A Hybrid Annotation Method and A Reference Corpus for Hiking Descriptions

Abdelkrim Tafer[1,2,*], Mauro Gaio[1] and Javier Lacasta[2]

[1]Université de Pau et des Pays de l'Adour, CNRS, LMAP, Pau, France
[2]Universidad de Zaragoza, I3A, Zaragoza, Spain

Abstract

Reference corpora for spatial language understanding remain scarce, particularly for procedural movement texts where navigation relies on generic landmarks rather than well-known toponyms. This contribution is an attempt to fill that gap. We present a 3,686-sentence corpus[1] derived from French hiking trail descriptions and annotated with a two-level scheme. Level 1 identifies atomic span types (NE, NOE, NNE, ACTION, OFFSET, MEASURE, and MISC), and Level 2 captures discourse functions (GEOMOTION, GEOVISION, GEODESC, and MISC). The corpus was built using an active learning workflow combining zero-shot *GLiNER* bootstrapping with manual correction in *Argilla*, augmented by 738 synthetic non-spatial sentences for negative sampling. Statistical analysis reveals that Nominal Entities (NOE) consistently outnumber Named Entities (NE) across all semantic contexts, with NEs rarely appearing without an accompanying NOE. This finding challenges the traditional NER focus on toponyms and highlights the importance of recognizing common spatial nouns for route reconstruction. Within a single sentence, Level 2 spans of different types can partially overlap, as the same tokens may simultaneously convey a motion instruction and an environment description.

Keywords

NER corpus, spatial entity recognition, hiking descriptions, nested named entities, active learning

1. Introduction

Real human movement can be expressed in various ways in natural language. More generally, motion is a fundamental linguistic concept; owing to its intrinsic properties, it has been the focus of both theoretical and practical research in linguistics and geomatics for several decades; they all agree that movement is a semantic foundation ([1], [2], [3]). It is therefore important to identify the various forms of expression of movement in the text and their characteristics. Consequently, much research has focused on this identification and the associated semantics, going so far as to propose formal coding schemas ([4], [5], [6], [7]). The core of these coding schemas lies on a triptych: action, landmark and spatial relationship. However, to the best of our knowledge, since these proposals were put forward, no major study has resulted in the creation of an annotated reference corpus based on such coding schemas and available to the community. In particular, in contexts where it is necessary to highlight the relationship between spatial activity and one or more locations.

We are attempting to address this shortcoming by developing a comprehensive process that greatly simplifies the creation of this kind of corpus. In order to illustrate this process with a concrete example, this paper describes the entire process, focusing on three specific aspects: (i) an intermediate annotation scheme, designed to group atomic elements to discourse-level semantic functions; (ii) a reference corpus based on hiking descriptions in French, one of the distinctive features of which is the predominance of spatial landmarks composed of nominal entities (NOE) rather than named entities (NE), thereby challenging the named entity recognition (NER) approach centred on place names; and (iii) a hybrid annotation method combining zero-shot *GLiNER* bootstrapping with iterative human correction and

[1]The dataset is available upon request to the corresponding author in JSON format with train/dev/test splits.

GeoExT 2026: Fourth International Workshop on Geographic Information Extraction from Texts at ECIR 2026, April 2, 2026, Delft, The Netherlands

*Corresponding author.
atafer@univ-pau.fr (A. Tafer); mauro.gaio@univ-pau.fr (M. Gaio); jlacasta@unizar.es (J. Lacasta)
0009-0000-9803-9648 (A. Tafer); 0000-0002-8041-4240 (M. Gaio); 0000-0003-3071-5819 (J. Lacasta)

synthetic negative sampling, designed to facilitate corpus extension to additional languages.

2. Related Work

Spatial annotation schemas. During the 2000s, a lack of annotation schemas was identified for encoding geographic information in natural language into precise formalised representations — geographically located spatial landmarks and qualitative spatial relationships (topological and orientational). Markup languages such as SpatialML [7], ISO-space [4] and TEI [5] were therefore proposed to mark up places in text, link them to geographical nomenclatures, and associate formal models with the spatial relationships described. These languages support entity encapsulation essential for modelling locations, but impose a single tree-based hierarchical structure which may be too restrictive for annotations at the discourse level. Some mature languages also support tags specific to the language of expression, such as POS (part of speech).

Their aim is to provide a formal framework for annotating spatial facts in text, enabling automatic interpretation or inventorying how spatial information is expressed in natural language. Let's focus on ISO_Space because designed for standardisation, provides a set of tags (PLACE, PATH, SPATIAL_ENTITY, MOTION, …) and relational links (MOVELINK, QSLINK, OLINK) covering both static and dynamic actions. ISO-space may therefore appear ideally suited to our objectives. We explain below why we chose to produce an intermediate annotation as an essential transitional step.

