recently discovered Buddha of Berenike



# Typological review - I

- Using our phonological transcription, as discussed later
- Phoneme inventory
  - Consonants: Includes both aspirated /k<sup>h</sup>/, /p<sup>h</sup>/ and unaspirated (k, p) forms
  - Vowels and diphthongs
  - Long vowels /o:/ and /e:/
- Syllable structure
  - Both open and closed syllables present
  - Varieties like CV, CVC, CCV are prevalent
- Phonotactic rules
  - Initials include aspirated and unaspirated plosives, hinting at phonemic distinction
  - Various consonant clusters
  - Aspirated phonemes commonly at the start of words
  - Specific restrictions on phoneme "β"

# Typological review - II

- Recurring Patterns
  - Sequences like "a β a", "k o t: o: s" may represent significant words or morphemes.
  - "t<sup>h</sup> ei" and "t<sup>h</sup> i" could denote specific morphological components.
- Morphological Indicators
  - Variants like "β r a t<sup>h</sup> i s" vs. "β r a t<sup>h</sup> ei s" hint at inflection and different forms of root words.
  - Sequences like "m ai m ai" showcase reduplication, potentially indicating plurals, intensifiers, etc.
  - Repetitive structures like "k<sup>h</sup> o r β o n o r β o t<sup>h</sup> o r β a" may represent a song, chant, or ritualistic expression.

# Data Analysis Method - I

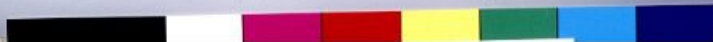
- ***Data Collection:*** Obtained transcriptions in the Greek alphabet.
- ***Phonological Transcriptions:*** Constructed phonological transcriptions of the unknown language based on expected phonetic values from a play from Oxyrinchus during that era.
- ***Translation Creation:*** Developed potential English translations for the unknown segments using context, Greek literature insights, and intuition.
- ***Multilingual Translations:*** Procured translations of these potential translations in various languages through manual and automated tools (e.g., Google Translate, ChatGPT, llama.cpp).

413

The image shows a single page from the Voynich manuscript, characterized by its unique script. The parchment is heavily aged and discolored. A significant portion of the left edge is missing, creating a large, jagged tear that exposes the inner binding of the book. The text is arranged in several horizontal lines, with some lines appearing more densely packed than others. The script itself is a series of loops, curves, and straight lines, which are not recognizable as any known human language. The overall appearance is one of great age and historical mystery.

[illegible]

This image shows a fragment of an ancient papyrus scroll. The text is written in a cursive script, likely from the Hellenistic or Roman period, but is largely illegible due to the fragment's narrow width and significant damage. The papyrus is heavily discolored, showing a range of brown and tan hues, and has large, irregular sections missing, particularly along the right edge and in the center. The remaining text is arranged in several lines, with some characters appearing to be Greek or Latin letters. The fragment is set against a plain, light-colored background.





## Popular Mime

LCL 225: 378-379

GO TO SECTION ▼

Go to page:  [GO](#)

[Find in a Library](#) [View cloth edition](#)

### POPULAR MIME

] KOI. αβορατον Z  
 ]μαλαγαβρουδιτακοτα  
 ]ρασαβαδιναραπρουτινα  
 ],. [ .... ]ακρατιεντιγα  
 15 ]μα  
 ]ωσαδω[, ]χαριμμα Z  
 B. ό πρω]κτός μου άπεσφήνω-  
 ται έν τ]ώι πελάγει χει-  
 μών ]αι έρεγμὸν  
 20 κ]ατείδαν αὐτῶν

### FRAGMENT 6

]ALL. aboraton Z  
 ]malalagabroudittakota  
 ]rasabadinaraproulinna  
 ]akratieutiga  
 15 ]ma  
 ]osado[, ]charimma Z  
 B.  
 ] my bum is fully compressed  
 ] a storm at sea  
 ] bellowing

Tools

Search within work



κροτ.  
 αι) B.

