Fictive motion extraction and classification

Ekaterina Egorova, Ludovic Moncla, Mauro Gaio, Christophe Claramunt, Ross Purves

▶ **To cite this version:**

Ekaterina Egorova, Ludovic Moncla, Mauro Gaio, Christophe Claramunt, Ross Purves. Fictive motion extraction and classification. International Journal of Geographical Information Science, 2018, 32 (11), pp.2247-2271. 10.1080/13658816.2018.1498503 . hal-02139019

HAL Id: hal-02139019
https://hal.science/hal-02139019

Submitted on 24 May 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE
2018, VOL. 32, NO. 11, 2247–2271
https://doi.org/10.1080/13658816.2018.1498503

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

OPEN ACCESS Check for updates

Fictive motion extraction and classification

Ekaterina Egorova[a], Ludovic Moncla[b], Mauro Gaio[c], Christophe Claramunt[b] and Ross S. Purves[a]

[a]Department of Geography, University of Zurich, Zurich, Switzerland; [b]French Naval Academy Research Institute – IRENav, Brest, France; [c]Computer Science Department, University of Pau, Pau, France

ABSTRACT

Fictive motion (e.g. 'The highway runs along the coast') is a pervasive phenomenon in language that can imply both a static and a moving observer. In a corpus of alpine narratives, it is used in three types of spatial descriptions: conveying the actual motion of the observer, describing a vista and communicating encyclopaedic spatial knowledge. This study takes a knowledge-based approach to develop rules for automated extraction and classification of these types based on an annotated corpus of fictive motion instances. In particular, we identify the differences in the set of concepts involved into the production of the three types of descriptions, followed by their linguistic operationalization. Based on that, we build a set of rules that classify fictive motion with an overall precision of 0.87 and recall of 0.71. The article highlights the importance of examining spatially rich, naturally occurring corpora for the lines of work dealing with the automated interpretation of spatial information in texts, as well as, more broadly, investigation of spatial language involved into various types of spatial discourse.

ARTICLE HISTORY
Received 9 December 2017
Accepted 5 July 2018

KEYWORDS
Fictive motion; automated extraction; automated classification; knowledge-based approach

1. Introduction

The emergence of large corpora of digitized texts and user-generated content has fuelled both thematic and methodological advances in Geographic Information Science, echoing developments in Digital Humanities, where qualitative methods and theories are combined with computer-based approaches that include natural language processing (NLP) and data mining (Piotrowski 2012). On the one hand, large corpora enable the exploration of research questions related to geography in ways which were previously only possible through empirical experiments (Xu *et al*. 2014, Derungs and Purves 2016). On the other hand, working with such data requires the development of methods for the extraction and analysis of specifically geographic content from texts (Vasardani *et al*. 2013, Moncla *et al*. 2016, Kim *et al*. 2017).

To date, research has often focused on extracting locations from text and linking them with thematic and/or temporal information, with a wide variety of related tasks: for example, modelling regions associated with vernacular names (e.g. Davies 2013), examining differences in spatial language use (e.g. Xu *et al*. 2014), deriving spatial

CONTACT Ekaterina Egorova ekaterina.egorova@geo.uzh.ch

© 2018 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

folksonomies of landscape elements (e.g. Derungs and Purves 2016), and identifying and mapping hazard-related events (e.g. Wang and Stewart 2015). In most of this work, unstructured text is essentially treated as a bag of words: locations are linked to their properties through relatively simple (though often effective) approaches such as the analysis of word co-occurrences (e.g. Davies 2013, Bruggmann and Fabrikant 2016). Though such methods are a good starting point for extracting broad overviews of variations in space and time, they are less suited to the extraction of spatially nuanced information, e.g. reconstructing static spatial scenes or dynamic itineraries from linguistic descriptions.

In this article, we set out to make a thematic and methodological contribution by researching a particularly geographic use of language: that of *fictive motion*. Fictive motion (FM) is an example of *figurative* use of language, where the concepts encoded in an utterance cannot be interpreted literally. Thus, (1a)[1] does not imply movement of the valley itself but rather encodes information about the valley's spatial extendedness and location with respect to the sea. Furthermore, FM can be used to indirectly describe actual motion of an observer as in (1b), where motion as represented linguistically is fictive (since the highway is not moving) but implies the movement of the observer:

(1) a. The ***valley ran*** towards the sea.
b. The ***highway ran*** along the sea for twenty minutes.

Interpreting FM correctly is not trivial, but as Pustejovsky (2017, p. 992) eloquently states, 'motion can… have a lasting effect on the interpretation of a text with respect to spatial information'. This in turn has important implications for research efforts which attempt to extract spatial information from unstructured text.

One way of dealing with nuanced spatial language such as FM is the development of sets of rules, typically driven both empirically by the textual analysis and by a priori knowledge of the patterns being sought. The goal is usually to develop rules which are effective in a given corpus (but also robust across corpora with similar properties), as well as gain insights into the particular patterns under study. Examples of the use of such approaches include work from Moncla *et al*. (2016) to reconstruct trajectories from text sources and from Vasardani *et al*. (2013) to derive depictions of the environment from verbal descriptions. Underpinning this line of work are efforts at developing annotation schemes for capturing spatial information in texts (Pustejovsky 2017) and, more broadly, work on spatial semantics (Talmy 2000, Langacker 2005) and representation of spatial language and concepts in formal models (Zlatev 2007, Bateman *et al*. 2010, Tenbrink *et al*. 2013).

Despite awareness of metaphoric mappings from spatial language, its vagueness and situated interpretation (Herskovits 1987, Lakoff and Johnson 2008, Bateman *et al*. 2010), methods to extract concepts such as location and motion often implicitly take the standpoint that spatial language can be interpreted in a relatively straightforward way. There are two key reasons for this. First, the approaches taken have been broadly successful. Secondly, to date, many researchers in this domain have concentrated on text sources which are factual, for example, in the form of newspaper articles, scientific papers or route descriptions. In practice, however, even such texts use a wide variety of figurative linguistic structures both at the level of individual words (e.g. polysemy or

metonymy) and in discourse (e.g. irony or metaphor) (Langacker 2005, Lakoff and Johnson 2008).

We suggest that understanding, extracting and classifying phenomena such as FM are important not only at a conceptual and theoretical level but also in a pragmatic sense, as it should result in better insights on the semantics of spatial information in text. Thus, such research lies at the core of GIScience. Our article presents an interdisciplinary study that combines approaches from linguistics, computer science and GIScience to develop and explore a set of rules for the extraction and classification of FM from a text corpus.

The article is structured as follows. In Section 2, we outline the state of art in the area of spatial information annotation and extraction, and provide a theoretical basis for the phenomenon of FM and its types (static and dynamic), followed by the problem statement. The data and general methodology used are discussed in Section 3. Section 4.1 focuses on the process of developing rules for FM extraction, while Section 4.2 focuses on its classification by describing spatial language and concepts characteristic of each type of FM. The implementation of rules in Unitex[2] and their evaluation performance on the training and validation data, and types of errors are presented in Section 5, followed by the conclusion in Section 6.

