A Multitask Approach Based on RoBERTa, Temporal Transformer, and Person-Location Relation Extraction in Multilingual Historical Text*

Notebook for the HIPE Lab at CLEF 2026

Danileth Almanza-Gonzalez[1,*], Juan Carlos Martinez-Santos[1,†] and Edwin Puertas[1,†]

[1]Universidad Tecnológica de Bolívar, Cartagena, Colombia

Abstract

This study presents an approach to extracting temporal person-location relations in multilingual historical documents in the context of the HIPE-2026 task. We implemented an architecture that integrates *XLM-RoBERTa-base*, Named Entity Recognition (NER), Optuna-based hyperparameter optimization, balancing strategies, and a lightweight Temporal Transformer module aimed at temporal and contextual refinement. The proposed system aims to determine whether a person is associated with a location within the document's immediate temporal context or at some point in the past. Experimental results demonstrated that the progressive incorporation of temporal and contextual mechanisms improved system performance, particularly for the *isAt* relation. The best configuration achieved a Accuracy Profile Score of 0.4842 and a Generalization Profile Score of 0.3726 in the official HIPE-2026 evaluation. Furthermore, the VerbaNexAI I team achieved 29th place in the official precision profile and 9th place in the combined precision-efficiency profile, demonstrating competitive performance with a moderate size multilingual architecture and controlled computational cost.

Keywords

Named Entity Recognition, Temporal Reasoning, Multilingual Transformers

1. Introduction

The identification of temporal and spatial relationships in texts constitutes a fundamental challenge within Natural Language Processing (NLP), especially in contexts where historical, geographical, and contextual information plays a relevant role in semantic interpretation. The ability to recognize named entities, such as persons, locations, organizations, and events, as well as to understand the relationships established among them over time, is essential for multiple applications related to knowledge graphs, historical reconstruction, document analysis, and intelligent information retrieval systems. In this context, Named Entity Recognition (NER) and temporal relation extraction have gained increasing importance due to the need to interpret complex texts containing implicit and explicit references to events occurring at different historical moments and geographical locations [1].

One of the main challenges in temporal and spatial relation extraction is understanding not only the presence of entities within the text, but also the semantic and temporal context in which they interact. In historical documents, relationships between persons and locations are not explicitly stated but must be inferred from linguistic cues and narrative references. Additionally, factors such as language, orthographic variation, and the presence of multiple languages significantly hinder automatic systems ability to identify these relationships correctly. Consequently, the development of models capable of

CLEF 2026 Working Notes, 21 – 24 September 2026, Jena, Germany

*Corresponding author.

†These authors contributed equally.

daalmanza@utb.edu.co (D. Almanza-Gonzalez); jcmartinezs@utb.edu.co (J. C. Martinez-Santos); epuerta@utb.edu.co (E. Puertas)

https://github.com/Danileth (D. Almanza-Gonzalez)

0000-0002-1664-0008 (D. Almanza-Gonzalez); 0000-0003-2755-0718 (J. C. Martinez-Santos); 0000-0002-0758-1851 (E. Puertas)

integrating contextual, temporal, and semantic information has become a fundamental necessity for advancing tasks related to historical reconstruction, document analysis, and knowledge generation from digitized historical collections.

Due to this need, the shared task HIPE-2026: "Evaluating Accurate and Efficient Person–Place Relation Extraction from Multilingual Historical Texts", focused on the multilingual extraction of person-place relations in historical documents, aimed to develop systems capable of identifying whether a person was actually associated with a specific location and determining whether such a relationship occurred within the immediate temporal context of the document or at a past moment. The task sought to overcome the limitations of traditional approaches based solely on textual co-occurrence by promoting models that can understand the semantic, temporal, and geographical context of documents. The evaluation criteria are: accuracy profile, generalization, and accuracy-efficiency [2] [3] [4].

In this context, we addressed this by using a RoBERTa-based architecture that integrates Temporal Transformer Encoding and automatic hyperparameter optimization through Optuna. The proposed approach aimed to improve the contextual understanding of person-place relationships by incorporating explicit and implicit temporal information present in historical documents. The source code and the experiments conducted are publicly available through the repository: [1].

