MILRIT at HIPE 2026: From Generalist Relation Extraction to Multi-view Learning for Historical Person–Place Relations

Ha Ngan Pham[1], Jose G. Moreno[2,*] and Antoine Doucet[1,3]

[1]Laboratory of Computer Science, Image and Interaction (L3i), La Rochelle University, La Rochelle, France
[2]IRIT, University of Toulouse, Toulouse, France
[3]University of Ljubljana, Ljubljana, Slovenia

Abstract

This paper describes the MILRIT participation in the HIPE-2026 shared task on person–place relation extraction from multilingual historical newspapers. The task requires predicting whether a person is associated with a place through the `at` relation and whether this association holds within the temporal horizon of the document through the `isAt` relation. We submitted three systems exploring two methodological directions: generalist relation extraction with GLiREL and task-specific multilingual encoder fine-tuning. The first two runs use GLiREL: one fine-tunes DeBERTa as a component of GLiREL, and the other trains an Multilayer Perceptron (MLP) with GLiREL relation scores as input. The third run uses a multilingual DeBERTa-v3-base encoder with explicit entity marking, temporal anchoring, and multi-view training based on original and enriched input representations. Official results show that the multi-view mDeBERTa-v3 system achieved our best performance, with a mean score of 0.5951, ranking 15th among submitted runs and 8th at the team level in the Accuracy Profile. It also ranked first in both the Efficiency Profile and the Balanced Efficiency Profile, showing a favourable accuracy–efficiency trade-off. The results suggest that compact, task-oriented multilingual encoders with carefully designed input representations can be robust for historical person–place relation extraction, while non-fine-tuning of generalist relation extraction models may require further adaptation to generalize effectively.

Keywords

Historical newspapers, Relation extraction, Person–place relations, Temporal relation classification, Multilingual NLP, GLiREL, DeBERTa, HIPE-2026

1. Introduction

Historical newspapers are valuable sources for studying people, places, events, and mobility over time. However, extracting person–place relations from such material is challenging because historical documents often contain OCR noise, spelling variation, multilingual content, and indirect formulations. Moreover, the co-occurrence of a person and a place in the same article does not necessarily mean that a relation holds between them.

Relation extraction (RE) is a central task in information extraction, aiming to identify semantic relations between entities in text. Recent work has shown that language-model-based approaches dominate many RE settings, while large language models have opened new possibilities for few-shot, zero-shot, and generative information extraction [1, 2]. At the same time, document-level extraction remains difficult because relations may be expressed across sentence boundaries and may require reasoning over broader contexts, entity mentions, and implicit evidence [3]. These challenges are amplified in historical corpora, where OCR errors, archaic language, and incomplete contextual cues make relation classification particularly difficult.

The HIPE-2026 shared task [4, 5] focuses on person–place relation extraction from multilingual historical texts. Given a document with pre-identified person and place entities, systems must classify all possible person–place pairs according to two relation types. The `at` relation indicates whether the

CLEF 2026 Working Notes, 21–24 September 2026, Jena, Germany
*Corresponding author.
†These authors contributed equally.
ngan.pham@etudiant.univ-lr.fr (H. N. Pham); Jose.Moreno@irit.fr (J. G. Moreno)

person was ever associated with the place before or at the time of publication, while `isAt` indicates whether the person was located there in the immediate temporal context of the document. The `at` relation is classified as TRUE, PROBABLE, or FALSE, and the `isAt` relation is classified as TRUE or FALSE, depending on the strength of the textual evidence.

This task requires both relation extraction and temporal interpretation in multilingual historical newspaper articles in German, French, and English. A person may be connected to a place through residence, travel, public office, institutional affiliation, military command, or participation in an event. Some of these relations are long-term, while others are valid only for a short period. The distinction between `at` and `isAt` therefore requires systems to determine not only whether a relation exists, but also whether it holds close to the publication date.

Our participation investigates three submitted runs. Runs 1 and 2 are based on GLiREL, a generalist model for zero-shot relation extraction that predicts relation labels between multiple entities in a single forward pass [6]. Run 1 fine-tunes a DeBERTA encoder as a component of GLiREL on the available HIPE relation extraction data, while Run 2 uses GLiREL as a relation-evidence extractor and trains a lightweight MLP classifier. Run 3 uses a multilingual DeBERTa-v3-base [7] encoder with explicit entity marking [8, 9], temporal anchoring, and multi-view training [10]. These runs allow us to compare task-specific fine-tuning, direct transfer, and a specialized multilingual classification approach for historical person–place relation extraction.

