Extracting Spatial Entities Involved in the Description of a Movement Action Using Deep Learning Methods: A Comparative Study of Three Models

Abdelkrim Tafer[1,2], Mauro Gaio[1,*]

[1]University of Pau and the Adour Region, Laboratory of Mathematics and Their Applications, Pau, France
[2]University of Zaragoza, Aragón Institute for Engineering Research, Advanced Information Systems Laboratory, Zaragoza, Spain

Abstract

This paper proposes a methodology to automatically extract spatial information from itinerary descriptions in French. We compare three models: BiLSTM-CRF, CamemBERT, and GLINER, focusing on the recognition of nested spatial entities, motion verbs, spatial relations and spatial condition, and measures. Preliminary results demonstrate the potential of these models in accurately identifying and classifying spatial elements necessary for the annotation of movement actions evoked in textual descriptions.

Keywords

automatique annotation, classification, deep learning, nested spatial named entities

1. Introduction

The aim of this paper is to present the first results obtained in the feasibility study of deep learning methods for annotating and categorising geospatial expressions in narrative texts. In particular, in the case of descriptions of the various stages to be completed during a hike. A fundamental step in this process is the identification and labeling of spatial entities, offsets, and movement actions embedded within these texts. This task partially aligns with the well-established Named Entity Recognition (NER) problem [1], which aims to detect and classify specific entities in a text such as person, organization and location (i.e. spatial entity core). However, standard NERs need to be adapted to meet our fundamental need to extend the category of location. Firstly, by including the extraction of *weak spatial entities*. More generally, and according to [2], it is appropriate to differentiate between two categories of named entities, *strong named entities* and *weak named entities*. Secondly, the objective set also requires us to integrate, beyond the basic entities (weak or strong), a more complex category emerges by combining multiple basic entities, the *Nested Named Entity* (NNE) [3]. In addition to NNE, actions and relations or conditions are fundamental in interpreting text descriptions of the various stages of a route.

2. Related Work

Probabilistic models such as Conditional Random Fields (CRF) [4] have been widely used for structured sequence prediction tasks. When combined with recurrent neural networks such as Long Short-Term Memory (LSTM) networks [5], these models effectively capture local and contextual dependencies while improving the accuracy of named entity recognition (NER) [6].

Transformer-based language models [7] have significantly advanced the modeling of linguistic structures through large-scale pre-training on extensive text corpora. Among these, **BERT** [9] introduced a bidirectional transformer architecture that substantially improved performance across various NLP tasks and can be further specialized for NER through targeted fine-tuning. Additionally, newer approaches such as **GLiNER** [10] exploit pre-trained language models as backbone, such as Deberta v3

GeoExT 2025: Third International Workshop on Geographic Information Extraction from Texts at ECIR 2025, April 10, 2025, Lucca, Italy
*Corresponding author.
atafer@univ-pau.fr (A. Tafer); mauro.gaio@univ-pau.fr (M. Gaio)

[11] in the original paper, to develop low-resource NER systems that require minimal or no fine-tuning with state-of-the-art performance in zero-shot learning NER. Transformers have also been adapted for domain-specific applications, such as place name extraction from unstructured text [8].

Although recurrent neural networks offer moderate computational efficiency, their inherently sequential training and inference can limit parallelization and make it difficult to capture long-range dependencies. In contrast, transformer-based architectures leverage self-attention to process entire sequences in parallel, facilitating more effective modeling of distant context and exploiting modern GPU resources efficiently. However, transformers can become computationally demanding for very long inputs, as the self-attention mechanism scales quadratically with sequence length.

For these models, a labeled training corpus is required. Texts are first tokenized into word or subword units using various approaches, after that they are transformed into numerical vector representations [12, 13]. A classification layer is then applied to predict the token labels.