] they've seen their

ALL





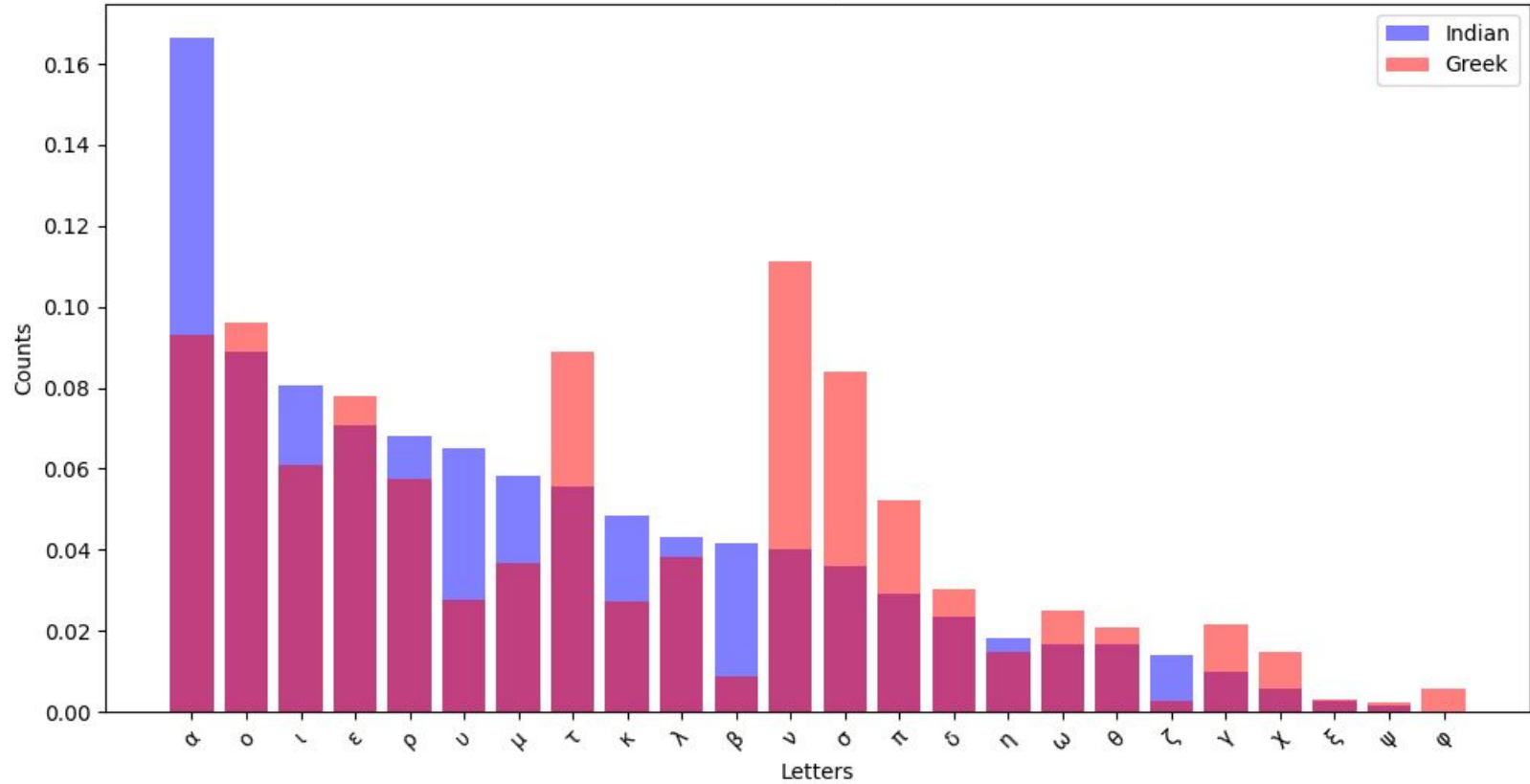
A9 μινει

	A
1	CHARITION
2	κρανου
3	λαλλε
4	λαπαλιαντα λαλλεαβ...αιγμ...
5	κοτακως αναβ...ωσαρα
6	λασπαθια
7	αλεμακα
8	αλεμακα...
9	μινει
10	βραθις
11	βραθεις
12	στουκεπαιρομελλοκοροκη
13	.ραθιε
14	βερη κονζει δαμυν πετρεκιω πακτει κορταμες βερη ιαλερω δεπωμενζι ετρεκιωδαμυτ κινζη παξει ζεβης λολω βια βραδισ κοττως
15	κοττως
16	ζοπιπ
17	ζεισουκορμωσηδε
18	σκαλμακαταβαπτειραγουμι