2. Related Work

The identification of temporal relations and named entity recognition has become increasingly relevant in the field of Natural Language Processing (NLP) due to the need to understand the contextual, spatial, and temporal information contained in large volumes of text. These tasks enable analysis of how people, places, events, and organizations are related across different historical, medical, and documentary contexts, contributing to the development of intelligent information retrieval systems, knowledge graph construction, and advanced semantic analysis. In this context, multiple studies have explored techniques based on deep learning, neural models, and contextual representations to improve temporal relation extraction in different domains. In one study, the authors implemented temporal information extraction techniques based on Natural Language Processing and deep learning to identify relations between medical events and the time of document creation. The proposed model integrated contextual mechanisms and dynamic representation through Bi-LSTM, sliding windows, and a dynamic routing algorithm inspired by Capsule Networks [5]. Similarly, another study employed techniques for extracting and normalizing temporal expressions in Chinese texts. The proposed method, called IMLLF, integrated Conditional Random Fields (CRF), linguistic features, knowledge-based rules, and temporal reasoning to identify time points and periods. They incorporated lexical, graphical, and co-occurrence features together with experimental comparisons against CRF and LSTM-CRF models [6]. Likewise, a study proposed the SDLG model, which combined latent dynamic graphs, Graph Convolutional Networks (GCN), attention mechanisms, and syntactic dependencies to identify temporal and hierarchical relations between events [7]. Another approach introduced the CPTRE (Contrastive Prototypical Temporal Relation Extraction) model, which integrated contrastive learning, prototypical networks, R-GCN, and contextual representations based on BERT and BigBird for temporal relation extraction between events [8]. Within this line of research, multitask approaches and hybrid systems focused on temporal reasoning have also been developed. An example of this is TemPrompt, a multitask framework for temporal relation extraction that combines prompt tuning, contrastive learning, and event temporal reasoning over pretrained models such as RoBERTa and T5 [9]. Complementarily, another study developed a system based on textual preprocessing techniques, temporal normalization, and rule-based parsing using the *reparse* library, to automatically identify temporal information in Italian clinical referrals related to medical follow-up examinations [10]. Likewise, a study proposed ABTE-NET, an encoder-decoder architecture based on ALBERT and Bi-LSTM to extract timelines from the life stories of older adults automatically. This approach incorporated summary extraction techniques, attention networks, RNN, Bi-LSTM, and ROUGE metrics to identify events, participants,

[1] https://github.com/VerbaNexAI/CLEF2026

time, and location within disordered biographical texts [11]. In this regard, various approaches have explored integrating temporality, contextual reasoning, and Named Entity Recognition across multiple domains and multilingual scenarios. Due to this growing interest, the HIPE initiative emerged to analyze digitized historical documents using advanced Natural Language Processing (NLP) techniques. In its HIPE-2020 edition, the primary objective was to evaluate and strengthen the recognition, classification, and linking of named entities in digitized historical newspapers across.

3. Data

The HIPE-2026 organizers provided a multilingual dataset aimed at person-place relation extraction in historical documents. The training corpus included documents in German, English, and French that had been previously annotated with temporal and spatial relations between person and place entities. The German subset consisted of 466 instances, the English subset of 307, and the French subset of 478, for a consolidated total of 1,251 instances across 104 historical documents. The dataset included two main classification tasks: the *at* relation, focused on determining whether a person resided in or visited a place before the publication of the document, and the *isAt* relation, which aims to determine whether a person was present at a specific location during the precise time frame described in the documen. For the *at* task, the dataset contained 690 instances labeled *FALSE*, 441 labeled *TRUE*, and 120 labeled *PROBABLE*. In contrast, the *isAt* task consisted of 972 *FALSE* instances and 279 *TRUE* instances. These annotations enabled the modeling of complex scenarios involving temporal inference, contextual reasoning, and semantic understanding in multilingual historical texts affected by data imbalance and linguistic variations characteristic of different historical periods [12] [13] [14].

4. Architecture

Figure 1 presents the general architecture proposed for the task. The system was designed as a multilingual multitask framework for extracting person-location temporal relations in historical documents. The architecture integrates preprocessing, regularization, feature extraction through named entity recognition, hyperparameter optimization with Optuna, a Temporal Transformer module, and a multilingual encoder based on XLM-RoBERTa for contextual representation learning to improve the identification of *at* and *isAt* relations by incorporating semantic, temporal, and contextual information extracted from multilingual historical texts.