3. Method

In location category, strong named entity, hereafter simply called Named Entities (NE), is built from a toponym (i.e. a proper name, such as in Figure 1: "Saint-Ybars", "Porte de Mazet"). As weak spatial named entity is built from noun phrase describing the feature of the object to be referenced such as building, river, or path (e.g., "medieval street", "church tower"); for ease of reference, it is henceforth termed *nominal entity* (NoE). As mentioned earlier, the combination of the first two categories of spatial entities make up the category of spatial *Nested Named Entity* (NNE). For instance, in Figure 1, the phrase "hôtel de ville de Saint-Ybars" ('Saint-Ybars city hall') exemplifies an NNE, where the NoE "hôtel de ville" functions as the feature and the NE "Saint-Ybars"; the same applies to the NNE "le chocher de l'église" ('the church bell tower'), where the first NoE "clocher" acts as a feature for the second NoE "église".

In addition to NNE, movement verbs or movement verbal phrases such as in Figure 1: "traversez" ('cross') or "tourner pour descendre" ('turn down') delineate a moving action. Finally expressions like "à gauche" ('left'), "au bout" ('at the end of'), "à côté" ('next to'), and/or "200 m" provide fine grain spatial context, these expressions while be called hereafter *Offsets or Measures*.

Figure 1: Excerpt from a hike around Saint-Ybars, with annotations of key spatial information.

Rule-based approaches such as the PERDIDO system [14] have traditionally been employed for structured spatial tagging by combining morpho-syntactic and semantic constraints. Although effective for predefined structures, these methods are inherently limited in adaptability, often failing to detect variations in nominal entities and their relationships. This rigidity underscores the necessity for more flexible methodologies capable of dynamically learning entity representations and dependencies.

To address these challenges, deep-learning-based approaches offer a promising alternative. These models, trained on annotated corpora, exhibit strong generalization capabilities, allowing them to classify and extract spatial entities even in previously unseen contexts. Unlike rule-based systems,

deep learning models learn implicit representations of spatial languages and capture hierarchical dependencies and context-aware entity relationships.

The aim of this study is to evaluate three models for recognizing NNE and their contextual references. These models were selected based on their significance in Named Entity Recognition (NER) research, each representing a distinct approach to structured sequence prediction:

1. **Bidirectional Long Short-Term Memory with a Conditional Random Field Layer (BiLSTM-CRF)**: A well-established standard in NER using recurrent neural networks (RNNs).
2. **Pre-Trained Bidirectional Transformer (CamemBERT)**: A transformer-based bidirectional language model (BiLM) with a classification head for token labeling.
3. **Generalist Named Entity Recognition Using Bidirectional Transformers (GLiNER)**: An innovative zero-shot and few-shot learning model introducing a new paradigm for NER.

Each selected model represents a different paradigm in NER, providing a comparative analysis of their performance on structured sequence prediction tasks.

BiLSTM-CRF This model [6, 5, 4] is a widely adopted architecture for Named Entity Recognition (NER) and structured sequence labeling. It integrates a BiLSTM network with a CRF layer to efficiently capture contextual dependencies while enforcing valid label transitions.

The BiLSTM component processes input sequences in both forward and backward directions. Given a sequence of tokens $\mathbf{x} = \{x_1, x_2, \ldots, x_n\}$, two LSTM networks generate forward hidden states $\overrightarrow{h_t}$ and backward hidden states $\overleftarrow{h_t}$ for each token. The final representation is obtained by concatenating these states, yielding $h_t = \left[\overrightarrow{h_t}; \overleftarrow{h_t}\right]$ with a total hidden state dimension d. This bidirectional encoding allows the model to incorporate context from both past and future tokens.

A dense layer projects each hidden representation h_t into a score vector $s_t(y) \in \mathbb{R}^L$, where L is the number of possible labels. Instead of predicting labels independently, the CRF layer models dependencies between adjacent labels. The probability of a label sequence $\mathbf{y} = \{y_1, ..., y_n\}$ is defined as:

$$P(\mathbf{y}|\mathbf{x}) = \frac{1}{Z(\mathbf{x})} \prod_{t=1}^{n} \exp(A_{y_{t-1},y_t} + s_t(y_t)). \tag{1}$$

where A_{y_{t-1},y_t} is the transition score from label y_{t-1} to y_t, and $s_t(y_t)$ is the BiLSTM emission score at position t. The partition function $Z(\mathbf{x})$ normalizes over all possible sequences:

$$Z(\mathbf{x}) = \sum_{\mathbf{y}' \in \mathcal{Y}(\mathbf{x})} \prod_{t=1}^{n} \exp(A_{y'_{t-1},y'_t} + s_t(y'_t)). \tag{2}$$