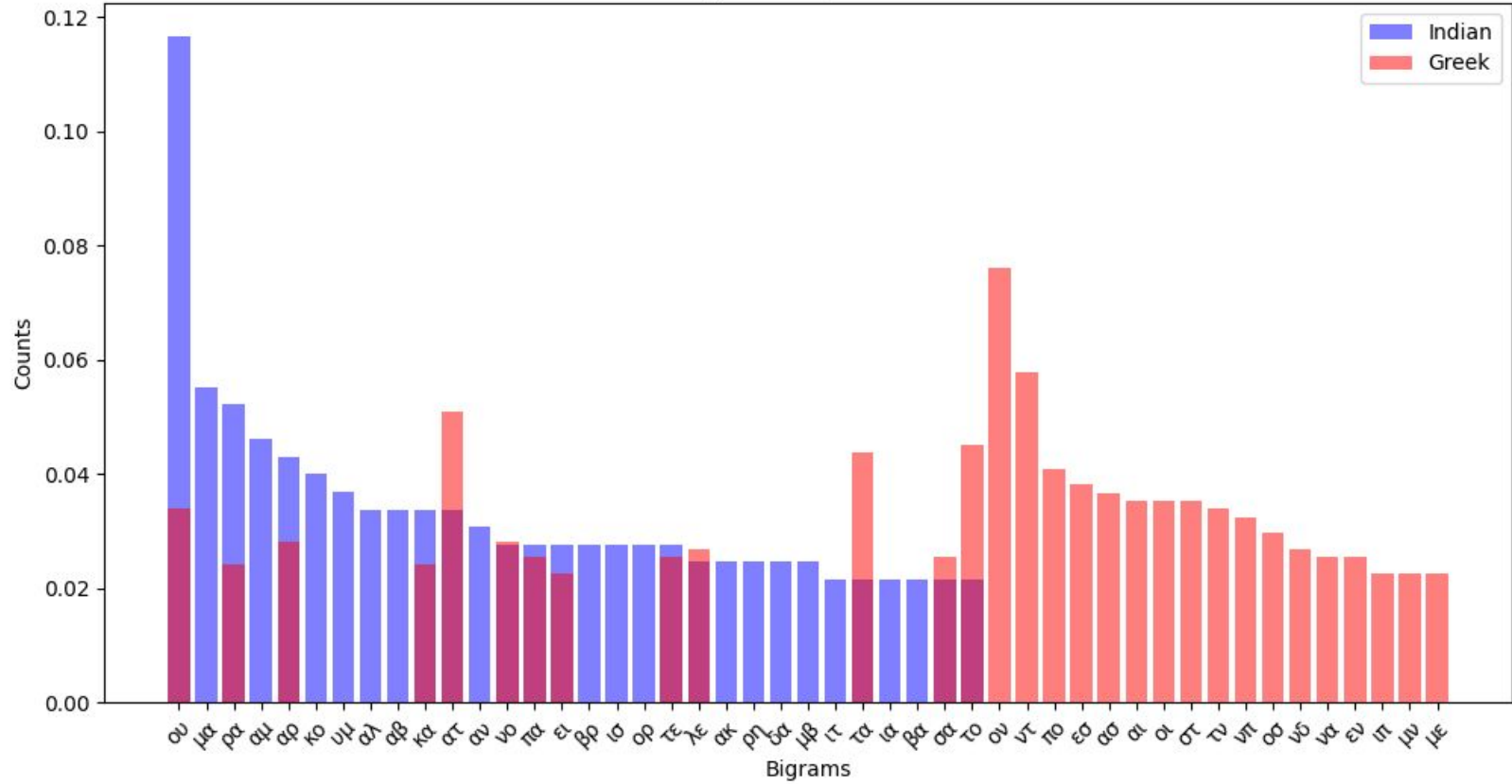
charition - Copy

Find Find All Formatted Display Match Case

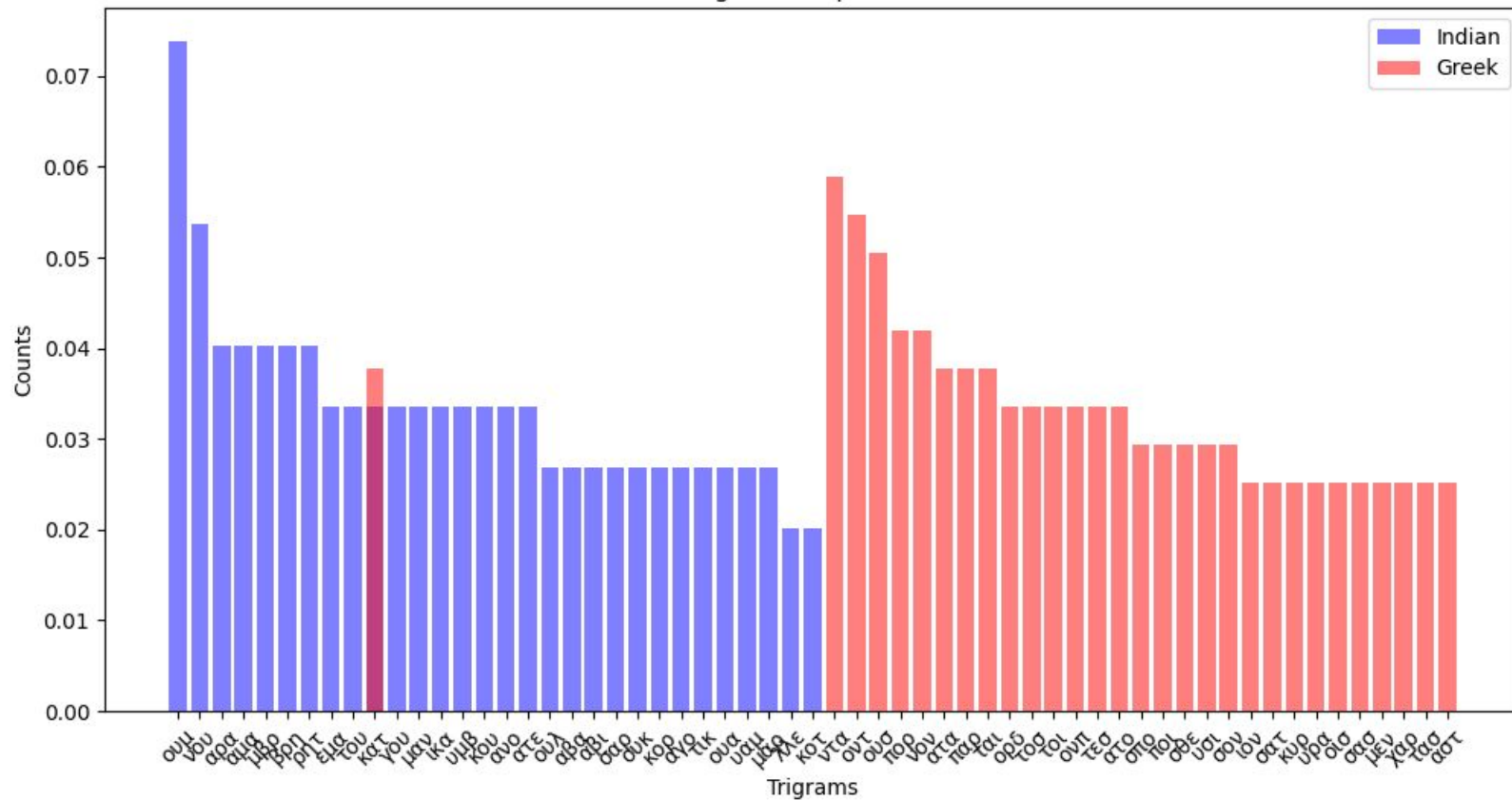
# Letter Frequencies



# Bigram Frequencies

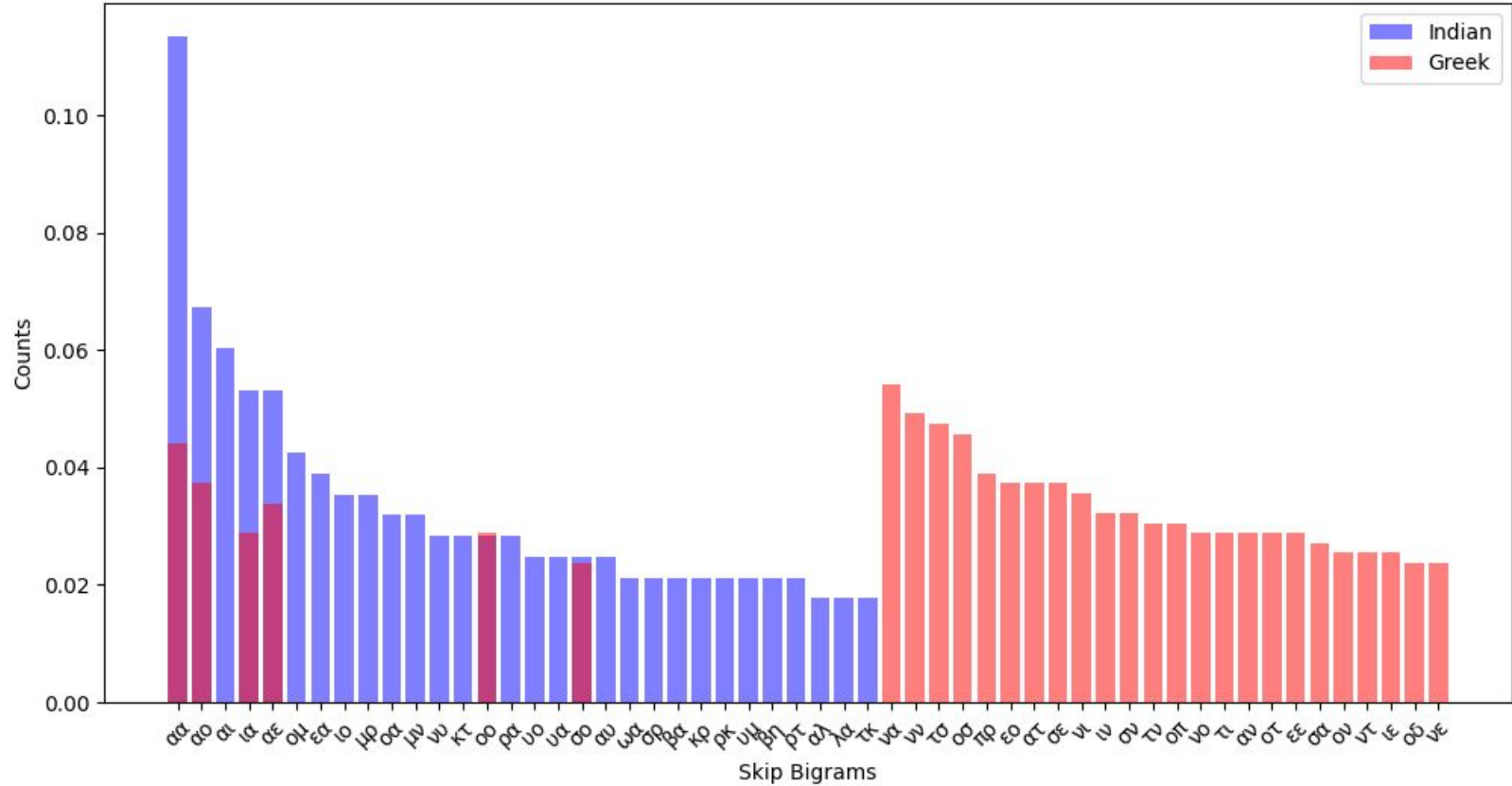


## Trigram Frequencies





Skip Bigram Frequencies



CHARITION	English
ζοπιτ	Magnificent!
ζεισουκορμοσηδε	You, take a sip!
σκαλμακαταβαπτειραγουμι	Savor this blessed beverage, don't hold back!
τουγουμι. νεκελεκεθρω.	Raise your cup, be one of us!
ιτουβελλετρα χουπτεραγουμι	Savor the drink's pure delight!
τραχουντερμανα	Let it dance in your mouth!

📄 Texto

🖼️ Imagens

📄 Documentos

🌐 Sites

Detectar idioma Inglês Português Espanhol ▼

Raise your cup, be one of us!



[Consulte os detalhes](#)



29 / 5.000



↔️ Tailandês Suaíli Espanhol ▼

ยกถ้วยของคุณมาเป็นหนึ่งในพวกเรา!