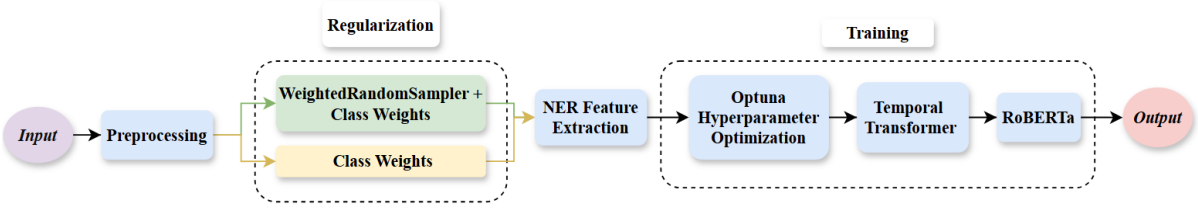

Figure 1: Architecture system.

4.1. Preprocessing

During preprocessing, historical documents underwent normalization to reduce noise and address heterogeneous structures typical of digitized historical texts. Key steps included text normalization, elimination of duplicate spaces, correction of fragmented words, date standardization, extraction of relevant metadata, and harmonization of multilingual textual formats. Furthermore, contextual windows containing textual evidence related to person and location mentions were extracted.

4.2. Regularization

We implemented techniques to address class imbalance in the dataset and used different regularization and balancing strategies during model training. Specifically, the *WeightedRandomSampler* technique was used to balance the sampling distribution of training instances according to the combination of labels associated with the *at* and *isAt* relations. Additionally, class weights were applied to the loss functions to reduce the impact of class imbalance. For the *at* relation, which contains three classes (*TRUE*, *FALSE*, and *PROBABLE*), a weighted *CrossEntropyLoss* function was employed. On the other hand, for the *isAt* relation, corresponding to a binary classification task, *BCEWithLogitsLoss* with positive class weighting was used to improve the identification of minority positive instances.

4.3. NER Feature Extraction

The architecture incorporates a Named Entity Recognition (NER)-based feature extraction module with the purpose of enriching the contextual representation of each candidate Person–Location pair. Although the HIPE corpus provides the target mentions already annotated, the NER module is not used to identify them again, but rather to obtain additional contextual information present within the evidence window surrounding both entities. This complementary information makes it possible to characterize the semantic environment in which each historical relation appears and to provide additional evidence for temporal reasoning.

From the evidence window, structural and semantic features related to the detected entities are automatically extracted, including the number of mentioned persons and locations, the presence of the target entities, the number of additional entities, and the co-occurrence patterns between persons and locations. Likewise, linguistic and temporal cues are identified, such as motion verbs, reporting expressions, spatial prepositions, past and present markers, the occurrence of both entities within the same sentence, and the textual distance between their mentions. These features are represented through binary variables, normalized counts, and discretized attributes, depending on their nature, allowing the explicit representation of the temporal and contextual structure of the historical document. [15][16]

Furthermore, the extracted features are integrated into a feature vector that is subsequently fused with the contextualized representations generated by XLM-RoBERTa within the Temporal Transformer module. In this way, the model combines contextual information learned through deep representations with explicit linguistic knowledge derived from the entities and the historical context, promoting a more informative representation for the classification of temporal relations between persons and locations.

4.4. Optuna Hyperparameter Optimization

During the Hyperparameter optimization process, we employed Optuna with Bayesian optimization based on the *Tree-structured Parzen Estimator* (TPE). The optimization process explored different configurations related to learning rate, dropout, weight decay, maximum sequence length, balancing coefficients, and positive class weights. In addition, we incorporated the *MedianPruner* strategy to eliminate unpromising trials during training, reducing computational cost and improving convergence efficiency. This process enabled the acquisition of more stable and generalizable configurations for the multilingual multitask architecture.

4.5. Temporal Transformer

The proposed architecture incorporates a refinement module called the Temporal Transformer on top of the multilingual encoder *XLM-RoBERTa-base* [17], with the aim of strengthening the modeling of temporal relations between person and location entities in multilingual historical documents. Initially, XLM-RoBERTa generates contextualized representations for each token in the document. Based on these representations and the document content itself, a specific temporal representation is constructed for each candidate Person–Location pair using the document date, the header, the dateline, and a context window containing both entities. From this evidence, temporal and contextual features related

to past and present markers, movement and reporting verbs, spatial relations between entities, textual distance between mentions, sentence-level co-occurrence, and features obtained through named entity recognition are automatically extracted, allowing the historical context associated with each relation to be represented explicitly.

