An automatic extraction method of static and dynamic spatial contexts from texts

Ludovic Moncla, Mauro Gaio, Ekaterina Egorova, Christophe Claramunt

An automatic extraction method of static and dynamic spatial contexts from texts

Ludovic Moncla* Mauro Gaio**
Ekaterina Egorova*** Christophe Claramunt*

*Naval Academy Research Institute, Brest, France
ludovic.moncla,christophe.claramunt@ecole-navale.fr
**LIUPPA, Université de Pau et des Pays de l'Adour, Pau, France
mauro.gaio@univ-pau.fr
***Department of Geography, University of Zurich, Zurich, Switzerland
ekaterina.egorova@geo.uzh.ch

Abstract. Spatial descriptions, with or without motion, are the main issues addressed by this paper. We describe construction grammars implemented in the PERDIDO platform with cascaded finite-state transducers which aims at marking and formalizing relations between extended named entities, geographical terms, spatial relations and motion verbs. These grammars can be seen as a computational synthesis of the work on the expression of space and motion in natural language. The proposed method for geographical information extraction has been tested for three different projects within the digital humanities using specific corpora. The first task deals with the extraction of place names from French novels, the second task deals with the extraction of motion events from hiking descriptions written in Romance languages (French, Spanish and Italian) and the third task aims at identifying fictive motion expressions in English alpine journals.

1 Introduction

It is established that one of the best ways to approach spatial semantics is through its representations in language. In all the viable representations, two subsets may be distinguished. Representations where the spatial situation described is motionless and representations where the spatial relation between entities changes over time. In other words, the spatial context may be static or dynamic. Whatever the context is, spatial descriptions involve three main components: a located entity called "target" (Vandeloise, 1991); a reference entity called "landmark" (Langacker, 1987; Vandeloise, 1991) and a spatial relation between these two entities.

For static descriptions, the relation is often carried by at least one adpositional element applied to the noun denoting the landmark (Levinson and Wilkins, 2006). For dynamic descriptions, motions events or displacements are introduced by one or more verbal and adpositional elements also applied to the noun representing landmark. Although these patterns are not unique (see e.g. Levinson and Wilkins (2006)), it has been observed that landmarks are

larger, more salient and stable than targets, since the main purpose of this kind of descriptions is to locate one entity with respect to another. Regarding dynamic scenes, a wide range of publications have specifically addressed the expression of motion in language (Talmy, 1985; Tenny and Pustejowsky, 1999; Talmy, 2000; Hickmann, 2006). In particular, lexicon-grammar approaches have significantly contributed to the represenation of dynamic spaces (Asher and Sablayrolles, 1995; Borillo and Sablayrolles, 1993; Laur, 1993; Muller and Sarda, 1998; Aurnague, 2011). Other studies oriented to an integration of additional spatial semantics have integrated the ontological nature of the landmark entities denoted by the nominal elements that propositions and verbal units select. Among the many phenomena tackled in this literature, an interesting pattern that appears is that some motion verbs and constructions are likely generate some static interpretations. This phenomenon is often called "fictive motion" (Talmy, 2000) or "non-actual motion" (Blomberg and Zlatev, 2014).

Spatial descriptions, with or without motion, are the main issues addressed by this paper. A first one mainly concerns the expression of the located entity called target. Thus, the task of "Named Entity Recognition and Classification" (i.e., NERC algorithms) is considered to play an important part in the processing of spatial descriptions. When considering such motionless spatial descriptions, a first experiment has been done to automatically recognize and extract the places mentioned in the context of French novels. This process also known as "geoparsing" can be non-straightforward for fictional texts because a novelist has often multiple ways of evoking a given place: either directly (by giving an explicit name) or more elusively (by using relative references, e.g., near, behind, or two blocks further, relative to other places mentioned before). Some places can be even deliberately disguised and others can be completely imaginary. These different cases, among many others, can be found in a same novel. Automatizing the recognition of all these kinds of places is even more difficult when referring to ancient texts (Matei-Chesnoiu, 2015). Additionally, the way real vs. fictive motion occurs in discourse is a quite unexplored question.