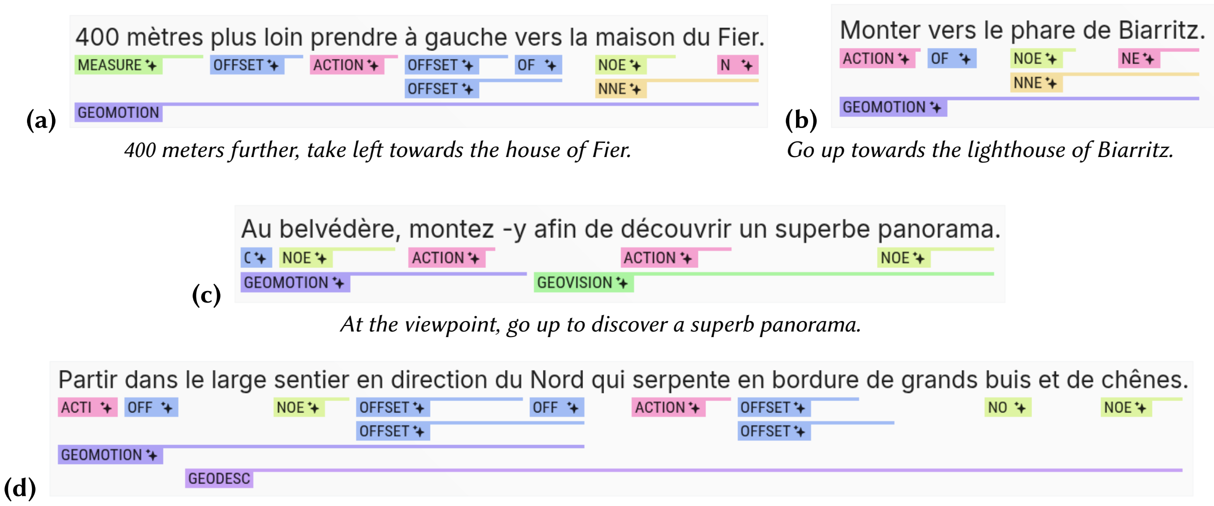
Geospatial corpora. Annotated resources for geographic entity recognition have primarily targeted news or historical text [8]. GeoEDdA [9] annotates 2,200 paragraphs from Diderot & d'Alembert's *Encyclopédie* with 12 geo-semantic span categories. In the outdoor domain, Kew et al. [10] annotate 8 fine-grained toponym types in a diachronic corpus of Swiss Alpine Club texts, while Rayson et al. [11] propose a deeply annotated testbed for the English Lake District with 19 geospatial categories. In the travelogue domain, Higashiyama et al. [12] annotate 200 Japanese travel documents (12,171 mentions) with named, nominal, and deictic geo-entity types, coreference relations, and links to OpenStreetMap. Notably, 48.4% of mentions are nominal or demonstrative, corroborating the dominance of non-named spatial references. Yamamoto et al. [13] extend this resource with graph-structured trajectory extraction, predicting visit status and visiting order. The Alpine and Lake District corpora focus on named entities without discourse-level functions, while the Arukikata dataset covers nominal mentions and coreference but does not model discourse-level semantic functions specific to procedural spatial text.

These corpora were each designed for a specific purpose, making repurposing difficult, and none has been annotated using the markup languages discussed above. These observations motivate our effort to streamline the process and reduce the time required to produce this kind of resource.

The implementation of our hybrid annotation method, which combines zero-shot NER with iterative human correction, required a form of annotation satisfying four constraints: (i) compatibility with both token-level (BERT-style [14]) and span-based (GLiNER-style [15]) NER input schemas; (ii) reduced cognitive load on the human annotator; (iii) explicit distinction between nominal and named entities; and (iv) a link between intra-phrasal atomic elements and discourse-level semantic functions, potentially inter-phrasal. The following section presents our intermediate annotation schema, designed to address these constraints.

3. An Intermediate Annotation Schema

The schema proposed here operates at two levels: Level 1 identifies spatial building blocks — Named Entities (NE), Nominal Entities (NOE), their nested combination (NNE), actions, offsets, and measures — whereas Level 2 groups them into semantic functions that characterise the action or actions in which the atomic elements are involved. The possibility that an atomic element may be involved in more than one action necessitates a more complex annotation schema, so that it can accommodate partial or complete overlap between spans. Figure 1 illustrates both levels. Note that this structural complexity



400 meters further, take left towards the house of Fier.