2. Related work

Two areas are central to our work. Research dealing with annotation and extraction of spatial information is important since it demonstrates the level of complexity with which spatial concepts are modelled in the state of the art. Investigations of FM in cognitive linguistics provide the necessary theoretical description of the linguistic phenomena of our study.

2.1. *Annotation and extraction of spatial information from text*

Extraction of spatial information from text relies on contributions from a number of related research directions including the development of annotation schemes (which spatial concepts are found across various corpora?) and the application of methods to automatically extract spatial information from text (how can we automatically extract spatial concepts from corpora?).

The tradition of annotation schemes is relatively long and can be related and traced back to the emergence of domain-specific languages, targeted at specific application areas and thus having an 'expressive power focused on a particular problem domain' (Van Deursen *et al*. 2000, p. 26). Development of domain-specific languages set central guidelines for acquiring knowledge and modelling specific domains and unveiled challenges, such as balancing the level of abstraction and expressiveness (Van Deursen and Klint 2002, Kosar *et al*. 2010). One example is Text Encoding Initiative (TEI),[3] a standard for the representation of texts in digital form commonly used in the humanities, social sciences and linguistics. Originating in the 1980s, the current version TEI P5 describes and defines nearly 500 well-documented textual distinctions, including a module for encoding names, dates, people and places (Burnard 2007).[4] Such encoding schemes require clear definitions that can be interpreted and followed by annotators, which can

be hard to achieve especially as annotation schemes become very rich. Thus, in Message Understanding Conferences (MUC) — another initiative, also dating back to the 1980s — annotators worked on corpora in parallel with the aim of populating given templates and comparing standard metrics, such as precision and recall (Grishman and Sundheim 1996). While formulating guidelines proved straightforward for named entities, formulating complete and consistent guidelines was remarkably difficult for some other elements like coreference (Grishman and Sundheim 1996).

Spatial markup languages or annotation schemes and ontologies are developed with the specific aim of enriching text with spatial annotations that go beyond locations. The GUM-Space ontology is an extension of the Generalized Upper Model used in NLP applications (Bateman *et al*. 2010). Grounded in grammar and spatial semantics, it does not cover elusive concepts represented by structures such as FM. Spatial markup languages such as SpatialML and ISO-Space, on the other hand, though mapped onto GUM-Space ontology, are typically based on the examination of real texts. SpatialML allows annotation of toponyms and other nominal references to place, as well as topological and orientational relations (Mani *et al*. 2010). ISO-Space (Pustejovsky 2017) adds motion into the annotation scheme, introducing corresponding tags and attributes (e.g. *source, goal* and *mover*). The development of such schemes is a highly iterative process characterized by challenges including the richness of spatial language, its potential variance across genres (dictating the need for the examination of diverse corpora, e.g. newswire, travel and blogs), multiple use cases for annotation schemes and the complexity and achievability of the annotation task (Pustejovsky and Moszkowicz 2012, Pustejovsky *et al*. 2012, Pustejovsky 2017). Grounded in the examination of naturally occurring discourse, such schemes exhibit more awareness of semantically nuanced spatial concepts. Thus, Pustejovsky and Yocum (2013) propose the *motion sense* attribute for different interpretations of *motion events*, including, but not limited to, *literal* and *fictive* values.

Automated extraction of spatial concepts is another distinct task. Here, apart from the extraction of references to locations in the form of toponyms, research has focused on two other important subdomains of spatial language – location (in the form of spatial relations between objects) and motion (Levinson 2003).

In the above-mentioned MUC, an important task was the extraction of named entities, including toponyms, and basic capabilities were developed quickly (Chinchor and Robinson 1997, Mikheev *et al*. 1999). Since then, a variety of methods from simple heuristics and gazetteer lookup to machine learning approaches have been applied to the problem of identifying and disambiguating toponyms in Geographic Information Retrieval; however, certain challenges (e.g. use of metonymic language and domain language diversity) still await solutions (Leveling and Hartrumpf 2008, Gritta *et al*. 2018).

Automatically extracting locations in the form of spatial relations (e.g. 'The vase is on the table') has also gained attention recently. Kordjamshidi *et al*. (2011) describe the Spatial Role Labelling task: by applying machine learning and using a set of features including lemmas, part-of-speech tags and semantic role labels, they assign roles to linguistic structures representing spatial concepts (e.g. *trajectors, landmarks, spatial indicators, motion indicators*). Related to this work is the parser developed by Khan *et al*. (2013), which relies on spatial prepositions and dependency roles to identify triplets representing locative expressions. *Motion events* and their extraction are

addressed by Moncla *et al*. (2016), who perform itinerary reconstruction using texts annotated with geo-semantic tags. The tags capture motion verbs, spatial relations in the form of prepositions and spatial named entities. The method combines quantitative and qualitative criteria based on information from the annotated text and data from geographic databases.

Both the development of spatial markup languages and the extraction of spatial concepts from text are grounded in spatial semantics, which explores the meaning of spatial language. On the one hand, it equips related research with key spatial semantic concepts such as *figure*,[5] *ground* or *motion event* and their encoding in language(s) (Langacker 1987, Talmy 2000). On the other hand, it examines the role that space plays in thinking, which is reflected in the pervasiveness of space-related metaphors (Lakoff and Johnson 2008), polysemy of spatial terms (Zlatev 2007) and, generally, figurative use of much of spatial language (Langacker 2005). FM, in particular, is a vivid example of our cognitive bias towards dynamism, dictated by our basic experience of moving in space (Talmy 2000). Automated interpretation of such phenomena and extraction of complex concepts, such as those encoded in the ISO-Space distinction between *literal* and *fictive* motion, has, however, proved elusive because they have been regarded as either too esoteric or too challenging given the current state of the art.

2.2. *Fictive motion and its types*

Against a common assumption that language describes events and situations primarily directly, much of our linguistic effort goes into the description of *virtual* entities, even if our concern is with actual ones. This results in systematic and fundamental departures from the direct description of *actuality* (Langacker 2005). FM is a linguistic structure that represents a static object as moving and is thus a vivid example of such a departure. To consider (2), our common sense representation of a highway is that of a linear, static object; the fictive representation encoded in the literal meaning of the sentence, though, is the one of the highway moving along its axis:

(2) The ***highway runs*** along the coast.

What makes FM specific is the fact that it can represent different types of scenes, with the essential difference attesting to the presence or absence of the actual motion of the observer. The use of these two different types of scenes is also motivated by different cognitive processes (Matsumoto 1996b, Langacker 2005).