A second set of issues addressed in this paper concerns the semantic and syntactic relationships between the verb and the possible adverbial elements appearing in motion descriptions. More precisely with respect to fictive or non-actual motions, our objective is oriented to the identification of the whole range of verbs that are likely to occur in fictive motion descriptions. The ontological nature of the target entities appearing in this kind of interpretation of motion verbs have been another tackled issue for which an in-depth analysis of French texts have provided valuable insights.

Two different experiments have been conducted to illustrate the potential of our method oriented towards an automatic information extraction of real and fictive motion expressions in texts. The first experiment focuses on motion and geo-spatial information extraction for itinerary reconstruction from texts written in Romance languages. The second experiment implements a similar method for automatic extraction and classification of fictive motion expressions in an English corpus.

2 Geoparsing places

2.1 Construction grammar for Extended Named Entities extraction

It is generally accepted that proper name is the most frequent component of a Named Entity (NE). As proper name is a linguistic issue still under discussion, we have adopted a definition according to various criteria. Words (one or several following each other) starting with an uppercase letter is the most commonly cited criterion (Fourour et al., 2002), especially in Romance languages. But this criterion alone is not discriminatory because it suffers of many exceptions (vallée de la mort or vallée de la Mort, gave de Pau, massif armoricain or Massif armoricain). Again in Romance languages the second criterion is morphosyntactic, it shows that most of the time a proper name does not involve determiners and could not be inflected, but this criterion also suffers of some exceptions on at least one of the two rules (Le Havre, La Rochelle, La Pierre Saint Martin, Gaves Réunis) or both (Les Deux Alpes). However, let us not confuse with expressions like l'Aquitaine, les Pyrénées, le Rhône, where the determiner is not really part of the proper name even if the rule could be considered as waived. A third criterion is the non-significance of a proper name, but here again this criteria suffers of exceptions (Archipel des Sanguinaires, Petit Mont Blanc, gare du Nord). A last criterion is the uniqueness of the proper name reference but as the others criteria some exceptions can be found and moreover some common names can have a single reference (the sun, the house on the left after the crossroad). As regards the syntactic shape we selected Jonasson's (Jonasson, 1994) categories of proper names. A proper name can be categorised as pure or descriptive. Pure proper names are simple (i.e., composed of a single lexeme) or complex (i.e., composed of several lexemes) and are composed of proper names only. Descriptive proper names refer to a composition of proper names and common names (i.e., descriptive expansion). In other words, a descriptive proper name overlaps a pure proper name and refers to NE built with a pure proper name and a descriptive expansion. This expansion can change the implicit type (e.g., location, person, etc.) of the initial pure proper name and then of the NE. However, the presence of a proper name in NE is not mandatory. In some specific contexts, la Ville Lumière, le pays du Soleil-Levant ou le sommet du Monde, could be considered as NE. As mentioned in (Kleiber, 1981) these expressions are part of what the author calls "la description définie" (the defined description) namely expressions having the ability to make reference to an entity identifiable as such in a given context.

Finally, according to these concepts of proper name and Named Entity, we introduced in a previous work the concept of Extended Named Entity (ENE) (Gaio and Moncla, 2017). The notion of "extended" is pretty near to the one named "mixed" proposed in (Fourour et al., 2002).

We defined several levels of overlapping (0, 1, 2, etc.) for the representation of ENE. Each level is encapsulated in the previous one. For instance, level 0 refers to pure proper names and can be seen as the core component of an ENE. Thus, we consider NE as a special kind of ENE. Then, level >0 refers to descriptive proper names composed of another descriptive proper name or of a pure proper name (i.e., an entity of level 0) and a common noun. Descriptive expansions may or may not change the implicit or default nature of the object described by the proper name. Indeed, when the associated term has not the same type of the intrinsic or default feature type of the pure proper name, it defines a new entity that overlaps the pure proper name one.