Go up towards the lighthouse of Biarritz.

At the viewpoint, go up to discover a superb panorama.

Leave on the wide path heading north, which winds along large boxwoods and oaks.

Figure 1: Annotated examples from the corpus showing both annotation levels. Examples (a–c) illustrate Level 1 entities grouped by Level 2 semantic functions. Example (d) shows partial overlap between Level 2 spans: the GEOMOTION and GEODESC spans share the tokens *le large sentier en direction du Nord*.

poses new challenges for existing models, as chosen NER architectures are not designed to handle partial overlaps between spans. This issue is addressed in Section 4.2.

The Level 1 entities share common ground with ISO-Space's entity inventory. In particular, NE and NOE correspond to typed PLACE elements and OFFSET to SPATIAL_SIGNALs, as for ACTION, the 'motion' subcategory has a direct correspondence with MOTION events. The Level 2 adds a discourse-function layer that groups these atomic entities into communicative units (motion instruction, visual cue, environment description), a dimension orthogonal to ISO-Space's pairwise relation links. Since ISO-Space's modular design allows partial adoption, a mapping between both schemes is feasible and is discussed as future work (Section 6).

Level 1: Atomic spatial entities.

- **NE** — Proper geographical names (toponyms), e.g., *Biarritz* and *Fier* (Figure 1b,a), or trail codes such as *GR10* (hiking trail reference) [16].
- **NOE** — Common nouns designating spatial objects, such as *maison (house)* and *phare (lighthouse)* (Figure 1a,b), or natural features like *col (mountain pass)* and *rivière (river)*. These "weak" spatial entities [16] form the majority of references in procedural texts [17].
- **NNE** — Compositional entities nesting two or more spatial sub-entities, most commonly an NOE with an NE (79.1%, e.g., *maison du Fier (house of Fier)* and *phare de Biarritz (lighthouse of Biarritz)* in Figure 1a,b), or two NOEs (18.7%, e.g., *allée de la station d'épuration (lane of the water treatment plant)*).
- **ACTION** — Verbs conveying actions directly related to the hike realization: motion (*prendre (take)*, *montez (go up)*, Figure 1a,c), visual perception (*découvrir (discover)*, Figure 1c), or states (*est (is)*).
- **OFFSET** — Terms evoking all or part of a spatial relation: cardinal direction (*au nord (to the north)*), egocentric (*à gauche (to the left)*, Figure 1a), prepositions (*vers (towards)*, Figure 1b), and qualifying adverbs (*longtemps (for a long time)*, *brièvement (briefly)*).
- **MEASURE** — Quantified distance or duration, e.g., *400 mètres* (Figure 1a), *pendant 20 mn* (*for 20 minutes*).
- **MISC** — MISC marks non-spatial content at both levels. At Level 1, it identifies non-spatial clauses in spatial sentences (e.g., *que l'on peut entendre (which can be heard)*) and, in synthetic negatives (Section 4), annotates verbs most likely confused with ACTION (e.g., *retrouver (rediscover)*).

Span boundary conventions. Determiners are included in NOE spans (e.g., *la maison (the house)*), while prepositions are typically annotated as OFFSET. NNE spans capture the two composition patterns described above (NOE+NE and NOE+NOE), where the nominal head specifies the type of the spatial referent.

Level 2: Semantic functions. Inspired by the distinction between motion events, perception, and static configurations [3], Level 2 groups Level 1 entities into four semantic functions:

- **GEOMOTION** — Motion instructions or itinerary steps, grouping an ACTION with spatial references. Examples (a–d) in Figure 1 illustrate GEOMOTION spans: an ACTION verb (*prendre (take)*, *Monter (go up)*, *montez (go up)*, *Partir (leave)*) combined with OFFSETs and spatial entities to form a complete motion instruction.
- **GEOVISION** — Visual or perceptual landmarks used as orientation cues, e.g., *afin de découvrir un superbe panorama (to discover a superb panorama)* in Figure 1c.
- **GEODESC** — Environment descriptions situating the walker without prescribing movement. In Figure 1d, the GEODESC span covers the description of the environment through which the walker moves.
- **MISC** — At Level 2, MISC groups Level 1 MISC spans and unannotated tokens into discourse-level non-spatial units; the 738 synthetic negatives are entirely covered by Level 2 MISC, providing 16.5% full negative examples.

