

# Yes but.. Can ChatGPT Identify Entities in Historical Documents?

**Carlos-Emiliano González-Gallardo**, Emanuela Boros, Nancy Girdhar,  
Ahmed Hamdi, José G. Moreno & Antoine Doucet

Université de la Rochelle, France



JCDL '23 - 28 June 2023

# Historical document datasets



ΑΘΗΝΑ.

'ΑΕ] μίν, ὦ παῖ Ἀλφρίου, δέδορκα σε  
πεῖραν τιν' ἐχθρῶν ἀρπάσαι θηράμενον  
καὶ νῦν ἐπὶ σκηναῖς σε ναυτικάς ὁρῶ  
Ἀλάντος, ἐνθα τάξιν ἐσχάτην ἔχει,  
πάλαι κυνηγετούδα καὶ μετρούμενον  
ἔχην τὰ κείνου νεοχάραχ', ὅπως ἰδῆς  
εἰτ' ἐνδόν εἰτ' οὐκ ἐνδόν, εἰ δέ σ' ἐκφέρει

[1 a.

5

1. Ἀλφρίου ΛΑ. Ἀλφρίου C' Vat. ac. Ἀλφρίου L'. 6. τὰ ἑσόνος L. τάξιον  
Παλ. νεοχάραχ' νεοχάραχ' L. ac. νεοχάραχ' C' L' Vat. ac.

1-3. Athena's eye is ever on Odysseus, and she is now come from Olympus to succour him. Infr. l. 36.  
4. ἴδω. ἔχει] The structure is paratactic; i.e. 'As I have ever seen thee... so now I see thee.' Essay on Language, 4, 36 p. 68.  
2. (1) 'In quest to snatch some exploit on a foe', i.e. seeking to effect some surprise against a foe. Or, (2) 'Seeking to foil (or detect) some enemy's attempt'. The latter (2) is simpler, and *νεῖρα* is used of the attempt of Ajax, infr. 290, 1027; but the former (1) is on the whole more probable. For Athena does not profess to know the circumstances until l. 36. She asks for information, and only assumes, what is evident, that Odysseus is engaged in some hostile adventure. This aspect of his character appears in the tenth Iliad. Cp. infr. 18, *ἔργον ἐφ' οὗ ἐν' ἀνδρὶ δορυμνεί* | *βίαν κυλοῦν*.  
*δορυμνεί* is to seize, i.e. 'to effect suddenly'. *θηράμενον* introduces the image of the huntsman continued in l. 5, and combined with that of the hound in ll. 7, 8.  
*ἀρπάσαι θηράμενον* is substituted for *μετρούμενον*, so as to convey the notion of surprise; *δορυμνεί* is an epexegetic infinitive, after which the accusative *εἶπεν* is to be resumed. The meaning

of *ἀρπάσαι* in (2) sup. viz. 'to arrest' is less natural than that given in (1).  
3. *σκηναῖς*] The *skeneia* of the Homeric hero. Cp. infr. 192-3.  
4. *ἴδω. ἔχει*] ll. 11, 7, 8; Eur. I. A. 292. This position of Ajax' tent enables him the more easily to steal forth unobserved at last, infr. 690 ff.  
5. *κυνηγετούδα*, which has no object, resumes *θηράμενον*.  
*μετρούμενον* 'Scanning attentively'. The middle voice marks the mental nature of the act; not measuring with a line, but scanning with the eye.  
6. *νεοχάραχ'*] Ajax has but recently returned, dragging the cattle with him, infr. 296. Odysseus has tracked him so far, but the confused struggle at the tent-door has made it uncertain whether he is not gone forth again.  
7, 8. *εἰ δέ σ' ἐκφέρει, κτλ.*] Odysseus is like a huntsman who is led to the right point by the scent of a keen Spartan hound. The dog is introduced to complete the image. Cp. infr. 19, 32. Is *εἶπεν* (1) nom. or (2) gen.? *εἶπεν* occurs in Aesch. Ag. 1093, and the authorities for *εἶπεν* are late. But the sentence is more balanced if the epithet is taken by hypallage with *βίαν*, and the abstract noun is some what abrupt by itself. Cp. the forms *εἶργον, εἶρμ' εἶργον, εἶρμ'.*