The model is optimized by minimizing the negative log-likelihood loss:

$$\mathcal{L} = -\sum_{t=1}^{n} (A_{y_{t-1},y_t} + s_t(y_t)) + \log Z(\mathbf{x}). \tag{3}$$

During inference, the CRF layer selects the most probable label sequence by considering both emission scores from the BiLSTM and transition scores from the CRF. Figure 2 presents an overview of the model architecture.

CamemBERT This model [15] is a transformer-based model designed specifically for the French language. Unlike BiLSTM-CRF, which processes sequences token by token, CamemBERT employs self-attention mechanisms that allow all tokens in a sequence to be processed in parallel, capturing long-range dependencies more efficiently.

Given an input sequence $\mathbf{x} = \{x_1, x_2, ..., x_n\}$, CamemBERT encodes each token using multiple transformer layers. At the core of its architecture is the self-attention mechanism, which computes

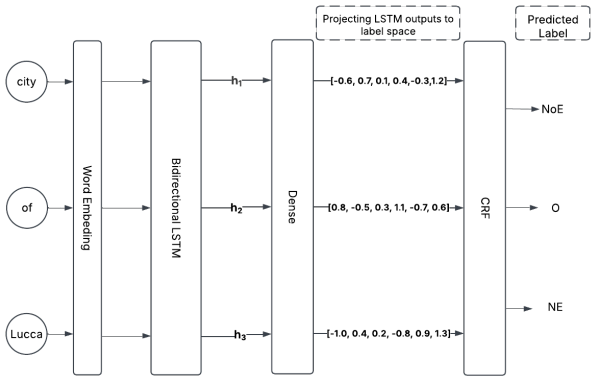

Figure 2: Schematic overview of the BiLSTM-CRF architecture for NER. Input tokens are first converted into word embeddings, then processed by a bidirectional LSTM to capture contextual information in both forward and backward directions. The resulting hidden states are projected into a label space through a dense layer, and finally, a CRF layer enforces valid label transitions during decoding.

contextualized representations by attending to all tokens in the sequence. The attention score between token i and token j is computed as 4 and the output representation is then obtained as 5:

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{n} \exp(e_{ik})}, \quad e_{ij} = \frac{\mathbf{q}_i \cdot \mathbf{k}_j}{\sqrt{d}} \tag{4}$$

$$h'_i = \sum_{j=1}^{n} \alpha_{ij} \mathbf{v}_j, \quad \mathbf{v}_j = W_v h_j \tag{5}$$

where $\mathbf{q}_i = W_q h_i$, $\mathbf{k}_j = W_k h_j$, and d is the head dimension, and W_q, W_k, and W_v are learnable projection matrices.

Unlike BiLSTM, which encodes sequential dependencies using recurrence, self-attention allows each token to directly incorporate information from all other tokens in a single operation.

For Named Entity Recognition (NER), CamemBERT employs a classification head that assigns labels to tokens. A dense layer maps the final hidden representation into logits $\mathbf{z}_t \in \mathbb{R}^L$, where L is the number of entity labels.

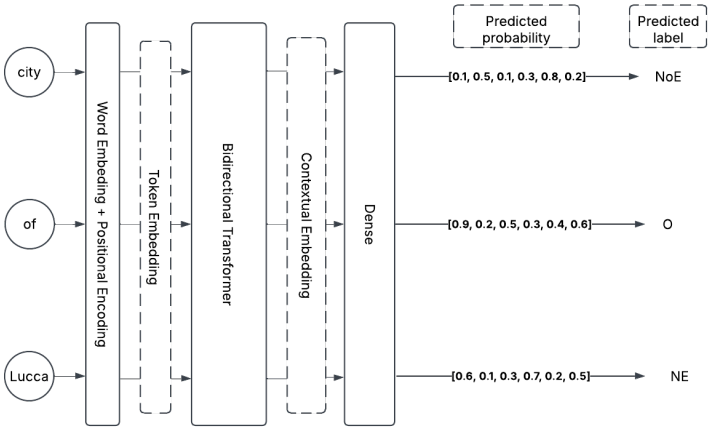

Figure 3: High-level overview of CamemBERT applied to token classification. Each token is first embedded, then fed into multiple transformer layers employing self-attention to capture contextual information across the entire sequence. A dense layer projects the contextual embeddings to produce token-level class probabilities, enabling the model to assign named entity labels.