Yk thŭy khŭng khun mā pĕn hŭnŭng nŭ phwk reā!

[Consulte os detalhes](#)



Enviar feedback



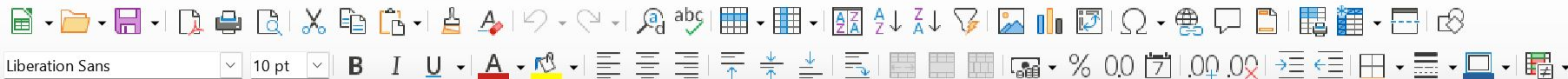
Histórico



Salvas



Contribuir



Liberation Sans

10 pt

B

I

U

A

A1 CHARITION


	A	
1	CHARITION	Afrikaans
2	κραυγου	Stop!
3	λαλλη	Kyk daar!
4	λαπαλιαντα λαλλεαβ...αιγμ...	Kyk na hulle, is hulle bekend?
5	κοτακως αναβ...ιωσαρα	Pasop, een van hulle kom.
6	λασπαθια	Vang hom!
7	αλεμακα	Moenie hom seermaak nie!
8	αλεμακα...	Sy beskerm hom.
9	μινει	Kom ons bad
10	βραθις	Sal ons baie gooi?
11	βραθεις	Ons sal gooi!
12	στουκεπαιρομελλοκοροκη	Vul die bekere, laat die viering begin!
13	.ραθιε	Aan die wyn!
14	βερη κονζει δαμυν πετρεκιω πακτηι κορταμες βερη ιαλερω δεπωμενζι ετρεκιωδαμυτ κινζη παξει ζεβης λολω βια βραδισ κοττως	Hierdie bekende drankie van ver af is nou voor or
15	κοττως	Cheers!
16	ζοπιτ	Magnificent!
17	ζεισουκορμοσηδε	Jy, neem 'n slukkie!
18	σκαλμακαταβαππειραγουμι	Geniet hierdie geseënde drank, moenie terughou
19	τουγουμι. νεκελεκεθρω.	Verhoog jou beker, wees een van ons!
20	πουβελλετα κοιπτεροουμι	Geniet die drank so suwer niosier!

charition - Copy

Find Find All Formatted Display Match Case

The Thai sentence you've provided is:

ยกถ้วยของคุณมาเป็นหนึ่งในพวกเรา!

 Copy code

Here's an approximate transcription:

csharp

[jók tʰûaj kʰǎːŋ kʰun māː pen nûŋ nāj pʰûak rɯː]

 Copy code

However, this is a rough transcription. Some things to note:

1. The transcription may vary depending on the actual intended pronunciation, especially regard to vowel length, tones, and exact vowel quality.

 Regenerate





# llama.cpp

The logo features the text "LLaMA" in a large, white, sans-serif font. To its right is a stylized orange "C" with two small orange flames above it, followed by two orange plus signs "++". The entire logo is set against a dark gray rectangular background.