Partial overlap between Level 2 spans. Level 2 spans may overlap: in Figure 1d, the tokens *le large sentier en direction du Nord* belong to **both** a GEOMOTION span (the motion instruction) and the GEODESC span (the environment description). This deliberate design choice allows each semantic function to be annotated as a self-contained communicative unit, avoiding linguistically complex decisions such as relative clause attachment (e.g., *qui serpente en bordure...*). This motivates the overlap-tolerant decoding described in Section 4.

4. Hybrid Annotation Method

The construction of a corpus follows five stages: (1) sourcing raw texts; (2) preprocessing via sentence segmentation, deduplication, and shuffling; (3) augmenting with synthetic non-spatial sentences generated by *GPT-OSS*; (4) annotating through an active-learning loop combining zero-shot *GLiNER* bootstrapping with manual correction in *Argilla* [18]; and (5) splitting into train/dev/test partitions. Figure 2 provides an overview of this process.

4.1. Our corpus composition

The core data consists of 2,948 sentences taken from raw trail description on the collaborative hiking platform *Visorando*[1]. Sentence segmentation was applied to multi-sentence trail descriptions, followed by random shuffling to prevent document-level leakage; duplicates were removed during preprocessing. We augmented this with 738 synthetic non-spatial sentences generated by *GPT-OSS* [19] following a two-pronged strategy. First, we compiled a lexicon of French motion verbs commonly found in hiking descriptions (e.g., *aborder (approach)*, *tourner (turn)*, *traverser (cross)*, *emprunter (take)*) and prompted the model to generate sentences that use these verbs in non-spatial contexts, due to their polysemy, figurative meaning, or metaphorical use — for instance, Example (1) repurposes the spatial verb *aborder (approach)* in its figurative meaning.

(1) *Aborder le sujet de la confiance nécessite une écoute sincère*
(Approaching the topic of trust requires sincere listening)

[1]https://www.visorando.com

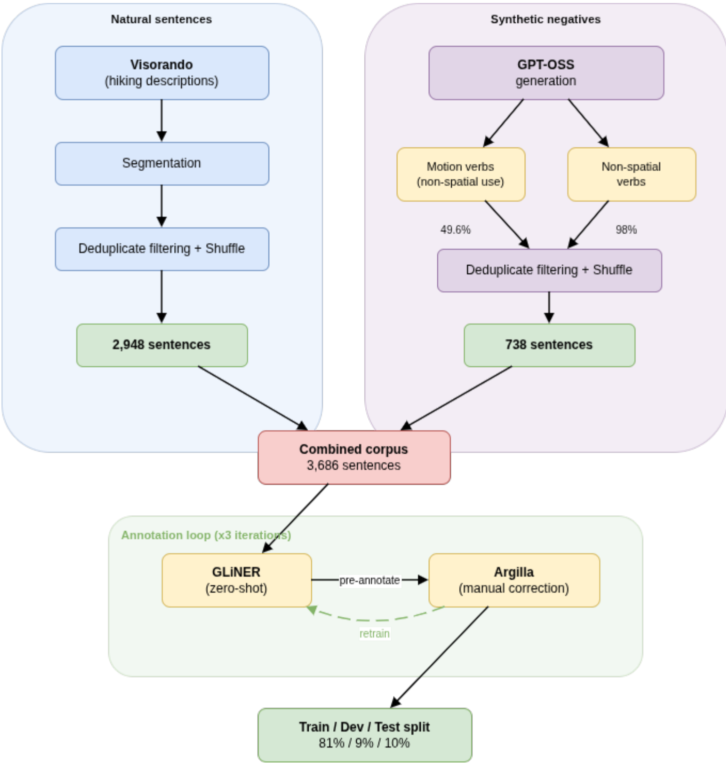


Figure 2: Dataset construction pipeline. Left: Visorando sentences undergo segmentation, deduplication, and shuffling. Right: GPT-OSS generates synthetic negatives via two strategies. Both streams merge and enter an iterative annotation loop before final splitting.

Table 1
Corpus overview by split.

	Train	Dev	Test	Total
Sentences	2,989	331	366	3,686
Visorando	2,389	265	294	2,948
Synthetic	600	66	72	738
Tokens	45,244	4,987	5,739	55,970
Spans	26,219	2,909	3,335	32,463
Avg tok/sent	15.1	15.1	15.7	15.2