The model is trained using the cross-entropy loss. Compared to BiLSTM-CRF, which explicitly models

label dependencies via a CRF layer, CamemBERT implicitly learns contextual relationships through self-attention. Figure 3 presents the overview of model architecture.

GLiNER This third and last model [10] is a transformer-based NER model that introduces span-based classification with zero-shot learning capabilities. By modeling spans instead of tokens, it allows more flexible boundary detection and can better handle nested structures. The token encoder processes a unified input consisting of both entity type tokens and the input text, generating contextualized representations. Let $\mathbf{p} = \{p_i\}_{i=0}^{M-1} \in \mathbb{R}^{M \times D}$ denote the entity type representations, where M is the number of entity types and D is the dimensionality of each representation. Similarly, let $\mathbf{h} = \{h_i\}_{i=0}^{N-1} \in \mathbb{R}^{N \times D}$ represent the contextual embeddings for each token in the input text, with N being the number of tokens. The entity representations are refined through a two-layer feedforward network, producing $\mathbf{q} = \{q_i\}_{i=0}^{M-1} \in \mathbb{R}^{M \times D}$.

The representation of a span from position i to j is computed as $\mathbf{S}_{ij} = \text{FFN}(h_i \otimes h_j)$, where $\otimes$ denotes concatenation. To determine whether a span (i, j) corresponds to entity type t, a matching score is computed as:

$$\phi(i, j, t) = \sigma\left(\mathbf{S}_{ij}^{\top} q_t\right), \tag{6}$$

where σ is the sigmoid activation function. This score represents the probability that the span (i, j) belongs to entity type t.

During training, the model distinguishes between positive pairs (spans correctly labeled with type t) and negative pairs (incorrect associations) using a binary cross-entropy loss:

$$\mathcal{L}_{BCE} = -\sum_{s \in \mathcal{S} \times T} \left[\mathbb{I}_{s \in \mathcal{P}} \log \phi(s) + \mathbb{I}_{s \in \mathcal{N}} \log\left(1 - \phi(s)\right)\right], \tag{7}$$

where $\mathbb{I}$ is the indicator function. This loss encourages high matching scores for correct span-type pairs while penalizing incorrect associations.

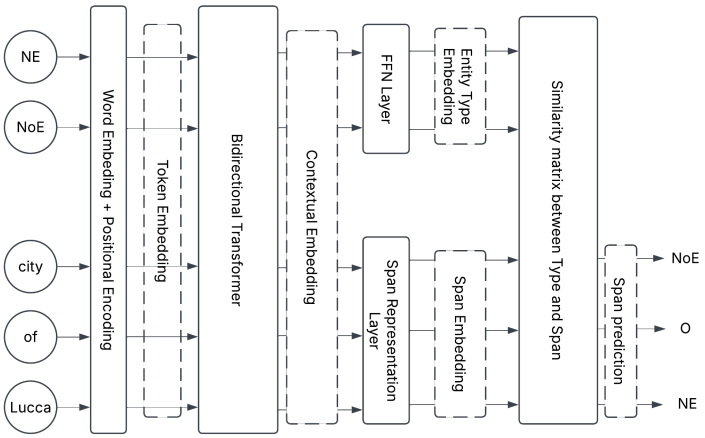

Figure 4: High-level depiction of the GLiNER architecture, illustrating its span-based approach to Named Entity Recognition. A unified input of both entity type tokens (e.g., NE, NoE) and the text is encoded by a bidirectional transformer to produce contextual embeddings. A feedforward network then derives embeddings for entity types and text spans, and a similarity matrix identifies which spans match each entity type, enabling flexible boundary detection and nested entity handling.