Actual motion. FM exemplified in (3a–b)[6] is based on the actual motion of the observer at a particular time and describes the entity as experienced by him or her. The motion component here has an experiential basis: a series of immediate fields of view of the moving observer are fictively construed as a single entity, experienced as moving through space itself (Langacker 2005):

(3) a. The ***road went*** up the hill. (as we proceeded)
b. This ***highway will enter*** California soon. (uttered by the driver)

Non-actual motion. FM based on non-actual motion, exemplified in (4a–b), represents a static description of an object whereby the actual motion of any entity is

completely absent. In cases like (4a), motion is purely fictive and is motivated by visual (or mental) scanning along the spatial entity (Langacker 2005). Cases like (4b) are based on 'hypothetical motion of an arbitrary entity at an arbitrary time' (Matsumoto 1996b, p. 361). The difference between the two subtypes is quite subtle. On the one hand, the *figure* in hypothetical motion FM has to be travellable (associated with a path-like entity, e.g. 'road', 'path'). On the other hand, only hypothetical motion allows the presence of *motion duration*, which becomes apparent in examples (4c–d) – the asterisk * indicates non-acceptability in (4c):

(4) a. The ***mountain range goes*** from Canada to Mexico. (mental scanning)
b. The ***highway enters*** California there. (hypothetical motion)
c. *The ***mountain range goes*** along the coast for some time. (mental scanning is incompatible with duration)
d. The ***highway runs*** along the river for a while. (hypothetical motion is compatible duration)

2.3. *Fictive motion types in an alpine corpus*

Egorova *et al*. (2018) investigated the way FM is used in natural spatial discourse, specifically in the 'Text + Berg' corpus consisting of 1484 texts (6,356,455 words) from the digitized Alpine Journal between 1968 and 2008 (Bubenhofer *et al*. 2015). First, based on mountaineering thesauri, the authors compiled a list of nouns representing spatial entities including both structural entities physically present in the environment (encoded by landscape terms, such as 'ridge') and functional entities that refer to behavioural patterns – nouns such as 'route', 'approach' and 'pitch' (Klippel 2003). Further, the corpus was queried for a noun representing a spatial entity followed by a verb, which produced 6530 candidate phrases. Manual inspection of candidate phrases resulted in a corpus of 981 FM instances, which was further investigated with respect to the presence of the two types of FM described above. Both types were found in the corpus; furthermore, the non-actual motion was found to be represented by two further subclasses – vista and encyclopaedic spatial knowledge. In the following, we provide definitions and examples of the use of FM.

Actual motion. This type encodes the actual motion of the mountaineers, as in (5a). Such cases are encountered in reports of completed ascents and represent descriptions of motion along a particular route segment:

(5) a. The latter ***part of the route crossed*** a steep snow slope.

Non-actual motion (vista). This type represents a description of a view encountered by a mountaineer and is also found in reports of completed ascents. While cases such as (6a) can be clearly linked to visual scanning, instances of the type (6b) also fit the definition of a hypothetical motion. The two examples make it clear that the dichotomy (visual scanning vs hypothetical motion) is not entirely unproblematic; although (6b) is grounded in visual scanning, the travellability of 'route' refers strongly to hypothetical motion. Neither Matsumoto (1996b) nor Langacker (2005) addresses such fuzzy cases.

Although the cognitive motivation (visual scanning or hypothetical motion) is of little importance to us, cases related to hypothetical motion are especially difficult to classify, since they can include markers associated with a prototypical actual *motion event*.

(6) a. Far off, a great red ***buttress rose*** steeply.
b. Realizing our ***route traversed*** much further to the right, I left a nut and carabiner behind and lowered myself to the tiny ledge.

Non-actual motion (encyclopaedic knowledge). This type encodes spatial knowledge about a larger geographic area, as in (7a), and often occurs at the beginning of a narrative and provides the general setting for the story. Alternatively, it can describe the general layout of a route, as in (7b). While the former type is grounded in mental scanning, the latter, again, has a strong tie to hypothetical motion:

(7) a. The ***range runs*** east and west and contains numerous peaks around 5700–5800 m.
b. The ***route takes*** the obvious pillar rising above the Charpoua glacier.

According to Egorova *et al*. (2018), the three major types of spatial descriptions encoded by FM differ in the spatial language used and in most cases can be distinguished based on the markers. However, in the absence of clear markers, only a broader context can disambiguate if the actual motion is present. This is particularly true for instances that can represent both hypothetical and actual motion. In (6b) above, 'route traversed' can be interpreted as actual motion. Only the second clause (describing the observer lowering herself to the ledge) indicates that the motion is hypothetical. Nonetheless, the reliability measure Krippendorff's alpha was found to be 0.802, which represents substantial agreement (Egorova *et al*. 2018).

We refer to the three types of FM discussed as 'actual motion FM', 'vista FM' and 'encyclopaedic knowledge FM' throughout this article.

2.4. *Research gaps and problem statement*

Interpretation of motion in text has a lasting effect on the interpretation of spatial information in general (Pustejovsky *et al*. 2012). Given the pervasiveness of FM in language, the ability to identify and classify its instances is an important step towards automated text understanding. As outlined above, ISO-Space now includes the *motion sense* tag, with *fictive* as one of its attributes, but overlooks the fact that FM can actually encode the motion of the observer. A specification of the conceptual inventory behind FM describing static scenes and those referring to the motion of the observer would contribute significantly to this line of work. Further, annotation schemes often remain at the level of abstract syntax (Pustejovsky and Yocum 2013) and the reader gets redirected to linguistically-grounded research (e.g. Bateman *et al*. 2010) for the potential linguistic encoding of the concepts. Egorova *et al*. (2018), while reporting on the types of scenes encoded by FM and corresponding markers, do not provide an exhaustive account of the structure and systematicity of the latter. There is thus also a need to link relevant concepts to their linguistic encoding. Working with the corpus of annotated

FM provided by Egorova *et al*. (2018), we aim to address these gaps by answering the following questions:

I. What are the concepts that capture the differences between the types of scenes described by FM?
II. To which extent are these concepts linguistically operationalizable? In other words, to which extent can the extraction and classification of FM be performed automatically?
III. How systematic are the markers? In other words, how many structures remain unclassified because of the absence of clear markers and a need of a broader context?

3. Data and methodology

We use two sets of data: a pre-processed (lemma- and part-of-speech-tagged) corpus of alpine texts 'Text + Berg' (Bubenhofer *et al*. 2015) and a corpus of sentences with FM from 'Text + Berg'. Sentences with FM were extracted semi-automatically and classified manually into actual motion, vista and encyclopaedic knowledge FM (Egorova *et al*. 2018). We treat this FM corpus as a gold standard.

The two sets of data were split into two subcorpora, one for training and one for validation. Texts written between 1969 and 1991 and a corpus of FM from those texts were used in the iterative training process of developing and refining rules for the extraction and classification of FM. Texts written between 1992 and 2008 and a corpus of FM from those texts were used for validation. The training data represent 543 cases of FM, and the validation data represent 442 cases of FM. Since we assume that FM as a linguistic structure is a thought-related phenomenon rather than a literary device, we also assume that it is not susceptible to changes in discourse over time.[7]

Out of 543 FM in the training data, 179 (33%) instances represent actual motion FM, 107 (19.7%) represent vista FM and 257 (47.3%) represent encyclopaedic knowledge FM. Similar orders of magnitude usage of FM are found in the validation data: 136 (30.8%) instances of actual motion FM, 98 (22.1%) instances of vista FM and 208 (47.1%) instances of encyclopaedic knowledge FM.

Methodologically, there are two main types of approaches to such tasks—data-driven and knowledge-based, and both are common in Information Extraction (IE) and NLP. Data-driven approaches are based on deep learning algorithms or machine learning methods (supervised or unsupervised) such as Hidden Markov Models, Maximum Entropy Models, Conditional Random Fields and Decision Trees. Knowledge-based approaches rely on heuristics and syntactic-semantic patterns (or ad hoc rules) developed manually by experts. Data-driven approaches require large volumes of annotated training data, while knowledge-based approaches focus on domain expertise and typically require less training data.