(1) a. *sindaco di* [*Genova*]
'mayor of Genoa'

b. *club de pelota vasca de la* [*ciudad de* [*San Sebastián*]]
'Basque pelota club of San Sebastian city'

c. *dépression karstique dans le* [territoire aride au sud de la [*région* d' [Aragon]]]
'karstic depression in the arid land south of the Aragon Region'

Examples (1a-1c) show that an entity may contain the name of another entity, and that the new entity may have a different feature type. For instance, 'Genoa' refers to a location whereas 'mayor of Genoa' refers to a person or a function (see example (1a)). Additionally, there is not really a limit to the overlapping. However, it is quite uncommon to find an ENE of a level greater than 3 (see example (1c)). We have considered the annotation of ENE as a shallow parsing and the grammar to be used as a specific construction. The core of the grammar [1] is set as follow:

$S \rightarrow ENE$
$ENE \rightarrow ENEA \quad | \quad (Term)\ ENER$
$ENER \rightarrow Offset\ ENEA \quad | \quad Offset\ ENER$
$ENEA \rightarrow (Term)\ ProperNoun \quad | \quad Term\ ENEA$
$Term \rightarrow Nominal\ Det$

This grammar integrated in the PERDIDO platform is implemented as a cascade of finite-state transducers using the CasSys program available in the Unitex/GramLab platform [2]. We developed an hybrid solution combining a preprocessing step for the disambiguation of grammatical categories (using part-of-speech taggers) and the cascade of transducers. The proposed PERDIDO NERC tool is based on a bottom-up strategy where each level of the ENE is marked, from the pure proper name to the whole ENE. It can distinguish between two types of ENE, 'absolute' referring to standard spatial ENE and 'relative' referring to spatial ENE associated with spatial relations (i.e., 'offset' and 'measure'). The cascaded finite-state transducers produce a generic annotation of ENE (i.e., ENE boundaries are identified but not classified). Therefore, for geocoding, PERDIDO implements a gazetteer lookup method to classify them and uses the local linguistic context (i.e., feature type within ENE), when available, to identify subtypes associated with ENE (e.g., city, street, church) to classify them and, more specifically to identify the spatial ones.

With respect to the specific problem of the NERC category of place names, one might move beyond reducing a place to a name and then geocoded with a single set of coordinates, a model that is still predominant in Geographic Information Science (Purves and Derungs, 2015). For instance, taking the example (1c), using the PERDIDO NERC tool, this produces the result represented in a feature structure form in Fig. 1. We argue that for a fine-grained task, especially in digital humanities, such as marking, classifying and disambiguating named entities, it is essential to consider ENE (1c) as a composition of entities. In a such case, standard NER

1. $Offset$ can be seen as an adverbial clause
2. http://unitexgramlab.org/

tools such as OpenCalais [3], OpenNLP [4] and Stanford-NER [5] consider only the entity 'Aragon Region', and therefore lead to inaccuracies in classification and/or disambiguation.

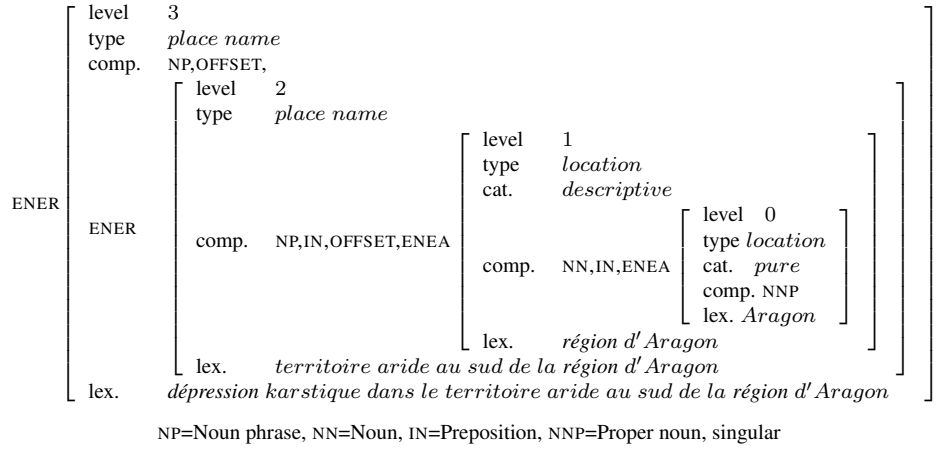


FIG. 1: Feature structure representation of ENE (1c)

2.2 Automatic extraction of place names from French novels

The method developed so far for automatically retrieving place names has been applied to French novels of the 19th century (Moncla et al., 2017). This work has been conducted in the context of a research project whose aims are to provide a method for the cartographic analysis of Paris street names in French novels. The corpus used for this experiment is composed of 31 French novels covering different periods of the 19th century centered on Paris.