The remainder of this paper describes the submitted systems, reports their results and efficiency characteristics, and discusses the main challenges observed during our participation.

2. System Descriptions

We submitted three runs. Runs 1 and 2 are based on the GLiREL architecture. Run 1 uses GLiREL after fine-tuning its DeBERTa component on the task data, whereas Run 2 keeps GLiREL as a relation-evidence extractor and uses its relation score vectors as input to a lightweight MLP classifier. Run 3 was based on a multilingual DeBERTa-v3 model with multi-view training and temporal/contextual enrichment.

2.1. Run 1: Fine-tuned DeBERTa–GLiREL

Run 1 fine-tunes the DeBERTa encoder inside GLiREL and then converts the resulting relation scores into `at` and `isAt` predictions through score aggregation and thresholding. The *Gold Train A* and *Silver Train A* sets are combined and split into two subsets: 80% is used for model training and 20% for threshold calibration. The *Silver Dev A* set is used for validation. During training, the DeBERTa encoder parameters were updated to adapt the model to the HIPE task.

GLiREL is trained with three positive relation labels:

$$\mathcal{R} = \{\texttt{at\_TRUE}, \texttt{at\_PROBABLE}, \texttt{isAt\_TRUE}\}.$$

Negative labels are are derived later when the corresponding positive scores are below the learned thresholds. For each sampled pair $(e_i^{\text{per}}, e_i^{\text{loc}})$, a context window x_i contains the two entities and the text between them, is extracted (referred to here as the entity–text–entity context window). The base model is loaded from the pretrained GLiREL checkpoint. The maximum input length is set to 512 tokens, and the relation label set is fixed to $\mathcal{R}$. The DeBERTa encoder produces embeddings:

$$H_i = \text{DeBERTa}_\theta(x_i).$$

GLiREL then scores each relation label $r \in \mathcal{R}$ for the target person–location pair:

$$s_{i,r} = f_\theta\left(H_i, e_i^{\text{per}}, e_i^{\text{loc}}, r\right), \qquad r \in \mathcal{R}.$$

Thus, each window is represented by a GLiREL score vector:

$$\mathbf{s}_i = \left[s_{i,\texttt{at\_TRUE}}, s_{i,\texttt{at\_PROBABLE}}, s_{i,\texttt{isAt\_TRUE}}\right].$$

The model is optimised using the binary cross-entropy loss function, which is also GLiREL's internal loss function:

$$\mathcal{L}_{\text{BCE}} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{r\in\mathcal{R}}\left[y_{i,r}\log s_{i,r} + (1-y_{i,r})\log(1-s_{i,r})\right],$$

where $y_{i,r} \in \{0,1\}$ indicates whether relation r is present for the target pair in window x_i.

The scores are converted into final labels using three thresholds chosen to maximise the macro recall on the *Silver Dev A* set:

$$\tau = (\tau_{\texttt{at\_TRUE}}, \tau_{\texttt{at\_PROBABLE}}, \tau_{\texttt{isAt\_TRUE}}).$$

The final labels are assigned by thresholding the GLiREL scores.

2.2. Run 2: GLiREL-Score MLP

Unlike Run 1, which fine-tunes the DeBERTa encoder within GLiREL, Run 2 keeps GLiREL fixed and uses it as a relation-evidence extractor. The resulting GLiREL score vectors are passed to a lightweight MLP composed of linear layers with ReLU activations and dropout, with a hidden dimension of 64. The MLP uses two task-specific output heads to predict the `at` and `isAt` labels. It is trained on the combined *Gold Train A* and *Silver Train A* sets, while the *Silver Dev A* set is used for validation, as in Run 1. This system extracts sentence-level context windows instead of entity–text–entity. Let $s(e)$ denote the sentence containing entity e. If the person and location mentions occur in the same sentence, the context is:

$$x_i = s(e_i^{\text{per}}) = s(e_i^{\text{loc}}).$$

If they occur in different sentences, the two sentences are concatenated in document order:

$$x_i = s(e_i^{\text{per}}) \parallel s(e_i^{\text{loc}}).$$

Inspired by [8], the target mentions are then marked:

$$x_i = \texttt{<PER>}\ e_i^{\text{per}}\ \texttt{</PER>}\ \ldots\ \texttt{<LOC>}\ e_i^{\text{loc}}\ \texttt{</LOC>}.$$

or the reverse order if the location appears first in the document.