Subsequently, this temporal information is integrated with the representations generated by XLM-RoBERTa through the application of Temporal Positional Encoding. Rather than relying solely on conventional positional information, the module incorporates temporal representations that enrich the contextualized sequence according to the temporal features extracted from the document. In this way, the representation of each token not only preserves the linguistic context learned by XLM-RoBERTa but also incorporates information related to the chronology and temporal structure of the document, enabling more precise discrimination between relations corresponding to past events, contemporary events, and historical references present within the same text.

The enriched representations are refined through a Transformer-based layer, which incorporates an attention mechanism complemented by semantic similarity information between tokens to strengthen long-range contextual dependencies. This refinement highlights the interactions between words that describe the same historical relation, even when they appear far apart within the document. The global representation obtained from the refined sequence is fused with the previously constructed temporal representation to generate a joint representation used by the *at* and *isAt* classification tasks. This strategy combines the multilingual contextual knowledge learned by XLM-RoBERTa with explicit temporal information derived from the document, promoting more accurate temporal reasoning over historical relations between people and locations.

4.6. Evaluation Metrics

The evaluation criteria defined in HIPE-2026 are: the Accuracy Profile, the Generalization Profile, which used unseen surprise data in French, and the Accuracy Profile, which assessed the balance between classification performance, computational cost, model size, and inference time. Furthermore, in the experimental design, we used Accuracy as the primary evaluation metric due to its ability to assess the performance of the *at* and *isAt* tasks while mitigating the impact of class imbalance.

5. Experiments Conducted and Training

The experimental design was structured through an incremental analysis of the proposed architecture evaluate the individual contribution of each component incorporated into the system. In all experiments, the same multilingual model, *XLM-RoBERTa-base*, was used as the foundation to ensure a comparison among configurations. The first configuration corresponded to the baseline model, which included the fundamental preprocessing, regularization, and balancing defined in the architecture. Subsequently, we incorporated the NER-based feature extraction module, the Optuna hyperparameter optimization process, and finally the Temporal Transformer refinement module.

Table 1 shows the performance of 0.3611 for the *at* task, 0.5002 for *isAt*, and a macro average of 0.4307. These values reflect the difficulty of the task, particularly for the *at* relation, which presents a more complex multiclass structure considering the labels *TRUE*, *FALSE*, and *PROBABLE*. After incorporating NER features, performance increased to 0.4003 for *at* and 0.5981 for *isAt*, yielding a macro average of 0.4992. This improvement demonstrates that named entity-based information contributed significantly to the recognition of person-location relations, especially in the binary *isAt* task.

The incorporation of Optuna primarily improved the performance of the *at* task, increasing from 0.4003 to 0.4415, while the *isAt* task showed a slight decrease, from 0.5981 to 0.5937. However, the macro average increased from 0.4992 to 0.5176, indicating that hyperparameter optimization favored a more balanced configuration between both tasks. This behavior suggests that Optuna enabled adjustment to sensitive training parameters, such as the learning rate, dropout, regularization weight, maximum sequence length, and class imbalance. Furthermore, to ensure the reproducibility of the training process, the final hyperparameters automatically selected by Optuna after exploring the predefined search

space were used. The best configuration corresponded to a maximum sequence length of 384 tokens, a learning rate of 1.24×10^{-5} for the XLM-RoBERTa encoder, and 7.51×10^{-5} for the classification and temporal refinement layers. These hyperparameters were used during the final training of the model and enabled a more balanced configuration between the *at* and *isAt* tasks, improving the stability of the learning process and the overall performance of the system.

The complete configuration, composed of the baseline model, NER, Optuna, and Temporal Transformer, achieved the best overall results, with 0.4690 for *at*, 0.6381 for *isAt*, and a macro average of 0.5535. This result shows that the Temporal Transformer provided an additional improvement by incorporating temporal and contextual information regarding the relationships between persons and locations. Compared to the baseline model, the complete architecture increased the macro average from 0.4307 to 0.5535, representing an absolute improvement of 0.1228. This confirms that the combination of linguistic features, automatic optimization, and temporal refinement enabled the construction of a more robust representation for temporal relation extraction in multilingual historical documents.