GLiNER differs fundamentally from BiLSTM-CRF, which explicitly models sequence dependencies via a CRF layer, and CamemBERT, which performs token-level classification. By employing span-based prediction and textual entailment-style classification, GLiNER enhances generalization across domains and under certain conditions, it enables entity recognition in low-resource and zero-shot settings. Figure 4 presents an overview of the model architecture.

By comparing these approaches, this study provides insights into the effectiveness of different NER paradigms in extracting spatial movement actions from descriptive texts.

4. Experiments

An initial pilot study aimed to assess the performance of these 3 models[1] in accurately annotating text segments with six predefined labels.

Training Dataset and Annotation Process

The dataset[2] consists of 1,897 french hiking descriptions, totaling 27,083 sentences and 569,214 tokens. Spatial expressions are categorized using the annotation labels given in Figure 1: strong named entities **Named Entities (NE)**, weak named entities **Nominal Entities (NoE)**, motion verbs or verbal phrases **(Motion)**, **expressions evoking spatial relation or condition (Offset)**, and **numerical expressions followed by a unit of measurement (Measure)** and finally, **Nested Named Entities (NNE)**. These labels are inspired by previous rule-based approaches [16].

It is well known that producing an annotated dataset is a cumbersome and time-consuming task. It was therefore decided to use Perdido [14] as the annotator for this first study. But Perdido was not designed to be able to annotate nominal entities directly, and it would be a real challenge to integrate it. It was decided that this annotation would go through two stages (Figure 5). Firstly, following the result of the annotation carried out by Perdido, all the words or phrases involved in the annotation of a spatial named entity and having received the part of speech label *"Noun"* were extracted. A dictionary was created from these words or phrases, which then enabled all occurrences of the lexical entries in this dictionary to be labelled in the dataset as **NoE**.

The result is a silver-standard corpus—potentially containing errors due to fully automated annotation.

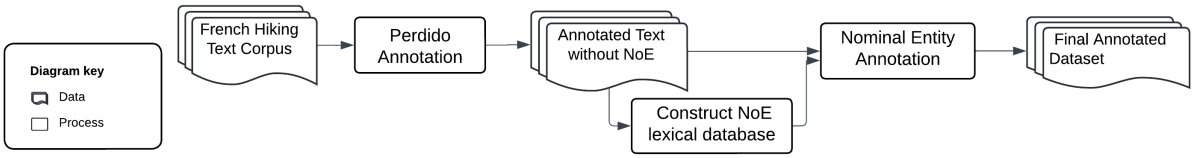

Figure 5: Training Dataset Annotation Flow

All models were trained and tested on an identical dataset extracted from the *silver-standard* corpus. Evaluation metrics include Precision, Recall, and micro F1-score, as summarized in Table 1.

Tokenization Tokenization is a crucial preprocessing step that can significantly affects model performance. In our experiments, each model uses a distinct strategy. The BiLSTM-CRF model employs rule-based, word-level tokenization with TreeTagger [17], configured for French. Camembert-base uses subword tokenization based on Byte Pair Encoding (BPE) [18, 19] as implemented by SentencePiece [20] to decompose rare and compound words. GLiNER, which leverages a multilingual DeBERTa backbone, adopts a unigram-based subword tokenization strategy [21] via SentencePiece.

Models Parameters

For each model, the following parameter settings were used without applying additional hyperparameter tuning techniques:

[1]Model implementation: https://git.univ-pau.fr/atafer/sner
[2]Dataset: https://git.univ-pau.fr/atafer/hiking-dataset

BiLSTM-CRF The BiLSTM-CRF model employs two LSTM cells (one for the forward and one for the backward direction) with an embedding size of 300 and a hidden dimension of 512 (256 per cell). The model is trained using a learning rate of 0.001.

Camembert-base Camembert-base is configured with an embedding/hidden size of 768, utilizes 12 transformer layers, and is trained with a learning rate of 2×10^{-5}.

GLiNER GLiNER utilizes the mDeBERTa-v3-large backbone—a multilingual variant of DeBERTa-v3—with an embedding/hidden size of 1024 and 12 transformer layers. The model is optimized using a learning rate of 5×10^{-6}.