As described in the previous section, the proposed construction grammar is applied to the extraction of "complex" place names such as example (2) [6]. This allows us to extract relative place names composed of spatial relations (*offset*) and ENE. This is particularly useful for the extraction of the static spatial context and not just standard place names. Moreover, this method can also be applied for the extraction of place names at different scales, such as places included inside other places (e.g., buildings or streets located inside a city or another administrative entity).

(2) *[...] se trouvait au coin de la* ***rue des Poissonniers*** *et du* ***boulevard de Rochechouard***
'[...] is located at the corner of ***Poissonniers Street*** *and* ***Rochechouard Boulevard***'

Figure 2 shows an excerpt of the XML/TEI annotation [7] of the place name shown in example (2). This figure shows the construction of the relative ENE (identified by the TEI element:

3. http://www.opencalais.com/
4. http://opennlp.apache.org/
5. http://nlp.stanford.edu/ner/
6. This sentence is extracted from the novel *L'Assommoir* written by Emile Zola in 1877.
7. The values of attributes *type* and *subtype* of the *geogName* element refer to GeoNames feature codes: http://www.geonames.org/export/codes.html

placeName) using an offset (*at the corner*) and two absolute ENE (*Poissonniers Street* and *Rochechouard Boulevard*).

```
<placeName type="relative" subtype="compound">
  <offset type="inclusion">
    <w lemma="au" type="PREPDET">au</w>
    <w lemma="coin" type="PREP">coin</w>
    <w lemma="de" type="PREP">de</w>
  </offset>
  <placeName n="1" type="absolute">
    <geogName type="R" subtype="ST">
      <w lemma="le" type="DET">la</w>
      <geogFeat>
        <w lemma="rue" type="N">rue</w>
      </geogFeat>
      <w lemma="de" type="PREP">des</w>
      <name>
        <w lemma="Poissonniers" type="NPr">Poissonniers</w>
      </name>
    </geogName>
  </placeName>
  [...]
</placeName>
```

FIG. 2: XML/TEI annotation of a place name using the PERDIDO NERC tool

According to the results provided by the PERDIDO NERC tool, 112 descriptive expansions of ENE referring to geographical feature types were found in the corpus. In particular, street is the most used geographical feature, which confirms the great interest in these novels for the cartographic analysis of Paris (Moncla et al., 2017) using street names. The proposed method can be used to generate diagrams (showing indicators such as the distribution of the number of occurrences of street names compared to the number of distinct streets mentioned) or maps built using geohistorical gazetteers (see Fig 3). Furthermore, the preliminary results described by Moncla et al. (2017), highlight the great interest for digital humanities in combining the PERDIDO NERC tool with a textometric analysis tool to provide automated analysis of novels based on spatial named entities. Indeed, the direct access to a corpus of texts through the use of place names significantly transforms the ways in which space and fictional landscapes can be explored. It becomes possible to interactively and simultaneously browse through geographical and literary space.

3 Retrieving the dynamic space context from texts

3.1 Construction grammar for motion expressions

For a better understanding of the spatial context, linguists have highlighted the importance of the use of motion verbs and spatial relations, especially in Romance languages (Aurnague, 2011). This leads us to take into account movement verbs and spatial offsets in the parsing process. The core of the 'VT' grammar proposed hereafter can be seen both as a specialisation and as an extension of the ENE construction grammar and it aims to be a computational attempt to provide a synthesis of previous works on how language expresses displacement (Talmy, 1983; Vandeloise, 1986), and on how movement verbs are used in some sentences (Pourcel

FIG. 3: Occurrences of street names represented using proportional lines (Moncla et al., 2017)

and Kopecka, 2005), and on how these verbs are combined with different prepositions (Boons, 1987). The core of the grammar is as follows:

$S \rightarrow V\ \ T$
$V \rightarrow Verb \mid Verb\ \ SO$
$C \rightarrow Conjonction \mid ,$
$LT \rightarrow ENE\ \ C\ \ T$
$T \rightarrow (SO)\ (det)\ ENE \mid (SO \mid ENE)\ T \mid (SO)\ LT$

The symbol V represents a set of movement verbs and the symbol T a set of n-tuples, i.e., a composition of elements belonging respectively to three sets: SO a set of spatial offsets (that can be seen as a spatial adverbial clause), TG a set of geographical noun phrases and E a set of ENE.

