Beyond Toponyms: Nominal Spatial Entities for Understanding Movement from Hiking Descriptions

Abdelkrim Tafer[1,2], Mauro Gaio[2], Javier Lacasta[1]
1 Advanced Information Systems Laboratory (IAAA),
Aragón Institute for Engineering Research (I3A),
Universidad de Zaragoza, Mariano Esquillor s/n, 50018, Zaragoza, Spain.
Tel. +34-976762707, e-mail: atafer@univ-pau.fr
2 Université de Pau et des Pays de l'Adour (UPPA).

Summary
Movement descriptions connect actions, places, landmarks, directions and distances. This work builds an annotated corpus of French route instructions and develops neural extraction methods to identify these cues, especially ordinary landmarks difficult to match to geographic databases. The objective is to support a system that reconstructs routes from text.

Introduction
Human displacement is important because mobility patterns are linked to urban planning, traffic forecasting, epidemic control, emergency response and behavioural modelling [1]. Most computational studies of movement rely on GPS traces or mobile-phone records. These sources give precise coordinates, but they do not explain how people describe a route, what landmarks they use, or which elements of the environment guide their decisions. Textual itineraries fill this gap: they encode movement as actions, places, landmarks, directions and distances.
This work studies hiking descriptions as a concrete case of real movement narratives. The data come from Visorando, a collaborative hiking platform that publishes route descriptions and GPX traces for outdoor walks.1 These descriptions are written for humans: "turn left", "cross the intersection", "continue on the path", "walk towards the church". For automatic processing, however, these expressions are not directly usable. The challenge is to transform them into structured spatial information while preserving the way humans describe movement.
Research on route directions has shown that spatial discourse combines actions and landmarks [2]. A route instruction is rarely a place name alone. It usually contains an action, such as turn or cross; a landmark, such as a bridge or a church; and a relation, such as left, after or towards. Our work is based on this idea: before reconstructing movement, we must first identify the linguistic elements that describe it.

What is a spatial entity?
In this work, a spatial entity is any textual unit that contributes to understanding where the walker is, what the walker does, or what the walker sees. We use six main categories. Named Entities (NE) are proper place names (named toponyms) such as Pau, Biarritz or GR10. Nominal Entities (NOE) are common spatial nouns such as path, bridge, church, square or intersection. Nested Named Entities (NNE) combine both, for example "the lighthouse of Biarritz" or "the church of Saint-Martin", but also "the peak of the mountain". ACTION marks movement or perception verbs such as turn, cross, continue or admire. OFFSET marks spatial relations and directions such as left, towards, after or north. MEASURE marks distances or durations such as 400 m or 20 minutes.
This distinction is necessary because place names are only part of the route. Named places help locate a route on a map, but nominal landmarks explain how the route is followed on the ground. In our reference corpus, NOE outnumber NE in all route-level contexts: 3.0 times more in movement instructions, 3.8 times more in environment descriptions and 2.7 times more in visual cues. Only 2% of sentences contain named entities alone. A system trained only to detect place names would therefore miss most spatial evidence.

Three cases of nominal entities
We identified three categories of NOE. The first is the type-of-named-toponym NOE. In "the church of Saint-Martin", the word church gives the geographic type of the named place. This case can often be geocoded through the associated toponym, but the NOE remains essential because it says what kind of object is involved.
The second case is the contextually anchored NOE. In "from the village of Laruns, turn right after the church", the church may not be named in the same sentence, but the preceding context can make it identifiable. This is a coreference problem: the system must link the nominal expression to an earlier NE, NNE or discourse anchor.
The third case is the non-geocodable NOE. Expressions such as "the red house", "the old stone wall", "the narrow street" or "the archway" may have no unique database entry. They cannot be resolved like classical toponyms, but they are nonetheless essential for navigation. This is why nominal entities are not secondary details: they are often the objects that make a route understandable.

1 Visorando: collaborative hiking platform and trail database, https://www.visorando.com/ (accessed 12 February 2025).

Work performed

The first experiment evaluated whether neural models could replace or complement automatic rule-based annotation. We scraped 1,897 hiking routes, including 1,098 in France, and extracted the French route descriptions and GPX files. After sentence segmentation, we obtained 27,083 sentences. An automatic annotator produced a silver-standard corpus for NE, NNE, ACTION, OFFSET and MEASURE; an 806-word dictionary was added for NOE.

We compared BiLSTM-CRF [3], CamemBERT [4] and GLiNER [5]. CamemBERT obtained the best overall F1 score on the silver-standard dataset, while GLiNER was useful because it predicts spans directly rather than assigning one tag per token. This matters for expressions such as "the church of Saint-Martin", where several labels can overlap.