Overall Analysis

Camembert-base achieved the highest overall performance with an F1-score of 0.9534, followed closely by GLiNER with an F1-score of 0.9355. The superior performance of Camembert-base Could perhaps be explained by its pre-training on French-language data [15], which enhances its ability to capture fine linguistic nuances inherent in the corpus. In contrast, GLiNER employs a backbone pretrained on the CC100 a multilingual corpus [22] where French comprises only about 3% of the tokens; this may partially explain its slightly lower performance on French-language data.

Interestingly, despite the absence of a dedicated pre-trained language model, the BiLSTM-CRF model effectively captured the specific characteristics of the hiking description corpus, achieving an F1-score of 0.9269 while maintaining a moderate number of parameters and lower computational cost.

Table 1
Overall model performance on the NER task.

Model	Precision	Recall	F1-score
BiLSTM-CRF	0.9425	0.9118	0.9269
Camembert-base	**0.9464**	**0.9605**	**0.9534**
GLiNER	0.9282	0.9428	0.9355

Label-Level Analysis

Table 2 indicates that the transformer-based Camembert-base model consistently outperforms the BiLSTM-CRF and GLiNER models across all six categories, particularly for NE. The BiLSTM-CRF model, however, performs poorly on NE due to its low recall. We hypothesize two main factors underlying this result. First, the model's reliance on word-level tokenization may exacerbate out-of-vocabulary (OOV) issues, especially for toponyms or alphanumeric place names that do not appear in the training vocabulary. For example, in one instance, "D218D" was misclassified, presumably because it did not match any known token from the training set. Second, the silver-standard dataset itself may contain annotation inconsistencies or errors, which can propagate into all models' outputs but disproportionately affect a model that already struggles with OOV tokens.

Table 2
Token-Level Model Performance for Named Entity Recognition (NER).

Label	BiLSTM-CRF			Camembert			GLiNER		
	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
MEASURE	0.9586	0.9980	0.9779	0.9856	0.9982	**0.9919**	0.8041	0.9333	0.8639
MOTION	0.9933	0.9901	0.9917	0.9906	0.9975	**0.9940**	0.9882	0.9972	0.9927
NE	0.7987	0.4683	0.5904	0.8320	0.8702	**0.8507**	0.7893	0.7571	0.7729
NNE	0.8350	0.8188	0.8268	0.9039	0.9043	**0.9041**	0.8502	0.8862	0.8678
NoE	0.9787	0.9847	0.9817	0.9809	0.9900	**0.9854**	0.9705	0.9826	0.9765
OFFSET	0.9649	0.9504	0.9577	0.9549	0.9825	**0.9685**	0.9378	0.9746	0.9559

Model Memory Footprint, Parameter Count, and Efficiency

Despite its compact architecture of approximately 8.73 million parameters and a minimal GPU memory allocation of 53.82 MB along with only 8.98 MB CPU memory during inference, the BiLSTM-CRF model demonstrates competitive performance relative to more complex transformer-based models. In contrast, CamemBERT-base, with 110.05 million parameters, requires substantially greater computational resources (430.07 MB allocated on the GPU and 324.73 MB on the CPU), achieving enhanced performance through richer language representations. The GLiNER model, which leverages a large multilingual DeBERTa backbone, comprises approximately 288.95 million parameters and incurs the highest memory demands (2206.75 MB allocated on the GPU and 1709.67 MB on the CPU).

These results highlight that small, specialized architectures such as BiLSTM-CRF can yield near-comparable performance with significantly lower memory and parameter footprints, making them particularly advantageous for deployment in resource-constrained settings, while the choice of a larger model backbone in GLiNER underlines the trade-off between resource investment and the potential for improved cross-lingual generalization. In addition, although our evaluation does not formally assess cross-lingual transfer performance, preliminary examples in English suggest that GLiNER's multilingual pretraining enables effective transfer of learned representations in french to other languages. Moreover, the GLiNER framework is inherently modular, allowing for the replacement of its resource-intensive multilingual DeBERTa backbone with alternatives such as CamemBERT, which offers a lower memory footprint. This flexibility provides a promising avenue for optimizing the balance between computational efficiency and performance in Named Entity Recognition tasks.