(3) *Descendre sur le territoire aride au sud de la région d'Aragon*
'Go down onto the arid land south of the Aragon region.'

Example (3) has the following VT structure= (v, t), with: v = *descendre*, t = *sur le territoire aride au sud de la région d'Aragon*. With t respectively composed of: $tg_3 = \emptyset$, so_3 = *sur*, ENE_2 = *territoire aride au sud de la région d'Aragon*, tg_2 = *territoire aride*, so_2 = *au sud de*, ENE_1 = *région d'Aragon*, tg_1 =*région*, $so_1 = \emptyset$, ENE_0 = *Aragon*.

The set SO of spatial offsets is composed of locative phrases in which, at least in verb-framed languages such as French, the role of prepositions is central. A large number of studies have shown that prepositions are involved in the operation of spatial tracking, or location. With respect to the location concept, following Talmy's work Talmy (1983) and Vandeloise's Vandeloise (1986) proposals, prepositions contribute significantly to reconciliate two entities: a locator and a localised entity (i.e., a landmark and a target in Vandeloise's terms). The part of the phrase used as locator must have spatial properties that facilitate its identification and the explanation of the spatial relationship in which it is involved. Boons (1987) proposed to classify motion verbs according to the aspectual properties of movement called 'aspectual polarity'. The three aspectual polarities are initial (e.g., to leave), median (e.g., to cross) and final

(e.g., to arrive). Without changing the intrinsic aspectual polarity of the verb, the preposition can change what could be called the focus of the displacement. More specifically, the association of a motion verb with a spatial preposition can change the focus of the displacement and take on the aspectual polarity of the preposition instead of that of the verb (Laur, 1993). Undeniably, 'leaving from Paris' and 'leaving for Paris' are two expressions with radically opposite focus of the displacement.

The bottom-up parser, based on the VT grammar and implemented with a cascade of transducers within the PERDIDO platform, can be viewed as searching through the space of possible parse trees to find the correct parse tree for a given 'VT' phrase.

3.2 Reconstruction of itineraries from texts

For experiment purposes, this construction grammar has been applied to the extraction of the dynamic space context from texts. More specifically, we try to use the information of motion expressed in texts to automatically reconstruct trajectories of displacements.

(4) a. *[Emprunter] successivement* ***rue des Capucins*** *et* ***rue de Compostelle***.
'Walk down Capucins Street and then Compostelle Street.'

b. *[Prendre] à gauche après l'entrée de* ***l'usine de Fontanille***.
'Turn left after the entry to the Fontanille factory.'

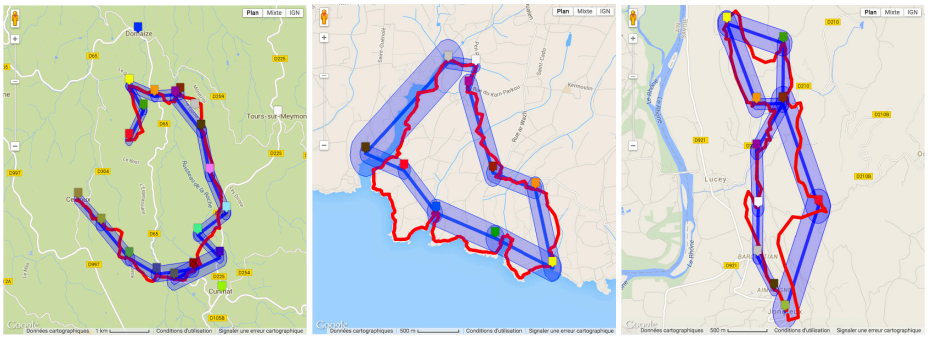
c. *[Suivre] la route depuis le* ***hameau Lic*** *jusqu'à la* ***Chapelle Saint-Roche***.
'Follow the road from the hamlet Lic to the Chapelle Saint-Roche.'

For the automatic reconstruction of itineraries from texts, we proposed a multi-criteria approach combining quantitative and qualitative criteria based on knowledge extracted from the text and geographic databases (Moncla et al., 2016). The proposed method builds a weighted complete graph using the multi-criteria approach where edges represent route segments and vertices represent locations. Then, in order to identify the sequence of waypoints (excluding landmarks) and build an approximation of a plausible footprint of the itinerary described, the graph is transformed into a directed acyclic graph using a minimum spanning tree and spatio-temporal information extracted from the text (see Fig. 4).