Second, we generated synthetic sentences using everyday non-spatial verbs (e.g., *manger (eat)*, *lire (read)*, *écouter (listen)*) as straightforward negative examples. Synthetic negatives serve two complementary purposes. First, non-spatial sentences in hiking guides are inherently rare; the synthetic sentences provide *domain-matched negative examples*, forcing the model to learn that motion-verb vocabulary alone does not signal spatial content — as in Example (1), where only the verb is annotated as MISC, not as ACTION. Second, the verb-only MISC strategy ensures every sentence carries at least one annotation span, preventing the model from learning to assign zero labels as a valid output. Manual filtering retained 98% of non-motion-verb sentences (490/500) but only 49.6% of motion-verb ones (248/500), reflecting the difficulty of generating genuinely non-spatial text when the verb itself carries spatial semantics. Larger language models may yield more diverse outputs and improve this retention rate.[2]

The final corpus comprises 3,686 sentences (55,970 tokens, 32,463 spans), of which 2,948 come from Visorando and 738 are synthetic. We split the data into train/dev/test sets of 2,989/331/366 sentences (≈81/9/10%); Table 1 details the breakdown.

[2]We acknowledge that LLM-generated negatives may exhibit structural repetition despite duplicate removal and manual validation.

Figure 3: Argilla annotation interface showing a hiking sentence with *GLiNER*-bootstrapped span annotations: *(Leave the alley when it turns left, bearing right onto the lane of the water treatment plant).* Annotators correct entity boundaries and types before the model is retrained.

4.2. Annotation method

Annotation followed an active learning workflow [20] using *Argilla* [18] (Figure 3), iterating through: (1) zero-shot *GLiNER* [15] bootstrapping, (2) manual correction, (3) *GLiNER* retraining on corrected data, (4) re-annotation of remaining sentences, and (5) evaluation until convergence.

Annotators see pre-annotated spans from *GLiNER* and correct boundaries and types before the model is retrained. *GLiNER* bootstrapping operates in zero-shot mode using the 10 label names as text prompts; after each correction round the model is fine-tuned on corrected data, progressively improving suggestions. However, *GLiNER's* default greedy decoding rejects partially overlapping spans, which is incompatible with our two-level scheme where Level 1 and Level 2 spans may partially overlap. We use *GLiNER* [15] with *CamemBERTa* [21], a French pretrained bidirectional encoder built on the *DeBERTaV3* architecture [22], as its backbone. Two adaptations were required to support our annotation scheme.

Overlap-tolerant decoding. *GLiNER's* default greedy search selects the highest-scoring candidate span, then discards every remaining candidate that overlaps with it. While *GLiNER* natively supports nested entities, this greedy strategy still prevents detecting *partially* overlapping spans — a structure required by our two-level scheme. For instance, in Figure 3: the GEOMOTION span *Abandonner la ruelle* and the GEODESC span *la ruelle lorsqu'elle tourne à gauche…* share the tokens *la ruelle*, creating a partial overlap that the default decoder would reject. We modified the decoding algorithm so that, for each unique (start, end, type) tuple, only the highest-scoring candidate is retained, but different entity types may overlap freely. This enables the model to predict the partially overlapping annotations required by our scheme.

Focal loss for class imbalance. The label distribution is highly skewed (OFFSET at 24.5% vs. GEOVISION at 0.7%). *GLiNER* uses focal loss [23], $\mathcal{L}_{\mathrm{FL}} = -\alpha_t\,(1-p_t)^{\gamma}\,\log(p_t)$, where α_t balances positive and negative class contributions and the modulating factor $(1-p_t)^{\gamma}$ down-weights well-classified examples to focus training on hard cases. With the default hyper-parameters (α=0.75, γ=0), the modulating factor equals 1 for all examples, so the loss reduces to α-balanced cross-entropy: class frequency is addressed, but easy, already well-predicted spans still dominate gradient updates. Setting $\gamma = 3.0$ activates the focusing component, sharply down-weighting confident predictions and concentrating updates on rare, hard-to-classify labels such as GEOVISION (0.7% of spans).

Iterative training. The annotation was refined through three iterations of the active learning loop. At each iteration, corrected data was split 90/10 for training and testing; the training portion was further divided 90/10 for optimisation and validation. After each cycle, the updated model re-annotated