	Tokens	Entities	PERS	LOC	ORG	HumanProd/ object	TIME/ date	SCOPE
NewsEye [19C-20C]	30,458	1,298	463	597	217	21	—	—
*hipe-2020 [19C-20C]	48,854	1,600	502	854	130	61	53	—
ajmc [19C]	5,390	360	139	9	—	80	3	129

pers work  
pers myth work.primlit scope

Others would render ἰσχυροῦς, mightysouldered," comparing the description of Ajax in ll. 3. 22 as ἑξοχος Ἀργείων κεφαλὴν ἢ δ' εὐρέας δυοῦν.

# Methodology (prompting ChatGPT)

NewsEye

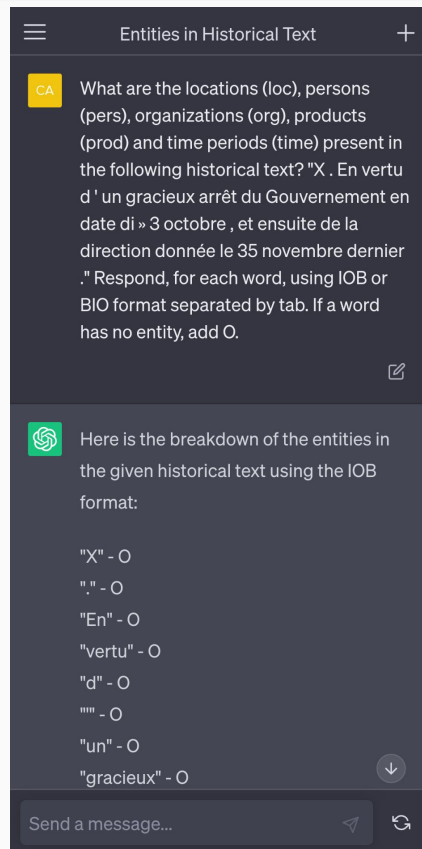
What are the locations (LOC), persons (PER), organizations (ORG) and human productions (HumanProd) present in the following historical text? **{PHRASE}**

hipe-2020

What are the locations (loc), persons (pers), organizations (org), products (prod) and time periods (time) present in the following historical text? **{PHRASE}**

ajmc

What are the locations (loc), persons (pers), time periods (date), human works (work), physical objects (object) and specific portion of works (scope) present in the following historical text? **{PHRASE}**



11th January -  
7th February  
2023

\*Respond, for each word, using IOB or BIO format separated by tab. If a word has no entity, add O.

# Findings

# Named Entity Definition

Reference:

*Le sens est donc : « Ai-je parlé trop peu clairement ? » Cf. Antigone, 405 : “Ap’ Eväênda xxi coepñ XéYw ; Eschyle, Agamemnon, 269 : ”H χορῶς λέγω ; [ajmc]*

HumanProd  
SCOPE

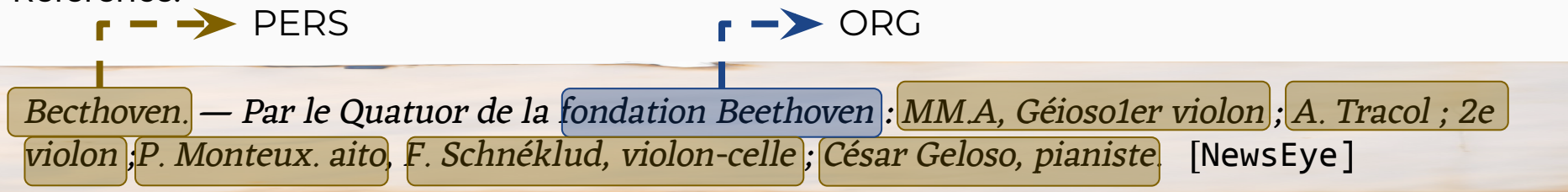
PERS

Response:

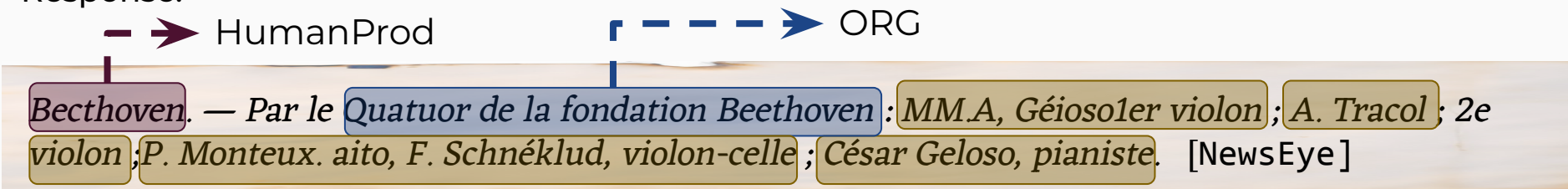
*Le sens est donc : « Ai-je parlé trop peu clairement ? » Cf. Antigone, 405 : “Ap’ Eväênda xxi coepñ XéYw ; Eschyle, Agamemnon, 269 : ”H χορῶς λέγω ; [ajmc]*

# Entity Complexity

Reference:



Response:



# Digitization Errors

Reference:

— ➔ PERS

« 13659 - 4360.Hxépra. ... . elloug. Ulysse paraît faire allusion à l'amertume des peroles que vient de prononcer À gamemnon ; Agamemnon répond comme si Ulysse avait eu en vue l'amertume de ses propres remontrances, 4866. 'EvOaë' Tlouet, j'en arriverai là, c'est-à-dire, je mourrai, .Dindorf : Kai' αὐτὸς ἴξομαι πρὸς τὸ θάπτειν αὐτόν. » [ajmc]

Response:

date

object

— ➔ PERS

— ➔ SCOPE

« 13659 - 4360.Hxépra. ... . elloug. Ulysse paraît faire allusion à l'amertume des peroles que vient de prononcer À gamemnon ; Agamemnon répond comme si Ulysse avait eu en vue l'amertume de ses propres remontrances, 4866. 'EvOaë' Tlouet, j'en arriverai là, c'est-à-dire, je mourrai, .Dindorf : Kai' αὐτὸς ἴξομαι πρὸς τὸ θάπτειν αὐτόν. » [ajmc]

# Code-switching



4890 . Μνήμων τ' Ἐρινύς . Οἱ *Eschyle* , *Prométhée* , *δι6* : Μοῖραι τρίμορφοι μνύ,μονές τ' Ἐρινύες ( εἰ εεῖ - ἁ - αῖγα suivant le Scholiaste : Αἱ μνημονεύουσαι τῶν παρ' ἀνθρώκοις καχῶν καὶ ἀντιδιδούσαι ) . [ajmc]

Response:



I'm sorry, but I'm unable to understand the text you have provided. The text appears to be a mixture of ancient Greek and French, with some references to ancient literature and annotations, which makes it difficult to extract meaningful information. Additionally, the text is not a historical text, but rather a literary text, which also makes it difficult to extract historical entities.





# Wrapping up...



Yes but.. Can ChatGPT Identify Entities in Historical Documents?



As an AI language model, I can certainly help you identify entities in historical documents to the best of my abilities. While I don't have direct access to real-time information beyond my knowledge cutoff in September 2021, I can still provide general guidance based on historical data and commonly known entities.



If you provide me with the historical document or relevant text, I can assist you in identifying and extracting entities such as names of people, places, organizations, dates, and other relevant information. However, please note that my responses are based on patterns and information available up until September 2021, so I might not be aware of specific events or developments that have occurred after that time.

Feel free to share the historical document or provide specific information from it, and I'll do my best to assist you in identifying the entities within it.