Additionally, Table 2 compares two balancing strategies using the same complete architecture, the model *XLM-RoBERTa-base* with NER, Optuna, and Temporal Transformer. The first strategy combined *WeightedRandomSampler* with class weights, whereas the second strategy used only class weights within the loss functions. With the combined strategy, scores of 0.4690 for *at*, 0.6381 for *isAt* were obtained with a macro average of 0.5535. In contrast, when the model used class weights, the performance for *at* increased to 0.5034, but that for *isAt* decreased to 0.5748, resulting in a macro average of 0.5391.

These results show that the exclusive use of class weights favored the *at* task, because it allowed a better adjustment of the contribution of each class within the loss function without modifying the sampling distribution during training. Nevertheless, the combination of *WeightedRandomSampler* and class weights produced better overall performance, especially for the *isAt* task, where it achieved 0.6381. Since the macro average was higher with the combined strategy, this configuration was selected as the final system setup. The results demonstrate that balancing does not affect both tasks in the same way: while *at* benefits more from the exclusive use of class weights, *isAt* achieves better results when weighted resampling is combined with weighting in the loss function.

The obtained results demonstrate that the progressive incorporation of specialized components improved the system's ability to model complex temporal relations in multilingual historical documents. In particular, the observed improvement in the *isAt* relation in the model was able to more accurately capture contextual signals associated with immediate temporality, such as reporting expressions, nearby geographical references, and linguistic constructions dependent on the documentary context. This is especially relevant in historical scenarios, where relationships between persons and places are not always explicitly formulated and often require temporal and semantic inferences from textual structures affected by OCR noise and historical linguistic variation. In this sense, the combined use of multilingual contextual representations, temporal cues, and semantic refinement mechanisms strengthened the system's generalization capability across multiple languages and document types.

Table 1
Effect of NER, Optuna, and Temporal Transformer on model performance evaluated on the training data.

Configuration	Accuracy Profile (AT)	Accuracy Profile (isAt)	Macro Average
Baseline Model	0.3611	0.5002	0.4307
Baseline + NER	0.4003	0.5981	0.4992
Baseline + NER + Optuna	0.4415	0.5937	0.5176
Baseline + NER + Optuna + Temporal Transformer	0.4690	0.6381	0.5535

Table 2
Comparison of balancing strategies evaluated on the training data using the full proposed architecture.

Balancing Strategy	Accuracy Profile (AT)	Accuracy Profile (isAt)	Macro Average
WeightedRandomSampler + Class Weights	0.4690	0.6381	0.5535
Class Weights Only	0.5034	0.5748	0.5391

6. Results

The official results show the performance of the proposed system on the test sets provided by the organizers. For the final evaluation, two official runs were submitted using test data with 19 instances of multilingual historical documents. Both configurations used the same base architecture, *XLM-RoBERTa-base*; however, one run incorporated the *Temporal Transformer* module, while the other used only the multitask architecture without additional temporal refinement. As shown in Table 3, the best configuration corresponded to the model integrating the temporal module, achieving a *Precision Profile Score* of 0.4842 and a *Generalization Profile Score* of 0.3726, consistently outperforming the configuration without temporal refinement.

In Table 3 show that incorporating the *Temporal Transformer* significantly improved the system's generalization to previously unseen historical documents. The improvement in the generalization profile is that temporal refinement mechanisms enabled the capture of more robust contextual dependencies related to implicit and explicit temporal references in historical texts. This behavior is particularly relevant given the complexity of the HIPE-2026 scenario, where person-location relations require distinguishing between past historical associations and locations in the immediate temporal context of the document. Likewise, the increase in the precision profile demonstrates that integrating temporal and semantic information reduced the contextual ambiguities present in documents affected by OCR noise, heterogeneous structures, and historical language variations.

On the other hand, the results from the efficiency profile show that the system maintained a competitive balance between performance and computational cost by using a relatively compact architecture compared to larger-scale models. Despite employing a model with approximately 355 million parameters and a size of approximately 1424 MB, the system achieved competitive results through by integrating multiple complementary strategies, including named entity recognition, hyperparameter optimization with Optuna, class balancing, and contextual temporal refinement. These results demonstrate that model size alone does not determine final performance; rather, it is the ability to integrate specialized mechanisms for temporal and semantic reasoning that determines final performance.