For the evaluation of our approach, we used a multilingual corpus (French, Spanish and Italian) of 90 hiking descriptions manually annotated. Each document in the corpus describes one trail and it is associated with the real trajectory (GPS) of the route (used as a comparison basis). Hiking descriptions are a specific type of document describing displacements using geographical information, such as toponyms, spatial and motion relations, and natural features or landscapes, such as shown on example (4a)-(4c).

	# of ENE	*Recall*	*Precision*	*SER*
French	660	95%	96%	17%
Spanish	421	97%	99%	15%
Italian	475	84%	98%	32%
total	1556	92%	97%	21%

TAB. 1: Evaluation of the NERC task with Perdido.



(a) (b) (c)

FIG. 4: Automatic reconstruction of itineraries (Moncla et al., 2016)

Table 1 shows recall, precision and SER scores for the NERC task according to the reference number of ENE in the Perdido gold-standard corpus. The SER (Slot Error Rate) (Makhoul et al., 1999) is the ratio of the total number of slot errors, i.e., insertions, deletions and substitutions (wrong classification and wrong boundaries), divided by the total number of relevant results in the reference. We compared the results obtained with Perdido and with the CasEN system (Friburger and Maurel, 2004) for the automatic annotation and classification of named entities. Although, CasEN obtains good results on a corpus of French newspapers, it obtains an SER score of 51% using the Perdido corpus of French hiking descriptions. This score is mainly explained by the fact that CasEN uses dictionaries of proper names whereas Perdido uses linked data resources. More details about the evaluation of the NERC task for the three languages on the Perdido corpus are given in (Moncla, 2015).

Additionally, the corpus analysis shows that only 2% of ENE are not spatial entities. Furthermore, 810 occurrences of spatial ENE are contained within a VT structure (i.e., 53%) and 47% are associated with feature types (i.e., 53% of spatial ENE belong to the level 0) and a very few number of spatial ENE (3%) are built with more than one expansion (level >1). Additionally, about 59% of verbs are motion verbs (i.e., 1985 occurrences). Median and final motion verbs are the most frequent ones and only 3% of verbs belonging to a VT structure refer to verbs of perception (i.e., 113 occurrences).

4 Fictive motion: static and dynamic scenes

Fictive motion (FM) is an example of the metaphoric nature of human thought and language (Lakoff and Johnson, 2008) that can represent a challenge in the task of automatic identification of motion events in text. Essentially, this linguistic structure represents a static spatial entity as moving, as in (5a) and cannot be interpreted literally. Moreover, there are two types of FM (Matsumoto, 1996). Type I is a static description of a spatial entity and it location in space, as in (5a). Type II is based on "the actual motion of a particular moving entity at a particular time" Matsumoto (1996, p. 361), as exemplified in 5b – imagine this sentences being uttered by a driver, who is actually moving (in the car).

(5) a. The mountain [range goes] from Canada to Mexico.
b. This [highway will enter] California soon.

Egorova et al. (2016) examined the use of FM in a corpus of alpine texts "Text+Berg" (Bubenhofer et al., 2015). They queried the corpus for a spatial entity (based on a list of terms) followed by a verb and manually created a corpus of FM out of candidate phrases. Both types of FM were found in alpine narratives, alongside with two distinct subcategories of Type I: description of a vista along the way (6a) and description of general spatial knowledge about a larger geographic area 6b). Type II, as stated in the definition, encodes the actual motion of the mountaineer, as in (6c). All the three uses of FM (Type I, vista; Type I, spatial knowledge; Type II) were subsequently annotated.

(6) a. Beyond, the desert [hills rose] to the Russian-China border. (Type I, vista)
b. This wonderful [chain runs] NE to SW for some 13 km. (Type I, spatial knowledge)
c. The [ridge went] on forever, but after what seemed an age... (Type II)

We use the "Text+Berg" corpus (Bubenhofer et al., 2015) and the subcorpus of annotated FM (Egorova et al., 2016) to automatically reproduce the half-automated extraction and manual annotation of FM into the three types performed by (Egorova et al., 2016).