5. Conclusion and Perspectives

In this study, we compared three deep learning models—our specialized BiLSTM-CRF model, CamemBERT-base, and GLiNER—for the extraction of spatial entities (nested or not, strong or weak) and movement actions from French itinerary descriptions. The experimental results indicate that transformer-based models, such as CamemBERT, effectively capture complex spatial patterns, while our specialized BiLSTM-CRF model, designed specifically for this task, offers a competitive alternative with substantially lower computational requirements. The efficiency of the BiLSTM-CRF model makes it well suited for resource-constrained environments, and incorporating subword tokenization could further enhance its ability to handle out-of-vocabulary terms—an issue highlighted by the misclassification of certain named entities.

The GLiNER model, which utilizes a large multilingual DeBERTa backbone, was not subjected to a detailed cross-lingual transfer analysis; however, its design suggests that multilingual pretraining may support transferring representations learned on French data to other languages. Moreover, its modular architecture permits the substitution of its resource-intensive backbone with alternatives such as CamemBERT, potentially reducing memory usage while hoping to maintain good performance.

Future work will focus on several key directions. First, the development of a gold-standard corpus (especially for the test dataset) with manually corrected annotations is essential to overcome the limitations of our current silver-standard dataset and to provide a more reliable benchmark. Second, integrating higher-level structural annotations—particularly syntax-semantic dependencies linking spatial entities with their contextual elements—could refine the extraction process. Lastly, we will continue to investigate and refine model architectures to optimize the automated extraction and categorization of spatial entities and movement actions from descriptive texts.

Declaration on Generative AI

During the preparation of this work, the authors used GPT-4 to Grammar and spelling check. After using these tool, the author reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] R. Grishman, B. Sundheim, Message understanding conference-6: a brief history, in: Proceedings of the 16th Conference on Computational Linguistics - Volume 1, COLING '96, Association for Computational Linguistics, USA, 1996, p. 466–471. URL: https://doi.org/10.3115/992628.992709. doi:10.3115/992628.992709.

[2] M. R. Vicente, La glose comme outil de désambiguïsation référentielle des noms propres purs, Corela. Cognition, représentation, langage (2005). URL: http://journals.openedition.org/corela/1212. doi:10.4000/corela.1212.

[3] J. R. Finkel, C. D. Manning, Nested named entity recognition, in: P. Koehn, R. Mihalcea (Eds.), Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Singapore, 2009, pp. 141–150. URL: https://aclanthology.org/D09-1015/.

[4] N. Patil, A. Patil, B. Pawar, Named entity recognition using conditional random fields, Procedia Computer Science 167 (2020) 1181–1188. URL: https://www.sciencedirect.com/science/article/pii/S1877050920308978. doi:https://doi.org/10.1016/j.procs.2020.03.431, international Conference on Computational Intelligence and Data Science.

[5] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (1997) 1735–1780. URL: https://doi.org/10.1162/neco.1997.9.8.1735. doi:10.1162/neco.1997.9.8.1735.

[6] G. Lample, M. Ballesteros, S. Subramanian, K. Kawakami, C. Dyer, Neural architectures for named entity recognition, in: K. Knight, A. Nenkova, O. Rambow (Eds.), Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Association for Computational Linguistics, San Diego, California, 2016, pp. 260–270. URL: https://aclanthology.org/N16-1030/. doi:10.18653/v1/N16-1030.

[7] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, in: Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17, Curran Associates Inc., Red Hook, NY, USA, 2017, p. 6000–6010.

[8] C. Berragan, A. Singleton, A. Calafiore, J. M. and, Transformer based named entity recognition for place name extraction from unstructured text, International Journal of Geographical Information Science 37 (2023) 747–766. URL: https://doi.org/10.1080/13658816.2022.2133125. doi:10.1080/13658816.2022.2133125. arXiv:https://doi.org/10.1080/13658816.2022.2133125.