Additionally, the experiments show that the combination of linguistic, temporal, and contextual techniques improved the performance of the multilingual model without the need for excessively large architectures or high-computational-cost foundational models. This demonstrates that moderate-sized architectures, when complemented with specialized modules and appropriate optimization strategies, can achieve solid results in complex multilingual historical relation extraction tasks. In this sense, the proposed approach demonstrates the potential of combining multilingual Transformers with lightweight temporal mechanisms to address historical and spatial reasoning problems within the context of digital humanities and historical knowledge extraction.

In this sense, the experiments reveal a difference in behavior between the *at* and *isAt* relations. While the *at* task exhibited a more complex behavior due to the presence of the intermediate *PROBABLE* class, the *isAt* task showed greater sensitivity to the immediate temporal signals incorporated by the model. This demonstrates that both relations require different levels of contextual and temporal reasoning. The identification of historical relations involves more ambiguous semantic inferences that are highly dependent on narrative context. In contrast, relations in the immediate context of the document are usually supported by more explicit linguistic cues. This difference explains why certain architectural and balancing strategies impacted the performance of each task differently.

Likewise, the results show that the progressive incorporation of techniques led to cumulative im-

provements within the system. The NER module strengthened the semantic representation of relevant entities, Optuna stabilized sensitive training configurations, and the *Temporal Transformer* improved the modeling of complex temporal dependencies. Finally, the experiments also highlight the importance of considering temporality as an explicit component within historical relation extraction systems. In many cases, the mere co-occurrence of a person and a location within the text does not guarantee the existence of a valid temporal relation. For this reason, incorporating mechanisms that model temporal dependencies and contextual signals helps reduce erroneous associations derived solely from textual co-occurrence. From this perspective, the proposed approach provides evidence for the need to integrate temporal and contextual reasoning into tasks for the automatic construction of historical knowledge from digitized multilingual documents.

Table 3
HIPE-2026 validation data results obtained for the Precision and Generalization Profiles.

Run	Accuracy Profile	Generalization Profile
VerbaNexAI I – Run 1	0.4628	0.3346
VerbaNexAI I – Run 2	0.4842	0.3726

Table 4
Efficiency-oriented evaluation on the HIPE-2026 validation data.

Run	Efficiency Rank Score	Accuracy Score	Parameter Count	Model Size
VerbaNexAI I – Run 1	18.0000	0.4628	355,000,000	1424 MB
VerbaNexAI I – Run 2	15.6667	0.4842	355,000,000	1424 MB

On the other hand, the results obtained by the best system configuration, corresponding to *VerbaNexAI I – Run 2*, are presented for each of the languages evaluated on the validation set. This configuration achieved the best overall system performance and ranked 29th among the 46 official participating submissions in the *Accuracy Profile*. As shown in Table 5, the best performance was obtained on the German documents, achieving an *Impresso Profile Score* of 0.5071. English ranked second with a score of 0.4808, while French obtained the lowest value of 0.4648. Although differences exist among the three languages, the results show a relatively consistent behavior of the model, demonstrating the ability of the architecture to generalize across multilingual historical documents with different linguistic characteristics.

Likewise, Table 6 presents the evaluation corresponding to the *Efficiency Profile*. The results show that the system maintained stable behavior across the three evaluated languages, obtaining a *Mean Efficiency Profile Rank* of 15.6667 for German, 17.0000 for English, and 16.6667 for French. Consistent with the *Accuracy Profile* results, German achieved the best *Mean Impresso Profile Score* of 0.5071, followed by English with 0.4808 and French with 0.4648. These results demonstrate that the proposed architecture achieved an appropriate balance between predictive performance and computational efficiency, showing that the incorporation of specialized modules, such as NER-based feature extraction and temporal refinement, improves the model's historical reasoning capability without significantly increasing its complexity.

Table 5
Language-specific Accuracy Profile evaluation results of VerbaNexAI I on the validation data.