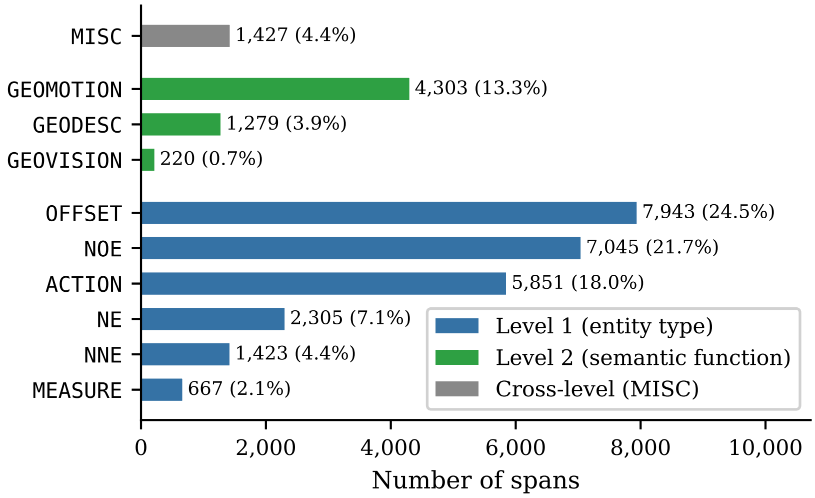


Figure 4: Distribution of all 10 labels, grouped by annotation level. Level 1 labels (blue) identify entity types; Level 2 labels (green) capture semantic functions. NOE outnumbers NE by 3.1×, reflecting the dominance of generic landmarks over toponyms.

remaining uncorrected sentences, progressively improving pre-annotation quality until convergence from the initial zero-shot baseline to a stable annotation assistant.

5. Proposed Corpus Properties

The corpus contains 32,463 annotated spans distributed across 3,686 sentences. As shown in Figure 4, OFFSET (7,943; 24.5%) and NOE (7,045; 21.7%) dominate the distribution, reflecting the density of directional cues and generic waypoints in hiking instructions. ACTION spans (5,851; 18.0%) form the third largest category, consistent with the procedural nature of the text. Mean sentence length is 15.2 whitespace tokens (std 7.68, median 13, 95th percentile 30). The high frequency of OFFSET and NOE reflects the inherently relational nature of hiking instructions, where walkers navigate by spatial relations to generic landmarks rather than by absolute coordinates.

The internal composition of Level 2 spans confirms their functional distinctiveness: 99% of GEOMOTION spans contain an ACTION verb, while 96% of GEODESC and 95% of GEOVISION spans contain an NOE. Motion instructions are verb-anchored, whereas descriptions and visual cues are landmark-anchored.

5.1. NE vs NOE Interplay

NOE consistently dominates NE across all semantic contexts (Figure 5).

NNE composition. Of 1,423 NNE spans, 79.1% combine NE+NOE (e.g., *phare de Biarritz (lighthouse of Biarritz)*), 18.7% contain only NOE, and just 1.9% only NE — toponyms are almost always contextualized by their geographic type.

Contextual ratios. NOE consistently outnumbers NE within Level 2 spans: 3.8× in GEODESC, 2.7× in GEOVISION, 3.0× in GEOMOTION (Figure 5(a)). Even in GEOVISION contexts, where named peaks might be expected, generic landmarks dominate. This consistent dominance has a practical consequence: NER systems trained exclusively on named entities would miss at least 73% of spatial references across all three semantic contexts.

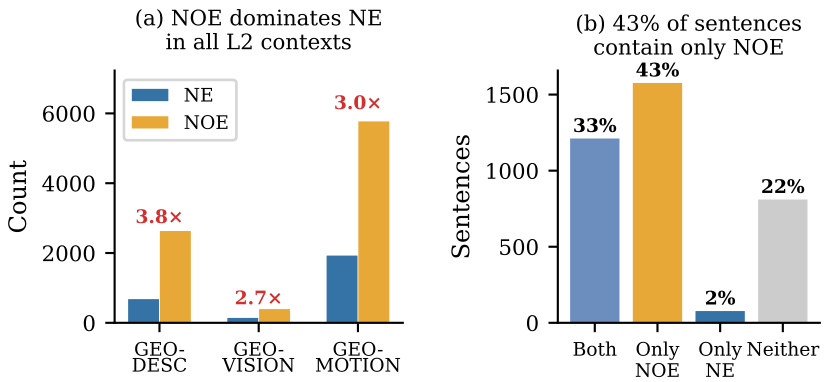


Figure 5: NE vs NOE interplay. (a) NOE outnumbers NE by 2.7–3.8× within every Level 2 context. (b) Sentence-level entity co-occurrence across all 3,686 sentences; 43% contain only NOE, and just 2% contain only NE.

Sentence-level co-occurrence. At the sentence level (Figure 5(b)), 43% of sentences contain only NOEs, 33% both NE and NOE, and merely 2% only NEs; 22% contain neither (primarily synthetic negatives). The rarity of NE-only sentences (2%) confirms that toponyms in hiking descriptions rarely appear without an accompanying generic spatial noun, reinforcing the need for NOE-aware annotation schemes.

