

Evaluating Large Language Models for Character Identification in Italian Renaissance Epics

A Case Study on “Orlando Furioso”

Gaudenzia Genoni¹, Alfio Ferrara^{1,2}, Stefano Montanelli¹

¹University of Milan, Department of Computer Science, Via Celoria, 18-20133 Milan, Italy

²University of Milan, Department of Literary Studies, Philology and Linguistics, Via Festa del Perdono, 7-20122 Milan, Italy



UNIVERSITÀ
DEGLI STUDI
DI MILANO



Introduction

Research Question

Can LLMs identify characters in the low-resource literary domain of Italian Renaissance epic poetry?

Challenges

- Linguistic complexity
- Segmentation into smaller textual units

Case Study

The first twelve cantos of *Orlando Furioso* (*The Frenzy of Orlando*).

Data

Annotation Process

- Proper names (PN)
- Nouns (N) / Noun Phrases (NP)
- Pronouns (P)

Constraints

- Narratological level
- Individual identity
- Proper-name attribution
- Entity type

Resulting Dataset

- 1,890 character occurrences
- 64 unique characters
- 106 stanzas with no character mentions

Annotation Example: Canto I, Stanza 5

Era *costui*_P quel *paladin*_N gagliardo,
figliuol d'*Amon*_{PN}^{NP}, signor di *Montalbano*_{NP},
a *cui*_P pur dianzi il suo destrier Baiardo
per strano caso uscito era di mano.
Come alla *donna*_N *egli*_P drizzò lo sguardo,
riconobbe, quantunque di lontano,
l'angelico sembiante e quel bel volto
ch'all'amorose reti *il*_P tenea involto.

English translation

This valiant *knight*_N was *Lord of Clarimount*_{NP},
*Duke Ammons*_{NP} *sonne*_{NP}, as you shall understand,
*Who*_P having lost his horse of good account,
That by mishap was slipt out of his hand,
*He*_P followd him, in hope againe to mount,
Untill this *Ladies*_N sight did make *him*_P stand,
*Whose*_P face and shape proportiōnd were so well,
They seeme the house where love itselfe did dwell.

Ground-Truth Entries

AMONE, *ANGELICA*, *RINALDO*

Methodology

The poem is split into batches of n stanzas and processed by a **character identification pipeline**.

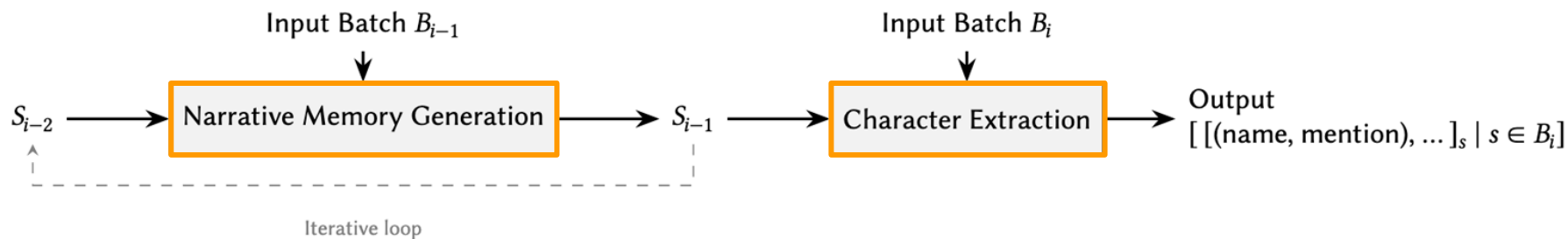


Figure 1. Character identification pipeline.

Experimental Results

Two NER baselines and four LLMs are evaluated; LLMs are tested at $n \in \{4, 8, 12\}$, with/without narrative memory.