 CI  passing  license  MIT

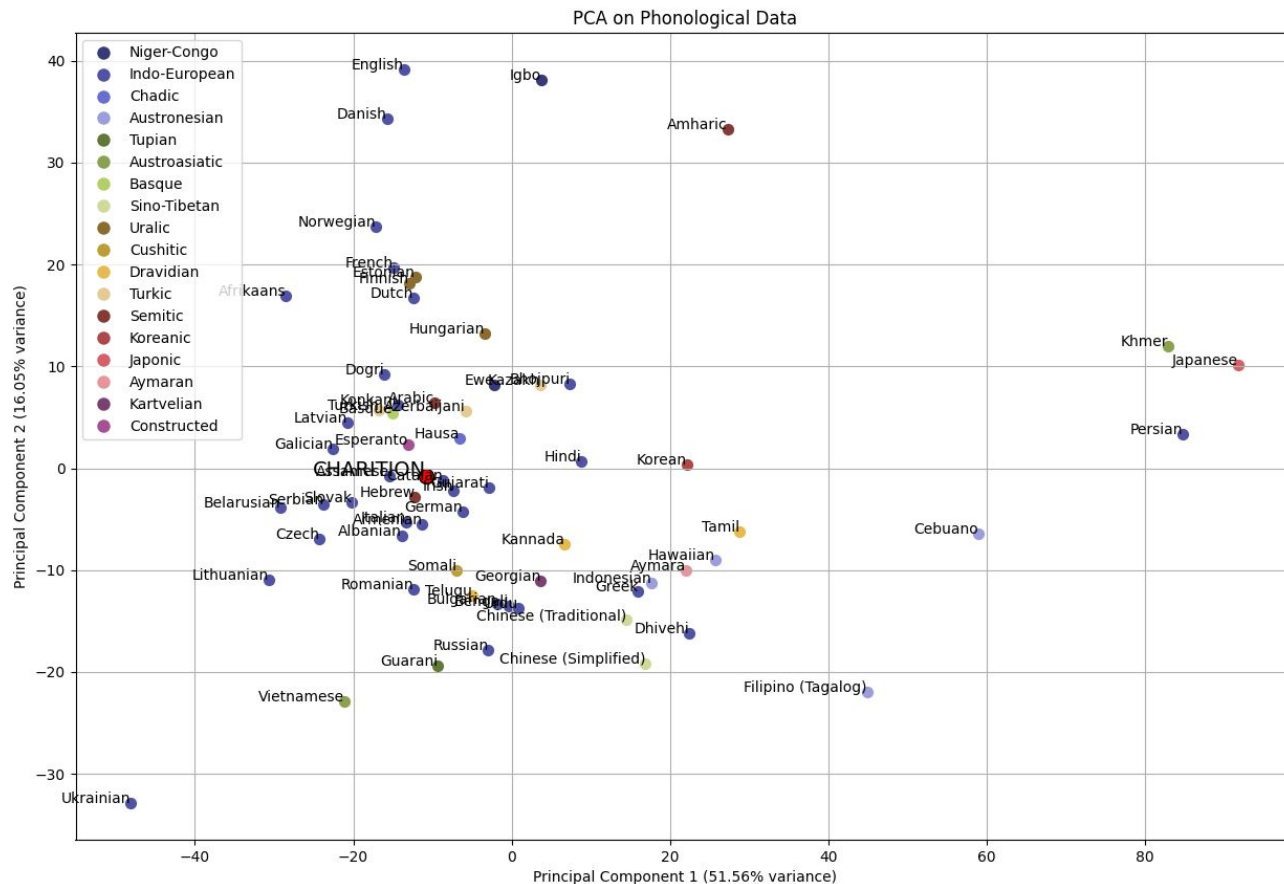
[Roadmap](#) / [Manifesto](#) / [ggml](#)

Inference of [LLaMA](#) model in pure C/C++

History

# Data Analysis Method - II

- ***Phonological Transcriptions:*** Utilized automated tools and conducted reviews for quality assurance.
- ***Tokenization:*** Tokenized transcriptions using a custom orthographic profile method, leveraging insights from CLTS and Lexibank for normalization.
- ***Vectors of Information:*** Extracted vectors from tokenized phonological sequences that encompass sound classes, distinctive features, and more.
- ***Additional Data Processing:*** Calculated entropy and a unique form of edit distance between all language pairs.
- ***Feature Reduction & Visualization:*** Reduced features to create a 2D plot to:
  - Evaluate if genetically and areally close languages cluster together.
  - Determine the placement of the unknown language (excluded during training)



# Limitations

- While the language feels natural, no immediate connection can be made to any other language or family
- Transcription of the Greek (Koine? Atticized Koine?)
- Comparison to modern languages only
- Needs better and more data, as well as “random” language as a control
- “In the present state of our knowledge there can never be certainty [about the language], but the question should surely be approached with an open mind” (Hall 2010)

# Conclusion

- ***Preliminary Insights:*** Work is ongoing, but early signs support the hypothesis that the language is *not* mere gibberish.
  - It is interesting that the language fits between Indo-European and Dravidian, and particularly that it is close to Esperanto
- ***Data release and community building***
- ***Automation's Impact:*** Tools like Google Translate and llama.cpp sped up tasks that would have taken years and many experts
  - This could potentially be used in many other linguistic tasks
- ***Broader Applications:*** By combining expert knowledge, automation, and statistics, this approach could help the research on other undeciphered texts