Given a fixed set of M relation labels, GLiREL produces a relation-score vector:

$$s_i = [s_{i1}, s_{i2}, \ldots, s_{iM}] \in \mathbb{R}^M,$$

The following candidate relation labels are used for GLiREL:

- is in
- is located in
- visited
- travelled to
- arrived in
- stayed in
- lived in
- worked in
- was born in
- died in
- is associated with
- is from

In this system, the classifier uses only the GLiREL score vector as input:

$$h_i = \text{MLP}_{\text{shared}}(s_i).$$

Two task-specific heads are then used to predict the two relations:

$$\hat{y}_i^{at} = \text{softmax}(W_{at}h_i + b_{at}),$$

$$\hat{y}_i^{isAt} = \text{softmax}(W_{isAt}h_i + b_{isAt}).$$

The `at` head performs three-class classification over TRUE, PROBABLE, and FALSE, while the `isAt` head performs binary classification over TRUE and FALSE.

The model is trained with a multi-task cross-entropy objective:

$$\mathcal{L} = \mathcal{L}_{at} + \mathcal{L}_{isAt}.$$

For one training batch $\mathcal{B}$, the losses are:

$$\mathcal{L}_{\text{at}} = \frac{1}{|\mathcal{B}|} \sum_{i \in \mathcal{B}} \log \hat{\mathbf{y}}^{\text{at}}_{i, y_i^{\text{at}}},$$
$$\mathcal{L}_{\text{isAt}} = \frac{1}{|\mathcal{B}|} \sum_{i \in \mathcal{B}} \log \hat{\mathbf{y}}^{\text{isAt}}_{i, y_i^{\text{isAt}}}.$$

This system aims to keep the system lightweight while benefiting from the GLiREL relation score. The MLP learns how patterns in the GLiREL score space correspond to the annotation scheme, without fine-tuning the entire underlying encoder.

2.3. Run 3: mDeBERTa-v3 with Multi-view Training

Run 3 uses a multilingual DeBERTa-v3-base encoder with a multi-view training strategy. This system differs from Runs 1 and 2 in both architecture and training procedure. Instead of using GLiREL, it uses a shared mDeBERTa-v3 encoder with classification heads for the two subtasks.

The input representation marks the mentions of the target person and location directly in the text. Person mentions are wrapped with section signs, while location mentions are wrapped with pilcrow signs. A date prefix is also added to provide temporal information for the `isAt` subtask as depicted in Figure 1. The special markers are chosen because they are rare in historical newspaper text and remain identifiable after tokenisation. They help the model distinguish the target pair from other entity mentions in the same document.

Inspired by [10], the mDeBERTa-v3 system uses a shared encoder and two classification heads: one for the three-way `at` classification task and one for the binary `isAt` classification task. The model is trained with weighted cross-entropy to compensate for class imbalance, since negative examples dominate both subtasks.

We used a standard classification, Categorical Cross-Entropy (CCE). However, in a multi-view setup, we calculate the loss for each view independently and sum them up:

$$\mathcal{L}_{CE} = \sum_{i=1}^{V} \mathcal{L}_{CE}^{i}(y, \hat{y}_i)$$

where V is the number of views, in our case, equal to 3, y is the true label, and $\hat{y}_i$ is the prediction from view i.

The model is penalised if the probability distribution of View 1 differs significantly from View 2. This forces the model to find features that are invariant across all views.

$$\mathcal{L}_{JSD} = D_{KL}(P_{view1} || P_{view2})$$

By using the InfoNCE (Information Noise-Contrastive Estimation) loss, the model is trained to recognise that View 1 and View 2 of a given document should be closer in vector space than View 1 of the same document and View 2 of any other document in the batch.

$$\mathcal{L}_{InfoNCE} = -\log \frac{\exp(\text{sim}(z_i, z_j)/\tau)}{\sum_{k=1}^{2N} \exp(\text{sim}(z_i, z_k)/\tau)}$$

In the multi-view setting, each candidate pair is represented through several complementary views as presented in Figure 1. The first view contains the original marked text. Additional views contain enriched versions of the input, including natural-language analyses generated using