5.2. Level 2 Span Overlap

Of the 2,911 sentences containing at least one Level 2 span, 899 (30.9%) exhibit partial overlap between two Level 2 spans. The dominant pattern is GEOMOTION–GEOMOTION (68% of overlap pairs), followed by GEODESC–GEOMOTION (24%). The median overlap extent is 5 tokens, with OFFSET (1,987) and NOE (1,978) as the most frequent Level 1 entities in overlap zones. This prevalence — affecting nearly one-third of spatial sentences — confirms that standard non-overlapping decoders would need to arbitrarily assign shared tokens to a single function, motivating the overlap-tolerant decoding of Section 4.

6. Conclusion

We propose to make available a hybrid annotation method, which combines zero-shot NER with iterative human correction which significantly simplifies the annotation phase, which is known to be time-consuming, tedious and sometimes complex. We also offer to the community a 3,686-sentence reference corpus for hiking descriptions, annotated with a two-level scheme linking atomic entities to their semantic function.

One of its distinctive properties is that Nominal Entities (NOE) consistently outnumber Named Entities (NE) across all contexts, with only 2% of sentences containing NEs in isolation. Combined with an active learning methodology and synthetic negative sampling, this corpus supports building models that move beyond toponym-centric NER.

Several limitations should be noted. Right now the corpus is monolingual (French) and restricted to hiking trail descriptions which may limit generalizability to other spatial discourse. Synthetic negatives were generated by a 120B-parameter model *GPT-OSS*; larger models may produce more diverse and higher-quality examples, particularly for the motion-verb subset where nearly half were discarded.

Future work includes developing architectures that jointly predict Level 1 and Level 2 spans with nesting constraints, extending the corpus to English and Spanish to enable cross-lingual transfer, exploring cross-domain transfer (urban navigation, literary narratives), and investigating alignment with ISO-Space [4]. Mapping Level 1 entities to ISO-Space's inventory (e.g., NE/NOE to typed PLACE, ACTION to MOTION) would facilitate interoperability, while Level 2 functions could extend the standard's relational layer.

Declaration on Generative AI

During the preparation of this work, the authors used *GPT-5* to check grammar and spelling. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] L. Talmy, Lexicalization patterns: Semantic structure in lexical forms, in: T. Shopen (Ed.), Language typology and syntactic description III: Grammatical categories and the lexicon, Cambridge University Press, 1985, pp. 36–149.

[2] C. Vandeloise, L'espace en français, Le Seuil, 1986.

[3] J. A. Bateman, J. Hois, R. Ross, T. Tenbrink, A linguistic ontology of space for natural language processing, Artificial Intelligence 174 (2010) 1027–1071. doi:10.1016/j.artint.2010.05.008.

[4] J. Pustejovsky, Z. Yocum, Capturing motion in ISO-SpaceBank, in: H. Bunt (Ed.), Proceedings of the 9th Joint ISO - ACL SIGSEM Workshop on Interoperable Semantic Annotation, Association for Computational Linguistics, Potsdam, Germany, 2013, pp. 25–34. URL: https://aclanthology.org/W13-0503/.

[5] L. Moncla, M. Gaio, A Multi-Layer Markup Language for Geospatial Semantic Annotations, in: R. S. Purves, C. B. Jones (Eds.), Proceedings of the 9th Workshop on Geographic Information Retrieval, GIR 2015, Paris, France, November 26-27, 2015, GIR '15, ACM, Paris, France, 2015, pp. 5:1–5:10. doi:10.1145/2837689.2837700.

[6] K. Lee, H. Bunt, J. Pustejovsky, A. C. Fang, C. Park, Representing ISO-annotated dynamic information in UMR, in: K. Lai, S. Wein (Eds.), Proceedings of the Sixth International Workshop on Designing Meaning Representations, Association for Computational Lingustics, Prague, Czechia, 2025, pp. 49–58. URL: https://aclanthology.org/2025.dmr-1.6/.

[7] I. Mani, J. Hitzeman, J. Richer, D. Harris, R. Quimby, B. Wellner, SpatialML: Annotation scheme, corpora, and tools, in: N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, D. Tapias (Eds.), Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08), European Language Resources Association (ELRA), Marrakech, Morocco, 2008. URL: https://aclanthology.org/L08-1017/.

[8] M. Gritta, M. T. Pilehvar, N. Collier, A pragmatic guide to geoparsing evaluation, Language Resources and Evaluation 54 (2020) 683–712. doi:10.1007/s10579-019-09475-3.