Thank you !!  
Gracias !!  
Merci !!

[https://github.com/cic4k/NERC\\_ChatGPT](https://github.com/cic4k/NERC_ChatGPT)

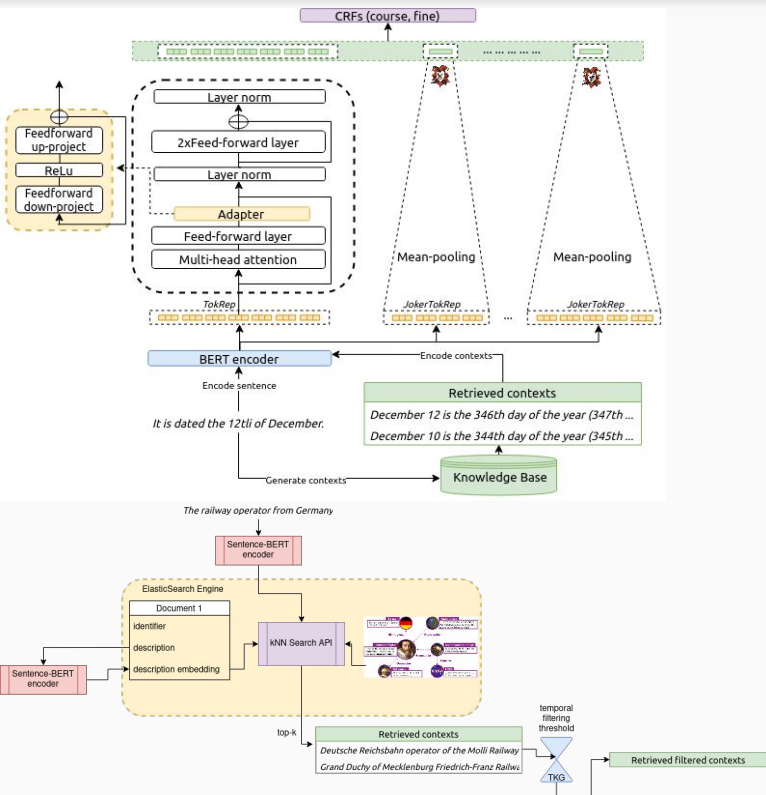
Carlos-Emiliano  
González-Gallardo

[carlos.gonzalez\\_gallardo@univ-lr.fr](mailto:carlos.gonzalez_gallardo@univ-lr.fr)



*"A robotic agent reading an ancient newspaper renaissance painting style 84137"*

# Some results...



**Table 3: Comparative results using the three datasets (micro).**

	NewsEye			hipe-2020			ajmc		
	P	R	F1	P	R	F1	P	R	F1
strict									
<i>Stacked NERC</i>	<b>75.0</b>	70.6	<b>72.7</b>	–	–	–	–	–	–
<i>Temporal NERC</i>	–	–	–	<b>76.5</b>	<b>76.5</b>	<b>76.5</b>	<b>84.8</b>	<b>83.9</b>	<b>84.4</b>
ChatGPT	70.9	<b>72.3</b>	71.6	32.5	50.0	39.4	21.8	26.1	23.8
fuzzy									
<i>Stacked NERC</i>	<b>85.4</b>	<b>80.5</b>	<b>82.9</b>	–	–	–	–	–	–
<i>Temporal NERC</i>	–	–	–	<b>86.7</b>	<b>86.7</b>	<b>86.7</b>	<b>90.2</b>	<b>89.2</b>	<b>89.7</b>
ChatGPT	77.8	79.4	78.6	49.0	75.4	59.4	25.5	30.6	27.8

*Stacked NERC*: Hamdi, Ahmed, et al. "A multilingual dataset for named entity recognition, entity linking and stance detection in historical newspapers.", SIGIR 2021.  
*Temporal NERC*: González-Gallardo, Carlos-Emiliano, et al. "Injecting Temporal-Aware Knowledge in Historical Named Entity Recognition.", ECIR 2023.