F1 score
0.758

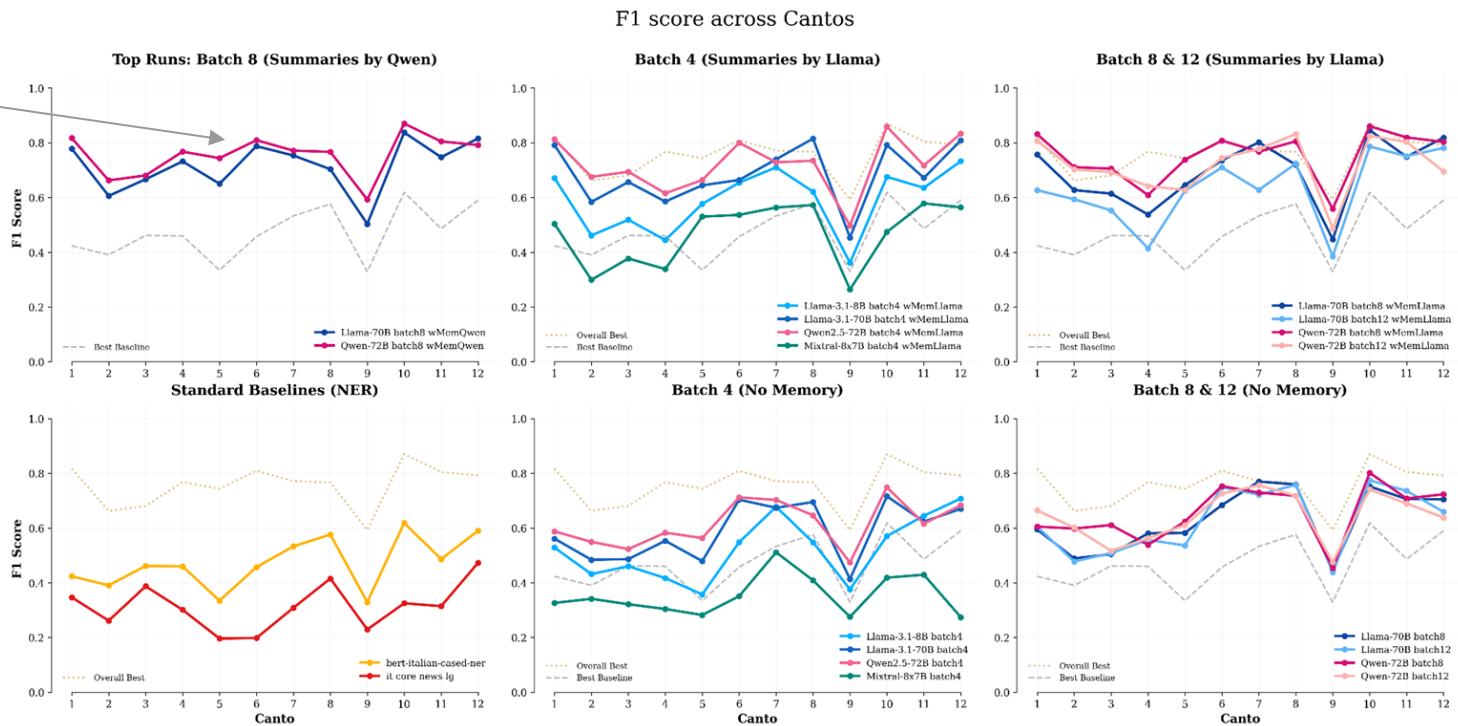


Figure 2. F1 score across Cantos I-XII for each experimental configuration.

Narrative Memory Impact

- Mean F1 improvement of +8.1 points.
- Improves **intra-batch** character identification for pre- and post-onomastic mentions.
- Enables **inter-batch** identification when a character's proper name never appears within the batch.

Mention Status	Without Memory			With Memory		
	TP	FN	Recall (%)	TP	FN	Recall (%)
➤ PN Not in batch	2	431	0.5%	118	315	27.3%
➤ Pre-onomastic	48	161	23.0%	146	63	69.9%
Onomastic	379	30	92.7%	377	32	92.2%
➤ Post-onomastic	664	173	79.3%	725	112	86.6%

Table 1. Impact of narrative memory injection for Qwen-2.5-72B ($n = 8$).

Error Analysis Insights

- ❑ Characters introduced anonymously generate **cataphoric references** whose resolution requires retrospective narrative integration not supported by the current pipeline.
- ❑ Hallucinations based on inferred character presence from narrative actions, in the absence of the required linguistic markers, point to the model's sensitivity to **character agency**.

	Error Type	Ratio
FN	Cataphoric Assignment Failure	129 / 208
	Undetected Nominal References	49 / 100
	Undetected Pronominal References	39 / 100
FP	Misresolution of Extracted Mentions	38 / 45
	Inclusion of Nameless Characters	28 / 28
	Character Misattribution	19 / 64
	Implicit Resolution	17 / 64
	Violation of Exclusion Constraints	13 / 13

Table 2. Error taxonomy for the top-ten FNs and FPs in the optimal configuration.

Future Work

- ❖ Redefining the task as character presence detection grounded in narrative agency, beyond nominal and pronominal markers.
- ❖ Tracking characters as anonymous entities linked to stable attributes, and resolving them through clustering and name resolution.
- ❖ Using structured, ontology-based narrative representations to model plot dynamics and story vs. narration order.

Main References

R. Campos, A. Jorge, A. Jatowt, S. Bhatia, M. Litvak, The 8th international workshop on narrative extraction from texts: Text2Story 2025, in: C. Hauff, C. Macdonald, D. Jannach, G. Kazai, F. M. Nardini, F. Pinelli, F. Silvestri, N. Tonello (Eds.), *Advances in Information Retrieval*, Springer Nature Switzerland, Cham, 2025, pp. 345–351.

G. Genoni, Un grafo per il “Furioso”. Prime annotazioni di una ricerca sperimentale, *AOQU (Achilles Orlando Quixote Ulysses). Rivista di epica* 3 (2022) 241–290.

J. Li, A. Sun, J. Han, C. Li, A survey on deep learning for named entity recognition, *IEEE Transactions on Knowledge and Data Engineering* 34 (2022) 50–70.

G. Martinelli, B. Gatti, R. Navigli, xCoRe: Cross-context coreference resolution, in: C. Christodoulopoulos, T. Chakraborty, C. Rose, V. Peng (Eds.), *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, Suzhou, China, 2025, pp. 34264–34278.

F. Pianzola, L. Cheng, F. Pannach, X. Yang, L. Scotti, The Golem ontology for narrative and fiction, *Humanities* 14 (2025).

P. Ranade, S. Dey, A. Joshi, T. Finin, Computational understanding of narratives: A survey, *IEEE Access* 10 (2022) 101575–101594.

B. Santana, R. Campos, E. Amorim, A. Jorge, P. A. Silvano, S. Nunes, A survey on narrative extraction from textual data, *Artificial Intelligence Review* 56 (2023) 8393–8435.

C. Santini, L. Melosi, E. Frontoni, Named entity recognition in historical Italian: The case of Giacomo Leopardi’s *Zibaldone*, 2025.

Thank you for your attention!