[9] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, in: J. Burstein, C. Doran, T. Solorio (Eds.), Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Association for Computational Linguistics, Minneapolis, Minnesota, 2019, pp. 4171–4186. URL: https://aclanthology.org/N19-1423. doi:10.18653/v1/N19-1423.

[10] U. Zaratiana, N. Tomeh, P. Holat, T. Charnois, GLiNER: Generalist model for named entity recognition using bidirectional transformer, in: K. Duh, H. Gomez, S. Bethard (Eds.), Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers), Association for Computational Linguistics, Mexico City, Mexico, 2024, pp. 5364–5376. URL: https://aclanthology.org/2024.naacl-long.300/. doi:10.18653/v1/2024.naacl-long.300.

[11] P. He, J. Gao, W. Chen, DeBERTaV3: Improving DeBERTa using ELECTRA-style pre-training with gradient-disentangled embedding sharing, 2023. URL: http://arxiv.org/abs/2111.09543. doi:10.48550/arXiv.2111.09543. arXiv:2111.09543 [cs].

[12] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, J. Dean, Distributed representations of words and phrases and their compositionality, in: R. Caruana, S. Lawrence, C. Giles (Eds.), Advances in Neural Information Processing Systems, volume 26, Curran Associates, Inc., 2013, pp. 3111–3119. Introduced the Skip-gram model and Negative Sampling, foundational for word embeddings.

[13] P. Bojanowski, E. Grave, A. Joulin, T. Mikolov, Enriching word vectors with subword information,

Transactions of the Association for Computational Linguistics 5 (2017) 135–146. URL: https://aclanthology.org/Q17-1010/. doi:10.1162/tacl_a_00051.

[14] L. Moncla, M. Gaio, Perdido: Python library for geoparsing and geocoding French texts, in: First International Workshop on Geographic Information Extraction from Texts (GeoExT), Dublin, Ireland, 2023. URL: https://hal.science/hal-04049794.

[15] L. Martin, B. Muller, P. J. Ortiz Suárez, Y. Dupont, L. Romary, É. de la Clergerie, D. Seddah, B. Sagot, CamemBERT: a tasty French language model, in: D. Jurafsky, J. Chai, N. Schluter, J. Tetreault (Eds.), Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Online, 2020, pp. 7203–7219. URL: https://aclanthology.org/2020.acl-main.645/. doi:10.18653/v1/2020.acl-main.645.

[16] M. Gaio, L. Moncla, Extended Named Entity Recognition Using Finite-State Transducers: An Application To Place Names, in: The Ninth International Conference on Advanced Geographic Information Systems, Applications, and Services (GEOProcessing 2017), Nice, France, 2017. URL: https://hal.science/hal-01492994.

[17] H. Schmid, Probabilistic part-of-speech tagging using decision trees, in: Proceedings of the International Conference on New Methods in Language Processing (NOLP 1994), 1994, pp. 44–49.

[18] P. Gage, A new algorithm for data compression, C Users J. 12 (1994) 23–38.

[19] R. Sennrich, B. Haddow, A. Birch, Neural machine translation of rare words with subword units, in: K. Erk, N. A. Smith (Eds.), Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Berlin, Germany, 2016, pp. 1715–1725. URL: https://aclanthology.org/P16-1162/. doi:10.18653/v1/P16-1162.

[20] T. Kudo, J. Richardson, SentencePiece: A simple and language independent subword tokenizer and detokenizer for neural text processing, in: E. Blanco, W. Lu (Eds.), Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, Association for Computational Linguistics, Brussels, Belgium, 2018, pp. 66–71. URL: https://aclanthology.org/D18-2012/. doi:10.18653/v1/D18-2012.

[21] T. Kudo, Subword regularization: Improving neural network translation models with multiple subword candidates, in: I. Gurevych, Y. Miyao (Eds.), Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Melbourne, Australia, 2018, pp. 66–75. URL: https://aclanthology.org/P18-1007/. doi:10.18653/v1/P18-1007.

[22] 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, in: D. Jurafsky, J. Chai, N. Schluter, J. Tetreault (Eds.), Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Online, 2020, pp. 8440–8451. URL: https://aclanthology.org/2020.acl-main.747/. doi:10.18653/v1/2020.acl-main.747.