In this work, we used a knowledge-based approach grounded in linguistic expertise which allowed us to iteratively use domain expertise and improve our understanding of the phenomena being analysed. We described labels (based on lexicons) and linguistic

rules (based on morpho-syntactic patterns) for identifying and classifying FM and implemented labels and rules using a finite-state transducers cascade.

In particular, we first developed rules for the extraction of FM from text. We classified false positives into four groups (e.g. polysemy) and iteratively developed a set of rules for dealing with them based on the training data. For the classification of extracted FM, we started with the markers outlined by Egorova *et al*. (2018). In the process of an iterative inspection of the training data and relevant literature, we first identified a system of concepts crucial for the differentiation between the types of FM (e.g. type of frame of reference). Further, we operationalized these concepts in terms of corresponding linguistic structures, relying on research in spatial language but also thesauri (e.g. expanding potential linguistic encodings of a concept through synonyms or words in the same semantic field). Finally, we transferred this conceptual and linguistic inventory into the syntax of Unitex in the form of labels and rules. At the linguistic level, we are working with lemmas, lexemes (words) and part-of-speech tags. Combinations of these lie at the heart of rules developed. In general, we follow Bunt and Pustejovsky (2010) in making a distinction between abstract syntax (concepts) and concrete syntax (labels and rules representing concepts in Unitex).

4. Identification and classification of fictive motion

4.1 *Extraction of fictive motion*

Based on the definition of FM, the major criteria for identifying FM was the presence of a motion verb related to a spatial entity as a subject, as in the prototypical 'The ridge runs north'. Thus, our starting rule for the extraction of FM is the presence of a spatial entity followed by a motion verb, based on the lists from Egorova *et al*. (2018). Applying this simple rule naturally retrieved a large number of false positives that could be classified into several types and reduced by additional rules.

Polysemy. One of the triggers of false positives is polysemy of basic motion verbs that often have further senses (Ravin and Leacock 2000, Zlatev 2007). We identified and operationalized contextual elements of non-motion senses of four verbs that particularly influenced the precision of FM identification: 'get', 'turn', 'take', 'follow'. Examples of disambiguation rules include: 'turn' and 'get' followed by an adjective for the change of state (e.g. 'the ice turned solid'); or 'take' followed by a time-related lemma (e.g. 'minute', 'hour', 'day') within a larger right window to discard cases such as 'these pitches took me all day to lead'.

Complex noun phrases. Another type of false positives is represented by cases where the noun denoting a spatial entity is part of a complex noun phrase representing the subject, as in (8a). Since the noun phrase can also represent a spatial entity, see examples (8b–d), we have to identify the subject correctly. Our algorithm searches for a preposition in the left window of the FM structure.[8] The presence of a preposition – e.g. 'along' in (8a), 'on' in (8b) – signals a locative phrase; to find the main noun, the algorithm searches further left for a lemma referring to a spatial entity for cases like (8b), spatial part (e.g. 'part', 'side', 'edge', 'bottom') for cases like (5a), distance (e.g. 'feet', 'ft', 'mile(s)', 'm', 'meter(s)') for cases like (8c), collection (e.g. 'series', 'succession', 'couple', 'row', 'system', 'maze', 'tangle',

'most', 'several', 'some') for cases like (8d). Spatial entity, spatial part, distance and collection are the four types of subjects encountered in the FM corpus. The search is conducted through lists compiled for each of these types based on the training data and synonyms derived from thesauri.

(8) a. A climb along this ***ledge brought*** us to the second rock barrier.
b. The route on the rock ***face led*** diagonally upwards.
c. Haifa (*sic*!) a mile along the ***ridge brought*** us to a pleasant clearing.
d. Here, a succession of limestone ***cliffs rises*** sheer from the sea.

Factive motion. Some false positives represent 'factive motion' (Talmy 2000), mostly referring to actually falling entities, as in (9a). We capture these through a list of 'movable' entities (e.g. 'boulder(s)', 'rock(s)', 'stone(s)', 'flake(s)', 'snow', 'ice') followed by one of the verbs (e.g. 'fall', 'drop', 'come down', 'descend', 'come', 'roll', 'run', 'stop', 'rush', 'land'). A separate rule is created to capture the cases of 'glacier' followed by 'move' and 'descend'.

(9) a. I was about 20 ft away from the rock when great tons of ***ice came*** down.

Errors in part-of-speech tagging. Finally, another frequent type of false positives pertains to errors in part-of-speech tagging, as in (10a), where 'climbs' is tagged as a verb based on the Penn Treebank, used for the part-of-speech tagging of the corpus (Marcus *et al*. 1993, Bubenhofer *et al*. 2015). A close inspection unveiled the systematic presence of two types of error, which could be excluded based on simple heuristics: noun phrases describing the type of a climb (e.g. 'rock climbs') as well as noun phrases with a past participle (e.g. 'peaks climbed by the team', 'mountains ascended this year').

(10) a. These are now looked upon as probably the finest ***rock climbs*** in Kenya.

4.2. *Types of fictive motion and conceptual inventory behind*

As outlined above, the three types of FM represent fundamentally different spatial descriptions: actual motion of the observer, description of a vista somewhere along the way or encyclopaedic knowledge. These three types of descriptions differ from several perspectives. From the perspective of dynamism, the first type encodes actual motion, while the other two are essentially static, grounded in visual and mental scanning – although some may also evoke hypothetical motion, as we have seen (Matsumoto 1996a). From the perspective of scale, the three types of descriptions relate to environmental, vista and geographic scale classes, respectively (Montello 1993). Furthermore, actual motion FM can be treated as part of route descriptions, encyclopaedic knowledge descriptions take the bird's eye or survey perspective, while vista FM can be related to scene descriptions (Taylor and Tversky 1992, Denis *et al*. 1999, Pustejovsky 2017). These differences should be reflected in the way spatial information is treated and linguistically encoded.

Encyclopaedic knowledge FM structures convey 'permanent state' (or, 'permanent truth') and are thus encoded by verbs in the simple present tense. This resonates well with Langacker's claim that the simple present tense is diagnostic of non-actual motion

FM (Langacker 2005, p. 176). It is thus easy to identify encyclopaedic knowledge FM based on the present tense, while structures in the past tense can represent either actual motion or vista FM.[9]

The difference between the vista and actual motion FM is mirrored in the difference in spatial concepts involved (and thus spatial language), which can be used for developing rules for their differentiation. In what follows, we first discuss these two types of FM vista and actual motion separately, unveiling the specifics of the concepts involved into their production. Further, we describe corresponding labels and their linguistic operationalization, followed by the summary of classification rules.

4.2.1 *Vista FM*

Vista FM describes a static view from a point in space which the observer is occupying.