Language	Impresso Profile Score	At Macro Recall	At Accuracy	IsAt Macro Recall	IsAt Accuracy
German	0.5071	0.4220	0.4496	0.5923	0.6555
English	0.4808	0.3700	0.4506	0.5916	0.6173
French	0.4648	0.4332	0.5617	0.4965	0.6610

Table 6
Language-specific Efficiency Profile evaluation results of VerbaNexAI I on the validation data.

Language	Mean Efficiency Profile Rank	Rank Impresso Profile Score	Mean Impresso Profile Score
German	15.6667	26	0.5071
English	17.0000	28	0.4808
French	16.6667	29	0.4648

7. Conclusion

This work presented a multilingual multitask architecture for person-place temporal relation extraction within the context of the HIPE-2026 shared task. The proposed approach integrated contextual representations based on *XLM-RoBERTa-base*, features derived from Named Entity Recognition (NER), hyperparameter optimization with Optuna, balancing strategies, and a lightweight *Temporal Transformer* module for temporal and semantic refinement. The results demonstrated that the progressive incorporation of these components significantly improved system performance on the *at* and *isAt* tasks, particularly in scenarios requiring contextual and temporal reasoning over multilingual historical documents affected by OCR noise and linguistic variability.

The experimental analysis evidenced that temporal refinement mechanisms contributed positively to both the precision and generalization capability of the system, particularly in the *isAt* relation, where the immediate temporal context plays a fundamental role. Likewise, the experiments showed that balancing strategies affect each task differently, confirming the importance of combining weighted sampling techniques and class weights in multitask temporal relation extraction scenarios.

Additionally, all experimental configurations evaluated in this work were trained for 10 epochs, maintaining homogeneous training conditions to compare the different architectural variants and to balance strategies properly. This configuration enabled consistent analysis of the progressive contributions of NER, Optuna, temporal refinement, and balancing mechanisms to overall system performance. The proposed approach demonstrates the potential of integrating multilingual Transformers with lightweight temporal reasoning mechanisms to address complex historical relation extraction tasks within the context of digital humanities and historical knowledge graph construction.

Furthermore, the official competition results showed strong performance by the *VerbaNexAI I* team in the HIPE-2026 scenario. In the official precision profile, the best system run achieved position 29 out of 46 official submissions, while in the generalization profile, it obtained position 36. Likewise, in the combined precision and efficiency profile, *VerbaNexAI I – Run 2* ranked 9th, standing out for maintaining a competitive balance between performance and computational cost while using a considerably smaller architecture than several higher-ranked models. These results show that integrating different mechanisms enabled the development of an efficient and robust system for large-scale architectures containing billions of parameters. As future work, it would be relevant to explore larger-scale multilingual architectures and retrieval-based approaches. Likewise, incorporating document-level temporal reasoning mechanisms and implementing additional techniques could improve the identification of implicit person-place relations in complex multilingual historical collections.

Declaration on Generative AI

During the preparation of this investigation, ChatGPT (OpenAI) was used for the revision of translations into English, as well as for grammatical and spelling correction. After using this tool, the content was reviewed and edited as necessary, and full responsibility for the content of the publication is assumed.

Acknowledgment

The authors express their gratitude to the Call 933 “Training in National Doctorates with a Territorial, Ethnic and Gender Focus in the Framework of the Mission Policy — 2023” of the Ministry of Science, Technology and Innovation (Minciencia). In addition, we thank the team of the Artificial Intelligence Laboratory VerbaNex [2], affiliated with the UTB, for their contributions to this project.

References

[1] D. Wang, X. Tong, C. Dai, C. Guo, Y. Lei, C. Qiu, H. Li, Y. Sun, Voxel modeling and association of ubiquitous spatiotemporal information in natural language texts, International Journal of Digital Earth 16 (2023) 868–890. URL: https://doi.org/10.1080/17538947.2023.2185692. doi:10.1080/17538947.2023.2185692. arXiv:https://doi.org/10.1080/17538947.2023.2185692.

[2] J. Opitz, C. Raclé, E. Boros, A. Michail, M. Romanello, M. Ehrmann, S. Clematide, Clef hipe-2026: Evaluating accurate and efficient person-place relation extraction from multilingual historical texts, in: Advances in Information Retrieval, Springer Nature Switzerland, Cham, 2026, pp. 354–363. doi:10.1007/978-3-032-21321-1_46.

[3] 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, CEUR-WS, 2026. doi:10.5281/zenodo.20344461.