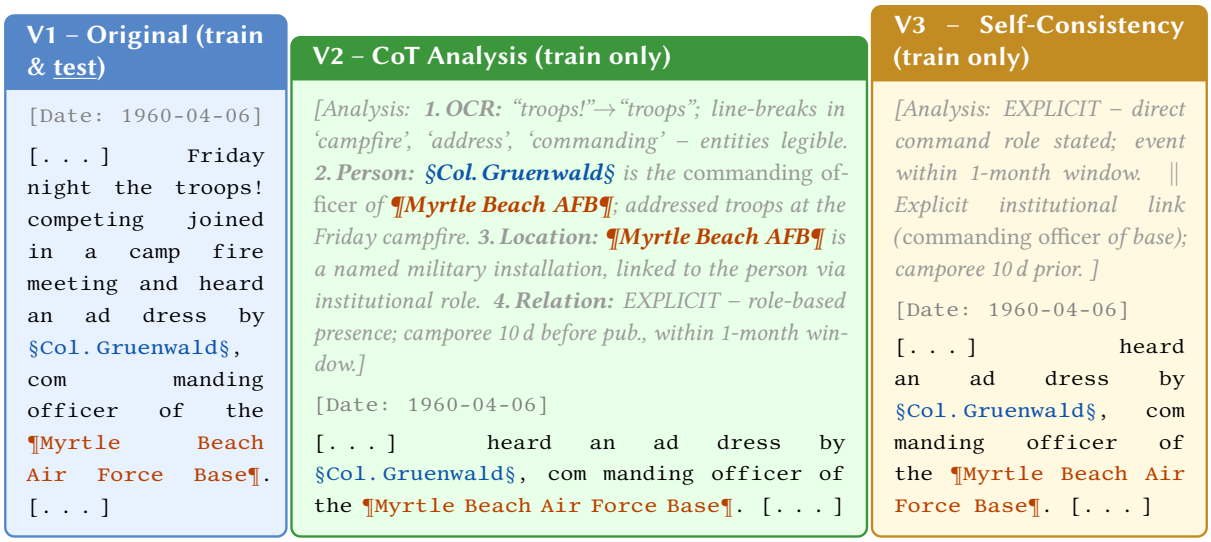

Figure 1: Multi-views used for Run 3. Left view correspond to original data with a marking strategy only and it is used during training and test. Center view is a Chain-of-Thoughts (CoT) extension of the V1 with V1 concatenated at the end. Right is the self-consistency view. Center and Right views are only used during training.

`Qwen/Qwen2.5-7B-Instruct` (4-bit quantized). These enriched views are intended to make implicit evidence more explicit, especially for temporal reasoning and for the distinction between TRUE, PROBABLE, and FALSE.

The multi-view model is trained with a combination of classification and consistency objectives. The training objective can be summarised as:

$$\mathcal{L}_{total} = \alpha\mathcal{L}_{CE} + \beta\mathcal{L}_{JSD} + \gamma\mathcal{L}_{InfoNCE}$$

where $\mathcal{L}_{CE}$ is the weighted cross-entropy loss, $\mathcal{L}_{JSD}$ [11] encourages prediction consistency across views, and $\mathcal{L}_{InfoNCE}$ [12] (with in-batch negatives) is a contrastive objective used to align representations of the same pair across different views. At inference time, the system uses the original marked text view only, making the final prediction pipeline more efficient.

2.4. Official Runs

Table 1 summarises the submitted systems.

Table 1
Overview of the submitted CLEF-HIPE systems.

Run	System	Description
Run 1	Fine-tuned DeBERTa–GLiREL	GLiREL relation extraction model with DeBERTA encoder fine-tuned on the CLEF-HIPE relation extraction data.
Run 2	GLiREL-Score MLP	GLiREL relation extraction model and MLP classifier trained on GLiREL output scores
Run 3	mDeBERTa-v3-base multi-view	Multilingual DeBERTa-v3-base encoder trained with marked entity spans, temporal input, and multi-view enriched representations.

3. Results

This section reports the official HIPE-2026 results for our team, identified as `MILRIT` in the official ranking files. The official *Accuracy Profile* is computed on the *Impresso* test files. For each language file,

the impresso_profile_score is defined as the mean of the at macro-recall and the isAt macro-recall. The final overall score is the mean of this profile score over the submitted language files. The *Efficiency Profile* combines the accuracy rank, the parameter-count rank, and the model-size rank, while the *Balanced Efficiency Profile* gives more weight to the accuracy rank. The values reported below are taken from the official HIPE-2026 evaluation results.