What makes such cases unique is, first, the *specific use of frames of reference* (Levinson 1999). On the one hand, vista FM can be marked by relative frames of reference where the observer is explicitly mentioned as a reference point, as 'below us' in (11a). On the other hand, only vista FM exploits absolute frames of reference that indicate *location* (answering the question 'where?') as opposed to *direction* (answering the question 'where to?'), characteristic of actual motion FM. Linguistically, the absolute frame of reference positioned to the left of FM can only encode location, as in (11b); cases such as (11c) are ungrammatical. Thus, an absolute frame of reference to the left of the FM structure refers to location and signals a vista FM.

(11) a. The main Shani ***glacier flowed*** below us to the south-east.

b. To the west, steep pine clad ***slopes rise*** up to Tukche peak.

c. ?? To the west, the ***route led*** us.

Another concept that is helpful to differentiate between vista and actual motion FM is *untravellability*. As outlined in Section 2.2, *figures* in actual motion FM must be travellable. Hence, an untravellable *figure* in our case always signals a vista. In linguistic literature, the term *untravellable* is used to describe linear objects that are normally not associated with human motion, such as walls, telephone lines, wires (Matsumoto 1996a, Rojo and Valenzuela 2009) – an easy differentiation, given the introspective nature of most studies. In our space-specific context, this concept is inherently linked to the spatial properties of a *figure*.

Thus, some spatial entities cannot represent route segments because of their geometry, e.g. 'peak' is associated with a point, as a rule. Similarly, large-scale spatial entities, e.g. 'massif', are highly unlikely to represent route segments. The concept of a large-scale entity is of course also arbitrary and is furthermore sensitive to plurality. Thus, it is difficult to project a route segment onto a massif, whether in singular or in plural. Nouns such as 'valley' or 'ridge', on the other hand, can represent route segments when in singular, compare examples (12a) and (12b):

(12) a. The ***ridge went*** on, now straightforward.

b. Rotondo's ***ridges curled*** in an embrace and covered us with their excess vapours.

Untravellability can be also constructed through the semantics of other linguistic features – verbs, adjectives, adverbs with strong connotation, especially in those vista descriptions that convey the sense of place (Egorova *et al*. 2016). In examples (13a–b), the strong connotation of 'precipitously' and 'plunged' excludes the possibility of FM to represent motion of the observer:

(13) a. The ***ridge dropped*** away precipitously.
b. Cathedral ***wall plunged*** to the west, while to the east the plateau extended.

Finally, according to our observations, vista descriptions can include an explicit reference to the *process of observation*, represented by a variety of related lexemes, as in (14a):

(14) a. At close quarters we could see that the lower ***part of the ridge rose*** in four steps.

4.2.2. *Actual motion FM*

Since this type of FM refers to the motion of the observer, it can be marked by elements of a typical *motion event*.

One such element is an explicit reference to the *moving entity (observer)*. In (15a), 'us' excludes the possibility of FM to be a vista. Further, the *figure* has to represent a *travellable* entity as mentioned above; an untravellable entity would signal a vista FM. According to our observations, functional entities (e.g. 'path', 'route') are mostly encountered in this type of FM and can thus serve as a marker of actual motion. Another element characteristic of this type of FM in the corpus is the *difficulty* of a route segment, as in (15b). Finally, *speed*, as exemplified in (15c), or *motion duration*, as exemplified in (15d), marks actual motion FM.

(15) a. The ***couloir debouched*** us on the flattened summit.
b. Below the steep pitch a rather easy ***ridge took*** us to a snow-covered gully.
c. Better acclimatized now, the ***route went*** quickly.
d. A gentle ***slope brought*** us in one hour to the main summit.

Again, a note should be made that these elements can also be encountered in vista FM in case the observer is describing a hypothetical motion, as in (6b) above in Section 2.3. However, in the training data, such cases are rare in comparison to those describing actual motion.

4.3 *Labels, their linguistic operationalization and rules*

To automatically classify FM into the three types, we examined the way concepts differentiating them are encoded linguistically. That allowed us to recognize and label them in text (see Table 1) and implement associated rules:

Table 1. Concepts/labels and their function.

Label	Type of FM
FUNCTIONAL ENTITY	actual motion
MOVING OBSERVER	actual motion
MOTION DURATION	actual motion
DIFFICULTY	actual motion
SPEED	actual motion
UNTRAVELLABILITY	vista
LARGE-SCALE ENTITY	vista
VISTA FRAME OF REF.	vista
VISION	vista

- FM with a verb in the present tense represents encyclopaedic knowledge.
- FM with a verb in the past tense and a label referring to difficulty, speed, motion duration, a moving observer or a functional entity represents actual movement of the observer.
- FM with a verb in the past tense and a label referring to a vista frame of reference, a large-scale entity, untravellablility or vision represents a description of a vista.

FUNCTIONAL ENTITY label annotates the following lexemes: 'route', 'line', 'approach' (if a noun), 'road', 'trail', 'traverse' (if a noun), 'pitch(es)'.

MOVING OBSERVER label captures personal pronouns in the accusative (e.g. 'us', 'me') as well as lexemes 'party', 'team' to the right of the FM structure and possessive pronouns (e.g. 'our', 'my') to its left.

MOTION DURATION label captures time-related nouns (e.g. 'day', 'afternoon', 'hour'), time-related phrases (e.g. 'on and on', 'forever', 'for some way', 'for a while').

DIFFICULTY label captures both qualitative and quantitative references to the difficulty of motion. Qualitative references are expressed by adjectives (e.g. 'easy', 'easier', 'hard', 'pleasant', 'bad'), adverbs and adverbial phrases (e.g. 'easily', 'without difficulty', 'free'). Quantitative references represent the numerical difficulty scale and are operationalized as a sequence of the preposition 'at' and a number between 1 and 9.

SPEED label captures adjectives (e.g. 'quick', 'fast', 'slow') and adverbs (e.g. 'quickly', 'slowly').

UNTRAVELLABILITY label captures a number of linguistic features constructing space as untravellable: adverbs (e.g. 'vertically', 'abruptly', 'sheer', 'precipitously', 'madly', 'magnificently', 'away', 'off', 'skywards'), adjectives (e.g. 'enormous', 'huge', 'massive'), verbs that are semantically incompatible with human motion ('soar', 'sweep', 'emerge', 'rear', 'sprawl', 'shoot').

LARGE-SCALE ENTITY label annotates the following lexemes: 'mountain', 'mountains', 'massif', 'massifs', 'faces', 'walls', 'glaciers', 'routes', 'cliffs', 'ridges', 'hills', 'valleys', 'channels'.

VISTA FRAME OF REFERENCE label annotates, firstly, relative frames of references that explicitly mention the observer and are operationalized through spatial prepositions corresponding to the observer-related axial structure ('above', 'below', 'beneath', 'under', 'underneath', 'in front (of)', 'before', 'behind') (Landau and Jackendoff 1993), followed by a personal pronoun in the accusative or a possessive pronoun to capture cases like 'above us' or 'beneath my feet'. Another combination is 'on' followed by either a possessive pronoun or a definite article and 'left' or 'right' to capture cases such as 'on

my right' or 'on the left'. Secondly, this label captures absolute frames of references positioned to the left of FM and describing location (as opposed to direction, as outlined in Section 4.2) based on the list of corresponding terms (e.g. 'south', 'south-east').