[9] L. Moncla, D. Vigier, K. McDonough, GeoEDdA: A gold standard dataset for geo-semantic annotation of Diderot & d'Alembert's Encyclopédie, in: Proceedings of the 2nd International Workshop on Geographic Information Extraction from Texts (GeoExT 2024) at ECIR 2024, CEUR-WS, Glasgow, United Kingdom, 2024, pp. 1–10.

[10] T. Kew, A. Shaitarova, I. Meraner, J. Goldzycher, S. Clematide, M. Volk, Geotagging a diachronic corpus of alpine texts: Comparing distinct approaches to toponym recognition, in: Proceedings of the Workshop on Language Technology for Digital Historical Archives at RANLP 2019, Varna, Bulgaria, 2019, pp. 11–18. doi:10.26615/978-954-452-059-5_003.

[11] P. Rayson, A. Sherwood, R. Sherwood, J. Butler, C. Donaldson, I. Gregory, J. Taylor, A deeply annotated testbed for geographical text analysis: The corpus of Lake District writing, in: Proceedings of the 1st ACM SIGSPATIAL Workshop on Geospatial Humanities (GeoHumanities'17), ACM, Redondo Beach, CA, USA, 2017, pp. 9–15. doi:10.1145/3149858.3149865.

[12] S. Higashiyama, H. Ouchi, H. Teranishi, H. Otomo, Y. Ide, A. Yamamoto, H. Shindo, Y. Matsuda, S. Wakamiya, N. Inoue, I. Yamada, T. Watanabe, Arukikata travelogue dataset with geographic entity mention, coreference, and link annotation, in: Findings of the Association for Computational Linguistics: EACL 2024, Association for Computational Linguistics, St. Julian's, Malta, 2024, pp. 513–532.

[13] A. Yamamoto, H. Otomo, H. Ouchi, S. Higashiyama, H. Teranishi, H. Shindo, T. Watanabe, Graph-

structured trajectory extraction from travelogues, in: Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (ACL 2025), Association for Computational Linguistics, Vienna, Austria, 2025, pp. 14116–14132.

[14] L. Martin, B. Muller, P. J. Ortiz Suárez, Y. Dupont, L. Romary, Éric de la Clergerie, D. Seddah, B. Sagot, CamemBERT: a tasty French language model, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL 2020), Association for Computational Linguistics, 2020, pp. 7203–7219. doi:10.18653/v1/2020.acl-main.645.

[15] U. Zaratiana, N. Tomeh, P. Holat, T. Charnois, GLiNER: Generalist model for named entity recognition using bidirectional transformer, in: Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL 2024), Association for Computational Linguistics, Mexico City, Mexico, 2024, pp. 5364–5376. doi:10.18653/v1/2024.naacl-long.300.

[16] M. Gaio, C. Sallaberry, P. Etcheverry, C. Marquesuzaa, J. Lesbegueries, A global process to access documents' contents from a geographical point of view, Journal of Visual Languages & Computing 19 (2008) 3–23. URL: https://www.sciencedirect.com/science/article/pii/S1045926X07000456. doi:https://doi.org/10.1016/j.jvlc.2007.08.010, spatial and Image-based Information Systems.

[17] I. Mani, C. Doran, D. Harris, J. Hitzeman, R. Quimby, J. Richer, B. Wellner, S. Mardis, S. Clancy, SpatialML: Annotation scheme, resources, and evaluation, Language Resources and Evaluation 44 (2010) 263–280. doi:10.1007/s10579-010-9121-0.

[18] D. Vila-Suero, F. Aranda, Argilla – open-source framework for data-centric NLP, 2023. URL: https://github.com/argilla-io/argilla.

[19] OpenAI, gpt-oss-120b & gpt-oss-20b model card, 2025. URL: https://arxiv.org/abs/2508.10925. arXiv:2508.10925.

[20] B. Settles, Active Learning, Synthesis Lectures on Artificial Intelligence and Machine Learning, Morgan & Claypool Publishers, 2012. doi:10.2200/S00429ED1V01Y201207AIM018.

[21] W. Antoun, B. Sagot, D. Seddah, Data-efficient French language modeling with CamemBERTa, in: Findings of the Association for Computational Linguistics: ACL 2023, Association for Computational Linguistics, Toronto, Canada, 2023, pp. 5174–5185. doi:10.18653/v1/2023.findings-acl.320.

[22] P. He, J. Gao, W. Chen, DeBERTaV3: Improving DeBERTa using ELECTRA-style pre-training with gradient-disentangled embedding sharing, in: Proceedings of the 11th International Conference on Learning Representations (ICLR 2023), 2023.

[23] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV 2017), 2017, pp. 2980–2988. doi:10.1109/ICCV.2017.324.