[4] J. Opitz, M. Ehrmann, C. Raclé, A. Michail, M. Romanello, 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] S. Zhao, L. Li, Temporal information extraction with the scalable cross-sentence context for electronic health records, Journal of Biomedical Informatics 128 (2022) 104052. URL: https://www.sciencedirect.com/science/article/pii/S1532046422000685. doi:https://doi.org/10.1016/j.jbi.2022.104052.

[6] S. Wang, R. Li, H. Wu, Integrating machine learning with linguistic features: A universal method for extraction and normalization of temporal expressions in chinese texts, Computer Methods and Programs in Biomedicine 233 (2023) 107474. URL: https://www.sciencedirect.com/science/article/pii/S0169260723001402. doi:https://doi.org/10.1016/j.cmpb.2023.107474.

[7] L. Zhuang, H. Fei, P. Hu, Syntax-based dynamic latent graph for event relation extraction, Information Processing & Management 60 (2023) 103469. URL: https://www.sciencedirect.com/science/article/pii/S0306457323002066. doi:https://doi.org/10.1016/j.ipm.2023.103469.

[8] C. Yuan, Q. Xie, S. Ananiadou, Temporal relation extraction with contrastive prototypical sampling, Knowledge-Based Systems 286 (2024) 111410. URL: https://www.sciencedirect.com/science/article/pii/S0950705124000455. doi:https://doi.org/10.1016/j.knosys.2024.111410.

[9] J. Yang, Y. Zhao, L. Yang, X. Wang, L. Chen, F.-Y. Wang, Temprompt: Multi-task prompt learning for temporal relation extraction in rag-based crowdsourcing systems, Neurocomputing 656 (2025) 131494. URL: https://www.sciencedirect.com/science/article/pii/S0925231225021666. doi:https://doi.org/10.1016/j.neucom.2025.131494.

[10] V. Torri, M. Ercolanoni, F. Bortolan, O. Leoni, F. Ieva, A nlp-based semi-automatic identification system for delays in follow-up examinations: An italian case study on clinical referrals, BMC Medical Informatics and Decision Making 24 (2024) 107. URL: https://doi.org/10.1186/s12911-024-02506-2. doi:10.1186/s12911-024-02506-2.

[2] https://github.com/VerbaNexAI

[11] N. An, F. Gui, L. Jin, H. Ming, J. Yang, Toward better understanding older adults: A biography brief timeline extraction approach, International Journal of Human–Computer Interaction 39 (2023) 1084–1095. URL: https://doi.org/10.1080/10447318.2022.2077278. doi:10.1080/10447318.2022.2077278. arXiv:https://doi.org/10.1080/10447318.2022.2077278.

[12] M. Ehrmann, G. Colavizza, Y. Rochat, F. Kaplan, Diachronic evaluation of ner systems on old newspapers, in: Proceedings of the 13th Conference on Natural Language Processing (KONVENS 2016), Bochum, 2016, pp. 97–107.

[13] A. Hamdi, et al., A multilingual dataset for named entity recognition, entity linking and stance detection in historical newspapers, in: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, New York, NY, USA, 2021, pp. 2328–2334.

[14] S. Menzel, et al., Named entity linking mit wikidata und gnd – das potenzial handkuratierter und strukturierter datenquellen für die semantische anreicherung von volltexten, in: Named Entity Linking mit Wikidata und GND – Das Potenzial handkuratierter und strukturierter Datenquellen für die semantische Anreicherung von Volltexten, De Gruyter Saur, 2021, pp. 229–258. doi:10.1515/9783110691597-012.

[15] S. Conia, M. Li, R. Navigli, S. Potdar, Semeval-2025 task 2: Entity-aware machine translation, in: Proceedings of the 19th International Workshop on Semantic Evaluation (SemEval-2025), 2025, pp. 2535–2557.

[16] D. P. Gnecco, J. Serrano, E. Puertas, J. C. Martinez-Santos, Hybrid re-ranking for biomedical entity linking using sapbert embeddings: a high-performance system for bionne-l 2025-1, 2025.

[17] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, V. Stoyanov, Unsupervised cross-lingual representation learning at scale, CoRR abs/1911.02116 (2019). URL: http://arxiv.org/abs/1911.02116. arXiv:1911.02116.