VISION label captures lemmas related to the process of observation: 'see', 'appear', 'look', 'seem', 'glance', 'view', 'watch', 'detect', 'examine', 'notice', 'observe', 'identify', 'spot', 'glimpse', 'sight', 'stare', 'distant'.

Some important notes should be made concerning the set of concepts and their linguistic operationalization. First, some of the concepts could be deconstructed and specified further. For example, *untravellability* could be decomposed into *verticality* and *spatial extension*. For our specific problem, however, we found this balance between abstraction and expressiveness optimal. Second, concepts vary in the degree of their vagueness. While *moving observer* is an example of a rather crisp concept, *untravellability* is exceptionally vague in the context of alpine space and mountaineering. Finally, from the perspective of linguistic representation, we have, on the one hand, concepts that are linked to Talmy's closed class (Talmy 2000), e.g. *vista frame of reference* ('beneath our feet'). On the other hand, there are concepts represented by the open class, e.g. *vision*, which are difficult to operationalize exhaustively.

5 Implementation and evaluation

5.1 *Automated annotation of fictive motion*

To implement the rules, we used the Unitex platform. Unitex grammars are variants of context-free grammars and recursive transition networks. These grammars contain the notion of transduction derived from the field of finite state automata, enabling a transducer (i.e. a grammar) to produce some output. Essentially, transducers are graphs that make annotations, replacements and deletions in texts, exploiting morpho-syntactic patterns and lexicons (Abney 1996). Importantly, they can be used in a cascade (Friburger and Maurel 2004) where transducers can use annotations performed by previous transducers in the chain. Unitex thus provides a highly intuitive and powerful framework for implementing shallow parsers (i.e. recognizing and annotating certain segments in texts) and building grammars (i.e. adding further levels of annotation based on the outputs of parsers), which is optimal for tasks similar to ours.

Since the 'Text + Berg' corpus already contains part-of-speech and lemma information (Figure 1), a preprocessing step consists of extracting text content, part-of-speech tags and lemmas from the XML files of the 'Text + Berg' corpus and transforming them into a format compatible with Unitex.

Our approach performs two tasks: segmentation and classification. The segmentation recognizes FM structures, while the classification assigns to them a corresponding label (i.e. a class). The cascade executes nine main transducers in a specific order, for both the segmentation and classification of FM structures (Table 2).

The order of transducers in a cascade is crucial. On the one hand, each transducer can use annotations added by previous transducers; on the other hand, tokens included in annotated patterns cannot be recognized by the following transducers. This determines the crucial role of the order in which rules are executed. The following order was found optimal for the segmentation and classification of FM.

```
<s lang="en" n="a0-s2040">
  <w lemma="our" n="a0-s2040-w1" pos="PP$">Our</w>
  <w lemma="propose" n="a0-s2040-w2" pos="VVN">proposed</w>
  <w lemma="route" n="a0-s2040-w3" pos="NN">route</w>
  <w lemma="follow" n="a0-s2040-w4" pos="VVD">followed</w>
  <w lemma="a" n="a0-s2040-w5" pos="DT">a</w>
  <w lemma="very" n="a0-s2040-w6" pos="RB">very</w>
  <w lemma="steep" n="a0-s2040-w7" pos="JJ">steep</w>
  <w lemma="mixed" n="a0-s2040-w8" pos="JJ">mixed</w>
  <w lemma="line" n="a0-s2040-w9" pos="NN">line</w>
  <w lemma="." n="a0-s2040-w10" pos="SENT">.</w>
</s>
```

Figure 1. Example of a sentence extracted from the "Text + Berg" corpus.

Table 2. Transducers of the cascade.

Order	Transducer
1	factive motion, errors in POS
2	spatial entities
3	first classification
4	polysemy
5	motion verbs
6	complex noun phrases
7	fictive motion
8	second classification
9	third classification

Transducer 1 annotates sequences representing the two types of false positives factive motion (see example (9a)) and structures that often contain part-of-speech tagging errors (see example (10a)). This prevents the next transducers from annotating these types of false positives. Transducer 2 annotates spatial entities (landscape terms as well as functional entities such as 'pitch') based on the list used by Egorova *et al*. (2018). It also adds a label to the annotation to make a distinction between (unmarked) entities, large-scale entities and functional entities. These annotations are used by the following transducers. Transducer 3 implements the first step of classification based on the semantics of a verb (verbs signalling untravellability, e.g. 'soar'). Transducer 4 deals with polysemy in order to avoid false positives as described in Section 4.1. Transducer 5 annotates motion verbs based on the list used by Egorova *et al*. (2016). Transducer 6 implements heuristics to exclude false positives related to complex noun phrases where the spatial entity is not the subject, as described in Section 4.1. Transducer 7 annotates FM structures taking into account the type of spatial entity (i.e. unmarked, large-scale, functional) and the tense of the verb. Transducers 8 and 9 refer to the second and third classification steps of the cascade. They implement all the heuristics described in Section 4.2. To illustrate the way linguistic rules are implemented within the Unitex platform using graphs, Figure 2 shows a simplified version of the transducer performing the second and main step of the classification process.

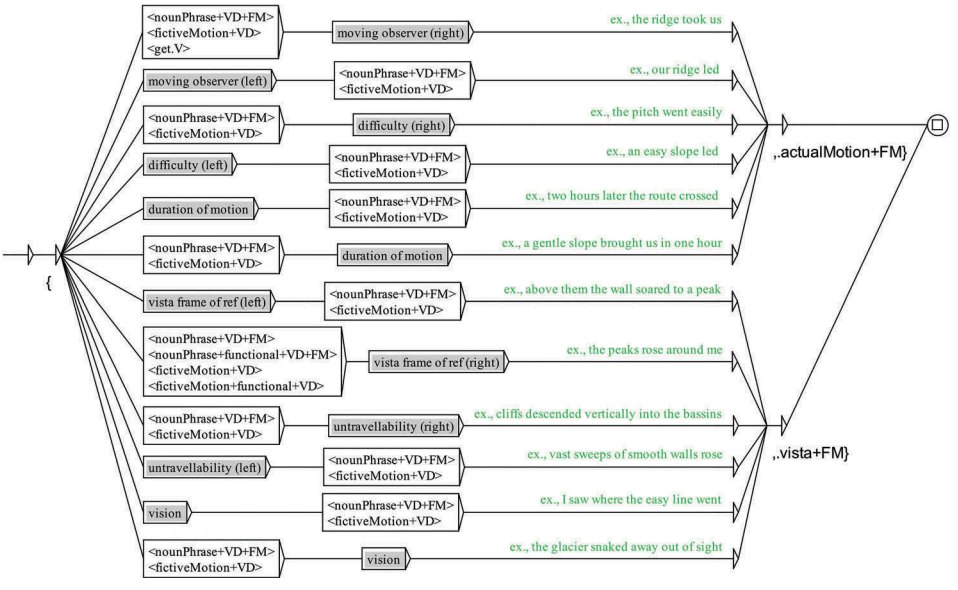

Figure 2. Transducer implementing the main step of classification. Grey boxes represent sub-graphs.

5.2. *Validation and error discussion*

We measure the performance of our approach using three commonly used metrics in identification and classification tasks: *precision* (*P*), *recall* (*R*) and F_1 *score* (F_1).