Table 2
Official *Accuracy Profile* ranking. The table reports the top run-level submissions for context and includes all three MILRIT runs. The team-rank column keeps only the best run per team; non-best runs from the same team are marked with "–". MILRIT runs are highlighted.

Run rank	Team rank	Team	Run	Mean Impresso score
1	1	Spinfo	run1	0.7479
2	–	Spinfo	run3	0.7289
3	2	MaxFo-Ajie	run1	0.7001
4	–	Spinfo	run2	0.6890
5	3	whereami	run1	0.6880
6	–	whereami	run2	0.6833
7	–	MaxFo-Ajie	run2	0.6690
8	4	Awakened	run3	0.6671
–	5	INSA Lyon	run1	0.6390
–	6	gipplab	run2	0.6271
–	7	Hansel&Gretel	run3	0.6221
15	8	MILRIT	run3	0.5951
21	–	MILRIT	run1	0.5623
44	–	MILRIT	run2	0.4264

3.1. Accuracy Profile

Table 2 reports the official overall *Accuracy Profile* results in a single view. The table keeps the strongest run-level submissions for context and includes all three MILRIT runs inline. Since teams could submit several runs, the run-rank column corresponds to the official ranking of individual submissions. The team-rank column gives the corresponding ranking when only the best run per team is retained.

Our best submission was MILRIT Run 3, which obtained a mean Impresso profile score of 0.5951 and ranked 15th among complete submitted runs. Under the team-level view, where only each team's best run is considered, MILRIT ranked 8th overall. Run 1 ranked 21st at the run level with a score of 0.5623, while Run 2 ranked 44th with a substantially lower score of 0.4264.

These results indicate that the mDeBERTa-v3-base multi-view approach used in Run 3 was the most robust of our submitted systems, followed by the fine-tuned DeBERTa–GLiREL setting used in Run 1. In contrast, the lower performance of Run 2 suggests that the non-fine-tuned GLiREL-Score MLP configuration did not generalise as expected on the official test set.

Table 3
Per-language official *Accuracy Profile* results for the MILRIT submissions. Detailed macro-recall scores are reported for Run 3, our best overall submission. MR denotes macro-recall.

Language	Run 1 score	Run 2 score	Run 3 score	Run 3 rank	Run 3 at MR	Run 3 isAt MR
German	0.5802	0.4365	0.5886	16	0.5339	0.6433
English	0.4808	0.4353	0.5881	15	0.5099	0.6663
French	0.6261	0.4073	0.6087	19	0.5428	0.6746

Table 3 reports the per-language *Accuracy Profile* results for the three MILRIT submissions, with detailed macro-recall scores for Run 3, our best overall submission. Run 3 obtained similar scores for German and English, with slightly higher performance on French. Both Run 1 and Run 3 performed best on French, while Run 2 was weaker across all languages.

3.2. Efficiency Profile

Table 4
Run-level *Efficiency Profile* ranking with corresponding team-level rank. Lower mean efficiency rank is better. The team-rank column keeps only the best run per team; non-best runs from the same team are marked with "–". MILRIT runs are highlighted using the same colour code as in the accuracy tables.

Run rank	Team rank	Team	Run	Mean eff. rank	Accuracy score	Params	Size MB
1	1	MILRIT	run3	10.6667	0.5951	277,730,309	1,111
2	2	FI-CODE	run2	11.3333	0.4734	0	0
3	3	DS@GT_HIPE	run1	11.6667	0.5142	2,087,375	87
4	–	DS@GT_HIPE	run2	13.0000	0.4836	2,087,375	87
5	–	DS@GT_HIPE	run3	13.3333	0.4771	2,087,375	87
6	–	MILRIT	run1	14.6667	0.5623	466,577,920	1,780

The official results also include an *Efficiency Profile*, which combines the accuracy rank with resource-related ranks, namely parameter count and model size. Table 4 reports the run-level ranking together with the corresponding team-level rank, where only the best run per team is retained. Our Run 3 ranked first at both levels, with a mean *Efficiency Profile* rank of 10.6667. Run 1 also appeared among the strongest efficiency-oriented submissions, ranking sixth at the run level.