For the extraction of FM, we use the lists of geographic entities and motion verbs from (Egorova et al., 2016). Although a list-based extraction is straightforward, resulting in high recall, dealing with various types of false positives – e.g. factive motion (as in 7a) or the use of motion verbs in metaphoric sense s (as in 7b) – represents an interesting task, requiring a set of additional rules that we develop within the PERDIDO platform.

(7) a. [Half the peak fell] in prehistoric times.
b. Out of sheer jealousy the mighty [mountain went on war] against Carihuairazo...

To classify the identified FM into the three types, we identify concepts that can be used for their differentiation. For example, (8) is a vista because of the explicit inclusion of the observer into the frame of reference. We further operationalize these concepts through linguistic structures, based on the literature and thesauri (e.g. expanding potential linguistic encodings of a concept through synonyms or words in the same semantic field).

(8) Down below us the [glacier snaked] away.

5 Conclusion

In this paper, we described a method for the automatic extraction of static and dynamic spatial context from texts. We have proposed construction grammars implemented in the PERDIDO platform with cascaded finite-state transducers which aims at marking and formalizing relations between ENE, geographical terms, spatial relations and motion verbs. These grammars can be seen as a computational synthesis of the work on the expression of space and motion in natural language.

The proposed geoparser tool has been tested for three different tasks (i.e., retrieving place names, motion events and fictive motion expressions) using several corpora (i.e., French novels, French, Spanish and Italian hiking descriptions and English Alpine journal) related with

the digital humanities. We first described the great interest of the concept of ENE for retrieving "complex" place names describing a static spatial context. Then, we also described how local geo-spatial information (referring to space and motion) extracted from the texts can be used for the construction and the representation of more complex geographical objects: here an itinerary. Finally, we have adapted our approach for the extraction of fictive motion expressions which can refer to both static and dynamic spatial descriptions.

The growth of digital corpora opens new perspectives regarding future work in digital humanities and more specifically developing natural language processing and data mining solutions. These few examples show the diversity of needs in the field of digital humanities but also a certain uniqueness in the way people refers to space in a static or dynamic context. This observation strengthen the idea of proposing to the community a set of tools, such as the Web services implemented in the PERDIDO platform, in order to build processing chains adapted to different tasks and to an important variety of needs.

References

Asher, N. and P. Sablayrolles (1995). A typology and discourse semantics for motion verbs and spatial PPs in french. *Journal of Semantics 12*(2), 163–209.

Aurnague, M. (2011). How motion verbs are spatial: The spatial foundations of intransitive motion verbs in French. *Lingvisticae Investigationes 34*(1), 1–34.

Blomberg, J. and J. Zlatev (2014). Actual and non-actual motion: Why experientialist semantics needs phenomenology. *Phenomenology and the cognitive sciences 13*(3), 395–418.

Boons, J.-P. (1987). La notion sémantique de déplacement dans une classification syntaxique des verbes locatifs. *Langue Française* (76), 5–40.

Borillo, M. and P. Sablayrolles (1993). The semantics of motion verbs in french. In *Proceedings of the 13th International Conference on Natural Language Processing of Avignon*, pp. 24–28.

Bubenhofer, N., M. Volk, F. Leuenberger, and D. Wüest (2015). Text+Berg-korpus (release 151_v01). XML-Format. The Alpine Journal 1969-2008.

Egorova, E., G. Boo, and R. S. Purves (2016). "the ridge went north": Did the observer go as well? corpus-driven investigation of fictive motion. In *International Conference on GIScience Short Paper Proceedings*.

Fourour, N., E. Morin, and B. Daille (2002). Incremental Recognition and Referential Categorization of French Proper Names. In *LREC 2002 Third International Conference on Language Ressources and Evaluation*, Las Palmas, Canary Islands.

Friburger, N. and D. Maurel (2004). Finite-state transducer cascades to extract named entities in texts. *Theoretical Computer Science 313*(1), 93–104.

Gaio, M. and L. Moncla (2017). 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)*.

Hickmann, M. (2006). The relativity of motion in first language acquisition. *Space across Languages: Linguistic Systems and Cognitive Categories. Amsterdam: Iohn Benjamins*,

281–308.

Jonasson, K. (1994). *Le nom propre*. Duculot, Belgique, Louvain-la-Neuve.

Kleiber, G. (1981). Problèmes de référence: descriptions définies et noms propres. *Centre d'Analyse Syntaxique de l'Université de Metz* (6), 538.