Precision *P* is defined as the ratio of the number of correctly annotated instances of FM (FM_C) divided by the total number of instances of FM annotated, i.e. both true positives (FM_C) and false positives (FM_{FP}).

$$P = \frac{FM_C}{FM_C + FM_{FP}}$$

Recall *R* is the ratio of the number of correctly annotated instances of FM (FM_C) to the number of FM instances in the collection, i.e. both true positives (FM_C) and false negatives (FM_{FN}).

$$R = \frac{FM_C}{FM_C + FM_{FN}}$$

The F_1 *score* F_1 is the harmonic mean of precision and recall and is defined as

$$F_1 = \frac{2 \times P \times R}{P + R}$$

The last run of the final set of rules on the training data retrieved 520 instances of FM, corresponding to the recall of 0.96 reported in Table 3. Our method was fairly discriminating in identifying FM, with a precision of 0.75 (i.e. for the 520 cases of FM identified, 176 false positives were returned). As for classification, the overall recall is 0.8 with a precision of 0.93, as reported in Table 4. Encyclopaedic knowledge FM is classified

Table 3. FM identification.

	Precision	Recall	F1
Training	0.75	0.96	0.84
Validation	0.62	0.96	0.76
Off: rule for polysemy	0.59	0.96	0.73
Off: rule for factive motion	0.59	0.96	0.73
Off: rule for POS-tagging errors	0.53	0.96	0.68
Off: rules for complex noun phrases	0.58	0.99	0.73

Table 4. FM classification.

		Precision	Recall	F1
Training	Average	0.93	0.80	0.86
	Actual motion	0.85	0.7	0.77
	Vista	0.86	0.48	0.61
	Encycl. knowl.	1	0.93	0.97
Validation	Average	0.87	0.71	0.78
	Actual motion	0.67	0.46	0.55
	Vista	0.75	0.41	0.53
	Encycl. knowl.	1	0.96	0.98
Run 1: nouns' and verbs' semantics only	Average	0.88	0.66	0.76
	Actual motion	0.7	0.43	0.53
	Vista	0.63	0.24	0.35
	Encycl. knowl.	1	0.96	0.98
Run 2: fine-tuned rules only	Average	0.90	0.59	0.71
	Actual motion	0.73	0.27	0.4
	Vista	0.68	0.28	0.4
	Encycl. knowl.	1	0.89	0.94

correctly in 100% of the cases, and vista and actual motion in 87% and 86% of the cases correspondingly.

From the 442 FM cases in the validation data, we automatically retrieved 424 (corresponding to a recall of 0.96 in Table 3). However, precision decreased to 0.62, with 293 false positives returned. Nonetheless, these results demonstrate that our approach was robust, with high recall and acceptable precision when retrieving instances of FM from an 'unseen' validation corpus. Performance of our processing chain for classification is represented in Table 4. Out of 424 retrieved instances of FM, 78 (18%) remain unclassified and 44 (10%) are classified wrongly. The overall recall in classification is 0.71, precision 0.87. The percentage of instances that were not classified and those that were classified wrongly increased by 4% in both cases. As for the types of FM, encyclopaedic knowledge FM, again, gets classified in 100% of the cases (which is not surprising given the robustness of the tense-based rule), and vista and actual motion in 76% and 68% of the cases correspondingly.

To better understand which rules were most useful in identifying and classifying FM, we carried out six sensitivity studies. For the identification of FM, we turned off individual rule sets related to the four types of false positives as described in Section 4.1. For the classification of FM, we decided to compare the performance of two simple rules based on the semantics of spatial entities alone and the set of 'fine-tuned' rules.

5.2.1. *FM extraction*

In what follows, we first report on false positive and false negatives affecting overall results, further reporting on the sensitivity analysis.

5.2.1.1. *False positives.* Inspection of this group reveals instances belonging to the four groups (discussed in Section 4.1) which were not captured through our rules because of the richness of discourse and impossibility of exhaustive operationalization of all patterns. Thus, polysemantic verbs, on the one hand, are not restricted to the four verbs we focused on. On the other hand, exhaustive operationalization of all potential 'non-motion' senses is not feasible and we find new instances of verbs that we covered in the rules. A very simple example is the case of 'take' in (16), which was not captured, since the rule was targeted at 'take part', without foreseeing the possibility of negation. Similar examples can be found for the other three types of false positives.

(16) The ***mountain took*** no part in his death.

Importantly, we also encounter FM among false positives. This can be explained by the fact that the corpus of classified FM structures was produced manually and is thus prone to human error. Through the automated process, we find structures that have been overlooked by a human annotator.

5.2.1.2. *False negatives.* Apart from a few OCR errors,[10] the main reason why some FM structures are discarded (and are thus not found) is the rule aiming at identifying the complex noun phrases representing spatial entities. Since it is based on lists of nouns referring to spatial parts, distance and collection, some rather rare noun phrases are missed (e.g. 'a jumble of peaks', 'landscape of peaks', 'the snout of the glacier').

5.2.1.3. *Sensitivity analysis.* We performed four runs to explore the sensitivity of results to individual rules aimed at reducing the four types of false positives (see Table 3).

Recall remains high for all of these constellations, demonstrating that the individual rules work to filter false positives and thus improve precision rather than increase the number of extracted instances of FM. Recall increase to 0.99 in the last run (where we switch off the rule dealing with complex noun phrases) is explained by the complexity of the rule, resulting in the introduction of errors and loss of true positives in an attempt to deal with false positives as discussed above.

As for precision, rules related to polysemy, factive motion and complex noun phrases result in a rather small change in precision, which are of limited practical importance (from 0.62 to 0.59 and 0.58). However, the identification of errors in part-of-speech tagging is more important, resulting in a reduction of some 9% of false positives retrieved. This shows how systematic alpine discourse (as any discourse) is – those few phrases that we identified in the training data (e.g. 'peaks climbed') are frequent enough to have an impact on precision.

5.2.2. *FM classification*

Here, we first manually reviewed cases that remained unclassified or were classified wrongly. Further, a sensitivity analysis was performed to see the performance of groups of rules, through two additional runs (see Table 4), where some of the rules were switched off.

5.2.2.1. *Unclassified cases.* On the one hand, these represent sentences that contain no markers related to our rules and/or require a broader context for their interpretation. (We referred to this issue in Section 2.3) A number of cases were left unclassified despite the presence of markers. Thus, (17a) was not annotated as actual motion FM despite the explicit vista frame of reference. The reason is that the article 'the' left of the FM structure was not captured in the corresponding rule, since we mostly encountered cases of the type (11a) in the training data.

(17) a. Behind us the ***valley descended*** between serrated mountain walls…

5.2.2.2. *Wrongly classified cases.* These are, firstly, associated with tense. Although all encyclopaedic knowledge FM is identified by the algorithm, the latter also returns false positives: cases where discourse conventions are violated and the present tense is used for a narration about the past. Thus, (18a) was classified as encyclopaedic knowledge based on the tense, despite the vista frame of reference mentioning the observer (which signals vista in our rules). Another type of error relates to cases of vista FM related to hypothetical motion scenario, which we described exhaustively throughout the article (see Sections 2.3, 4.2.2, 4.3). For instance, (18b) was annotated as actual motion on the basis of 'line', representing a functional entity, which appears often in actual motion FM, but can also be encountered in cases of a vista, encoding hypothetical motion.