3.3. Balanced Efficiency Profile

The *Balanced Efficiency Profile* gives more weight to accuracy than the standard *Efficiency Profile*: 0.5 for accuracy rank, 0.25 for parameter-count rank, and 0.25 for model-size rank. As shown in Table 5, MILRIT Run 3 ranked first at both the run and team levels, while Run 1 ranked fifth at the run level.

Table 5
Run-level *Balanced Efficiency Profile* ranking with corresponding team-level rank. Lower balanced efficiency rank is better. The team-rank column keeps only the best run per team; non-best runs from the same team are marked with "–". MILRIT runs are highlighted using the same color code as in the previous tables.

Run rank	Team rank	Team	Run	Balanced eff. rank	Accuracy rank	Param. rank	Size rank
1	1	MILRIT	run3	10.25	12	8	9
2	2	whereami	run1	13.5	4	22	24
3	3	DS@GT_HIPE	run1	13.75	23	5	4
4	–	whereami	run2	14	5	22	24
5	–	MILRIT	run1	14.75	18	12	11

Overall, our best accuracy result was obtained by Run 3, the mDeBERTa-v3 multi-view system. It ranked 15th among individual runs and 8th at the team level in the overall *Accuracy Profile*, with a score of 0.5951. It also outperformed the official baseline score of 0.5818. Run 1, the fine-tuned DeBERTa within the GLiREL system, ranked 21st among runs with a score of 0.5623, while Run 2, the GLiREL-score MLP, ranked 44th with a score of 0.4264. The contrast between Run 1 and Run 2 suggests that fine-tuning GLiREL on the available data improved generalisation in this configuration. In contrast, Run 3 achieved the strongest accuracy–efficiency compromise, ranking first in both the *Efficiency Profile* and the *Balanced Efficiency Profile*.

4. Discussion

The submitted runs explore two complementary directions. Runs 1 and 2 evaluate whether a generalist relation extraction model can be adapted to historical person–place relation extraction or reused as a source of relation-evidence features. Run 1 tests task-specific adaptation through fine-tuning, while

Run 2 keeps GLiREl fixed and trains an MLP classifier using its relation score vectors. This comparison helps assess whether GLiREL's score space contains useful signals for the target labels without model fine-tuning.

Run 3 investigates a more specialised approach based on multilingual encoder fine-tuning, explicit entity marking, temporal anchoring, and multi-view learning. The use of entity markers helps the model focus on the target person–place pair, while the date prefix provides a temporal anchor for the `isAt` relation. Multi-view training further encourages the model to learn robust representations from both original and enriched inputs.

The main challenges remain OCR noise, class imbalance, and the ambiguity of person–place relations. In particular, distinguishing TRUE from PROBABLE is difficult because many historical relations are not stated explicitly but can be inferred from context. Similarly, `isAt` requires temporal interpretation: the system must decide not only whether a relation exists, but whether it holds close to the publication date.

Overall, the three runs allow us to compare encoder-level adaptation within a generalist relation extraction model, lightweight classification based on relation-score features, and a task-specific multilingual encoder approach. The results provide insight into how much task adaptation, temporal information, and enriched training views contribute to person–place relation extraction in historical newspapers.

5. Conclusions

This paper presents the MILRIT participation in the HIPE-2026 shared task on person–place relation extraction from multilingual historical newspapers. We submitted three runs covering two methodological directions: generalist relation extraction with GLiREL and task-specific multilingual encoder learning with multi-view training.

The comparison between our submissions suggests that model type and adaptation strategy strongly influence performance in historical relation extraction. The generalist DeBERTa–GLiREL model showed that direct fine-tuning improved generalisation in our setting. By contrast, the multi-view mDeBERTa-v3 approach, which combines a compact multilingual encoder with explicit entity marking and temporal anchoring, achieved our best accuracy and efficiency results. This suggests that, for noisy historical newspapers, task-oriented encoder models with carefully designed input representations may be more robust than directly adapting generalist relation extraction models.

The results also highlight the difficulty of the task. Historical newspaper text contains OCR noise, spelling variation, multilingual content, indirect relations, and temporally ambiguous evidence. In particular, distinguishing TRUE from PROBABLE remains challenging, since many relations are not explicitly stated but can only be inferred from context. The `isAt` relation adds an additional temporal dimension, requiring the system to decide whether a person–place relation holds close to the publication date.