The second stage produced a corrected reference corpus of 3,686 French hiking sentences. Level 1 annotates the atomic entities defined above. Level 2 groups them into semantic functions: GEOMOTION for movement instructions, GEODESC for environment descriptions and GEOVISION for visual cues. The corpus was built with zero-shot GLiNER bootstrapping, human correction and iterative retraining. We also added 738 synthetic negative sentences to separate spatial and non-spatial uses of similar verbs. For example, "follow the path" is spatial, but "follow the instructions" is not. These examples prevent the model from learning that every movement-like verb is automatically an ACTION.

The current modelling work adapts span-based NER to this two-level schema. Instead of assigning one label to each token, the model scores complete spans against the possible labels. This is better adapted to route descriptions, where expressions are often multi-word and overlapping. For example, "continue on the wide path towards the bridge" contains an ACTION, a NOE and an OFFSET, but it can also form one GEOMOTION instruction. The model therefore represents each candidate span using its boundaries and internal words, then uses Level 1 evidence to help classify Level 2 functions. The goal is to obtain reliable spatial evidence: what is named, what is described, what action is expressed, and which words form a movement instruction. Figure 1 reports entity counts and the Level 2 micro-F1 gain over GLiNER.

LLM and vision-language experiments

After extraction, the remaining challenge is entity resolution. We conducted exploratory experiments on three tasks. First, a language model receives the raw text and extracted candidates and returns a constrained geographic context: likely countries, regions, confidence and a short justification. This reduces ambiguity before geocoding. Second, a language model links contextually anchored NOE to their textual anchor, for example linking "this town" to Pau. Third, for non-geocodable NOE such as "archway", we test a vision-language setting where the text, nearby extracted entities and a local map frame are used together. These experiments target the central difficulty: many landmarks that guide movement are not standard database places.

Conclusions

This work shows that route reconstruction from text cannot rely on place names alone. We produced a corrected reference corpus of 3,686 French hiking sentences with 32,463 annotated spans, linking basic spatial cues to movement instructions, environmental descriptions and visual cues. The corpus confirms the central result of the work: nominal landmarks, such as paths, roads, churches and bridges, are more frequent than named places in every context; 43% of sentences contain only such landmarks, while only 2% contain named places alone.

These results define the next step toward route reconstruction. A useful system must extract actions, landmarks, directions and distances, then resolve ordinary landmarks through their type, textual context, map evidence or visual evidence. The final objective is to connect textual instructions to geographic candidates and reconstruct a route that remains explainable from the original description.

REFERENCES

[1]. GONZÁLEZ, M.C., HIDALGO, C.A. and BARABÁSI, A.-L. Understanding individual human mobility patterns. Nature. 2008, 453, 779-782. Available from: doi:10.1038/nature06958.

[2]. DENIS, M. The description of routes: a cognitive approach to the production of spatial discourse. Cahiers de Psychologie Cognitive. 1997, 16(4), 409-458.

[3]. LAMPLE, G., BALLESTEROS, M., SUBRAMANIAN, S., KAWAKAMI, K. and DYER, C. Neural architectures for named entity recognition. In: Proceedings of NAACL-HLT 2016. San Diego: ACL, 2016, pp. 260-270.

[4]. MARTIN, L., MULLER, B., ORTIZ SUÁREZ, P.J., DUPONT, Y., ROMARY, L., DE LA CLERGERIE, É., SEDDAH, D. and SAGOT, B. CamemBERT: a tasty French language model. In: Proceedings of ACL 2020. Online: ACL, 2020, pp. 7203-7219. Available from: doi:10.18653/v1/2020.acl-main.645.

[5]. ZARATIANA, U., TOMEH, N., HOLAT, P. and CHARNOIS, T. GLiNER: generalist model for named entity recognition using bidirectional transformer. In: Proceedings of NAACL 2024. Mexico City: ACL, 2024, pp. 5364-5376. Available from: doi:10.18653/v1/2024.naacl-long.300.

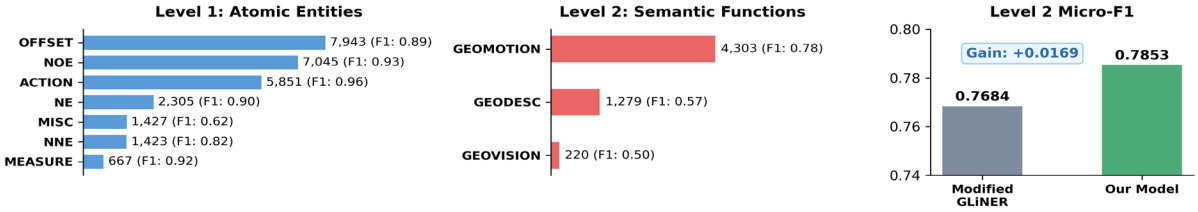

Figure 1. Entity counts and Level 2 performance gain.