Lakoff, G. and M. Johnson (2008). *Metaphors we live by*. University of Chicago press.

Langacker, R. W. (1987). *Foundations of cognitive grammar: Theoretical prerequisites*, Volume 1. Stanford university press.

Laur, D. (1993). La relation entre le verbe et la préposition dans la sémantique du déplacement. *Langages 27*(110), 47–67.

Levinson, S. C. and D. P. Wilkins (2006). *Grammars of space: Explorations in cognitive diversity*, Volume 6. Cambridge University Press.

Makhoul, J., F. Kubala, R. Schwartz, and R. Weischedel (1999). Performance measures for information extraction. In *In Proceedings of DARPA Broadcast News Workshop*, pp. 249–252.

Matei-Chesnoiu, M. (2015). *Geoparsing early modern English drama*. Springer.

Matsumoto, Y. (1996). How abstract is subjective motion?: a comparison of coverage path expressions and access path expressions. *Conceptual structure, discourse, and language*.

Moncla, L. (2015). *Automatic Reconstruction of Itineraries from Descriptive Texts*. Ph. D. thesis, Université de Pau et des Pays de l'Adour, France.

Moncla, L., M. Gaio, T. Joliveau, and Y.-F. Le Lay (2017). Automated geoparsing of paris street names in 19th century novels. In *1st ACM SIGSPATIAL Workshop on Geospatial Humanities*, Los Angeles Area, CA, USA. ACM.

Moncla, L., M. Gaio, J. Nogueras-Iso, and S. Mustière (2016). Reconstruction of itineraries from annotated text with an informed spanning tree algorithm. *International Journal of Geographical Information Science (IJGIS) 30*(6), 1137–1160.

Muller, P. and L. Sarda (1998). Représentation de la sémantique des verbes de déplacement transitif du français. *TAL. Traitement automatique des langues 39*(2), 127–147.

Pourcel, S. and A. Kopecka (2005). Motion expression in French: typological diversity. *Durham & Newcastle working papers in linguistics 11*, 139–153.

Purves, R. S. and C. Derungs (2015). From Space to Place: Place-Based Explorations of Text. *International Journal of Humanities and Arts Computing 9*(1), 74–94.
anglais

Talmy, L. (1983). *How language structures space*. Number 4 in Berkeley cognitive science report. Berkeley, CA, Etats-Unis: Cognitive Science Program, Institute of Cognitive Studies, University of California at Berkeley.

Talmy, L. (1985). *Lexicalization patterns: Semantic structure in lexical forms. Language typology and syntactic description, vol. 3, Grammatical categories and the lexicon, ed. by Timothy Shopen, 57-149*. Cambridge: Cambridge University Press.

Talmy, L. (2000). *Toward a Cognitive Semantics*. The MIT Press.

Tenny, C. and J. Pustejowsky (Eds.) (1999). *Events as grammatical objects: the converging perspectives of lexical semantics and syntax*. Stanford, CA: CSLI.

Vandeloise, C. (1986). *L'Espace en français. Sémantique des prépositions spatiales*. Editions du Seuil.

Vandeloise, C. (1991). *Spatial prepositions: A case study from French*. University of Chicago Press.

Résumé

L'espace statique par rapport à l'espace dynamique et plus spécifiquement les descriptions spatiales, avec ou sans mouvement, sont les principales questions abordées dans cet article. Nous présentons des grammaires de construction implémentées dans la plateforme PERDIDO à l'aide de cascades de transducteurs à états finis qui visent à marquer et formaliser les relations entre entités nommées étendues, termes géographiques, relations spatiales et verbes de mouvement. Ces grammaires peuvent être considérées comme une synthèse computationnelle du travail sur l'expression de l'espace et du mouvement en langage naturel. La méthode proposée pour l'extraction d'information géographique a été appliquée pour trois projets différents au sein des humanités numériques utilisant des corpus spécifiques. La première tâche concerne l'extraction des noms de lieux à partir de romans français du XIXe siècle, la deuxième tâche traite de l'extraction de déplacements et de trajectoires à partir des descriptions de randonnées écrites en langues romanes (français, espagnol et italien) et la troisième tâche vise à identifier les expressions du mouvement fictif en anglais dans des revues alpines.