Future work should further investigate temporal modeling, richer document-level context, and more robust strategies for handling noisy OCR. In particular, better integration of event time, publication time, and entity-specific historical context could help distinguish long-term associations from short-term location evidence.

Acknowledgments

This work has been co-funded by the French national research agency (ANR) through the project ANR-25-CE38-6695 (MILL-EHNAS), and by the European Union HORIZON-WIDERA-2023-TALENTS-01-01 grant 101186647 — AI4DH. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

Declaration of Generative AI Use

During the preparation of this work, the authors used ChatGPT in order to: Polish sentences, Rephrase and reword. Additionally, for run 3, Claude was used to implement the idea drafted by the authors. After using these tools, the authors reviewed, validated, and edited the content as needed and take full responsibility for the publication's content.

References

[1] J. A. Diaz-Garcia, J. A. D. Lopez, A survey on cutting-edge relation extraction techniques based on language models, Artificial Intelligence Review 58 (2025) 287.

[2] L. Xue, D. Zhang, Y. Dong, J. Tang, Autore: Document-level relation extraction with large language models, in: Proceedings of the 62nd annual meeting of the association for computational linguistics (volume 3: System demonstrations), 2024, pp. 211–220.

[3] H. Zheng, S. Wang, L. Huang, A comprehensive survey on document-level information extraction, in: Proceedings of the Workshop on the Future of Event Detection (FuturED), 2024, pp. 58–72.

[4] J. Opitz, C. Raclé, A. Michail, M. Romanello, M. Ehrmann, S. Clematide, Overview of HIPE-2026: Person–Place Relation Extraction from Multilingual Historical Texts, in: M. Hagen, M. Potthast, B. Stein, P. Schaer, E. Zangerle, S. MacAvaney, J. M. Struß, E. Sánchez Salido, A. Barrón-Cedeño, A. García Seco de Herrera (Eds.), Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the Seventeenth International Conference of the CLEF Association (CLEF 2026), Lecture Notes in Computer Science (LNCS), Springer, 2026.

[5] J. Opitz, C. Raclé, A. Michail, M. Romanello, E. Boros, S. Gabay, M. Ehrmann, S. Clematide, Extended Overview of HIPE-2026: Evaluating Accurate and Efficient Person–Place Relation Extraction from Multilingual Historical Texts, in: E. Sánchez Salido, A. Barrón-Cedeño, A. García Seco de Herrera, S. MacAvaney, J. M. Struß (Eds.), CLEF 2026 Working Notes, CEUR Workshop Proceedings, volume 3180, CEUR-WS, 2026.

[6] J. Boylan, C. Hokamp, D. G. Ghalandari, Glirel-generalist model for zero-shot relation extraction, in: Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers), 2025, pp. 8230–8245.

[7] P. He, J. Gao, W. Chen, Debertav3: Improving deberta using electra-style pre-training with gradient-disentangled embedding sharing, in: The Eleventh International Conference on Learning Representations, 2021.

[8] L. Baldini Soares, N. FitzGerald, J. Ling, T. Kwiatkowski, Matching the blanks: Distributional similarity for relation learning, in: A. Korhonen, D. Traum, L. Màrquez (Eds.), Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Florence, Italy, 2019, pp. 2895–2905. URL: https://aclanthology.org/P19-1279/. doi:10.18653/v1/P19-1279.

[9] J. G. Moreno, A. Doucet, B. Grau, Relation classification via relation validation, in: Proceedings of the 6th Workshop on Semantic Deep Learning (SemDeep-6), 2021, pp. 20–27.

[10] J.-B. Grill, F. Strub, F. Altché, C. Tallec, P. Richemond, E. Buchatskaya, C. Doersch, B. Avila Pires, Z. Guo, M. Gheshlaghi Azar, et al., Bootstrap your own latent-a new approach to self-supervised learning, Advances in neural information processing systems 33 (2020) 21271–21284.

[11] Y. Tian, C. Sun, B. Poole, D. Krishnan, C. Schmid, P. Isola, What makes for good views for contrastive learning?, Advances in neural information processing systems 33 (2020) 6827–6839.

[12] P. Khosla, P. Teterwak, C. Wang, A. Sarna, Y. Tian, P. Isola, A. Maschinot, C. Liu, D. Krishnan, Supervised contrastive learning, Advances in neural information processing systems 33 (2020) 18661–18673.