(18) a. The ***wall drops*** away below us for almost 4000 m.
b. A safe seemingly virgin ***line rose*** out of the wreckage like a vision.

5.2.2.3. *Sensitivity analysis.* The goal of the sensitivity analysis was to compare the role of rules of various complexity. In Run 1, we switched off all 'fine-tuned' rules represented in Figure 2 and thus performed the classification relying only on the semantics of nouns representing spatial entities (actual motion FM in case of a *functional* entity and vista FM in case of a *large-scale* entity) and motion verbs (vista FM if the verb signals untravellability, e.g. 'soar'). Run 2 performed the classification based only on the 'fine-tuned' rules, which are also more numerous (presence of the references to the *moving observer, difficulty, speed, motion duration*, *vista frame of reference* and *vision*).

F1 is highest for validation data with all rules active and decreases slightly for Run 1, more markedly for Run 2. By examining precision and recall in more detail, the underlying reasons for changes in performance become clearer.

Precision increases very slightly from 0.87 in the validation run to 0.88 and 0.9 in Runs 1 and 2 correspondingly. At the same time, recall deteriorates more markedly, from 0.71 in the validation run to 0.66 in Run 1 and further to 0.59 in Run 2. We suggest that this decrease in recall is because the occurrence of markers underlying the fine-tuned rules (Run 2) is less frequent than that of spatial entities annotated as *large-scale* and

functional (Run 1). Furthermore, the increase in precision for Run 1 and Run 2 suggests that combining more rules results in a greater proportion of false positives overall.

To better understand the reasons for these effects, we further examined precision for each individual type of FM classified. For encyclopaedic knowledge, precision is 1 for all runs, suggesting that the rule used is robust. In contrast, for actual motion we obtained precisions of 0.67, 0.70 and 0.73 for validation, Run 1 and Run 2, respectively. This illustrates that adding complexity to our rules appears to decrease precision. However, in practice, the increased precision comes at a cost – recall drops from 0.46 in the validation data to 0.27 in Run 2. Finally, for vista FM, reducing the complexity of our rules results not only in a decreasing recall (from 0.41 in validation to 0.28 in Run 2) but also in decreasing precision (0.75, 0.63, 0.67 for validation, Run 1 and Run 2, respectively). In sum, we argue that we achieve not only best performance for classification overall with all rules turned on, but that in practice for each individual FM case, on balance performance is more satisfying (as captured by F1 scores) with the full combination of rules active.

Generally, fewer cases of FM can be classified without the two sets of rules. Since F1 overall decreases as we suppress rules, we suggest that the slight increase in precision is outweighed in this case by the decline in recall.

6. Concluding discussion and future work

FM, representing a static spatial entity as moving, is pervasive in language (Talmy 2000, Langacker 2005). Given the fact that an accurate interpretation of *motion events* is key for the overall comprehension of spatial information in text, the question of how we should treat FM in text conceptually has been recently raised within the body of work developing spatial language annotation schemes (Pustejovsky and Yocum 2013). Egorova *et al*. (2018) investigated this question through the prism of a very specific corpus of alpine narratives and reported on three types of spatial descriptions encoded by FM – actual motion of the mountaineer, a vista along the way and encyclopaedic knowledge.

These three types of scenes differ profoundly from several aspects: cognitive motivations – actual motion versus visual or mental scanning (Matsumoto 1996a), scale – environmental, vista or geographic (Montello 1993), perspective on the scene – mental tour as in route description for actual motion FM, bird's eye or survey perspective for encyclopaedic spatial knowledge and scene description for vista FM (Taylor and Tversky 1992, Denis *et al*. 1999, Pustejovsky 2017).

In this article, we set out to examine the way these differences are reflected in the way spatial information is treated and represented linguistically. Pragmatically, our goal included the development of a set of rules for the automated extraction and classification of FM into these three types of spatial descriptions using the corpus of annotated FM provided by Egorova *et al*. (2016). Our contribution – the set of concepts and their linguistic operationalization, as well as resulting rules, is thus valuable for several lines of work, including the development of spatial annotation schemes (Pustejovsky 2017), automated reconstruction of spatial information from text (Vasardani *et al*. 2013, Moncla *et al*. 2016), as well as, more generally, research on spatial description and thinking strategies (Landau and Jackendoff 1993, Denis *et al*. 1999, Zlatev 2007).

Summarizing the findings, we first identified a set of concepts uniquely characterizing the types of scenes based on the inspection of data and previous literature. The concepts exhibit various degrees of vagueness. *Motion duration* is an example of a rather crisp concept, while *untravellability*, treated as a non-problematic concept in linguistic literature (Matsumoto 1996, Rojo and Valenzuela 2009), is an inherently vague concept in the context of alpine space and mountaineering. Secondly, we examined how these concepts are encoded linguistically; again, concepts vary from the perspective of linguistic representation. Some are related to Talmy's closed class (Talmy 2000); a good example is *vista frame of reference* ('beneath our feet') that can be operationalized exhaustively based on previous literature on spatial prepositions (Landau and Jackendoff 1993). Others, e.g. the above-mentioned *untravellability,* are related to the open class, are corpus-specific and, given the richness of the lexical subsystem of language, cannot be operationalized exhaustively.

The set of rules and their implementation require shallow data preprocessing (lemmatization and part-of-speech tagging). The good results obtained demonstrate the general robustness and simplicity of some of the rules, as well as regularity of certain discourse-specific patterns. It is especially important to note that the tense-based rule for identifying encyclopaedic knowledge and the high frequency of this type of FM in the corpus alone leads to high recall and compensates for highly ambiguous cases that remain unclassified or are classified wrongly. In particular, distinguishing between vista and actual motion FM in some cases is highly problematic even for human annotators. This is particularly true for vistas that refer to hypothetical motion (e.g. 'A steep snow slope led to a col between the twin peaks'). The fuzzy border between hypothetical vista and actual motion FM is further reflected in the fact that while some of the concepts are uniquely characteristic of only one type of FM, others are *more probable* in one but not excluded in another.

We chose to implement our extraction and classification tools using an iterative, knowledge-based approach. However, since we also annotated both training and validation corpora, it was possible to run a random forest classifier on the classification task using the features we identified. The results for this classifier based on our independent validation data were (results for our rule set are in brackets) a precision of 0.79 (0.87), recall of 0.78 (0.71) and an *F*1 score of 0.78 (0.78). Essentially, the random forest precision was lower than that of our rule set since it balanced precision with recall to give an identical *F*1 score. We suggest that this implies that the rule set chosen was similar (in performance terms) to the optimum tree generated by the random forest classifier. However, it is important to note that the data-driven approach relied firstly on annotated data and secondly on the markers that we identified through a meticulous linguistic analysis and based on the literature. Thus, the key advantage of our knowledge-based method is that expert, in this case linguistic, knowledge can be used iteratively to develop rules which can then be interpreted and transferred to other corpora.

Several research directions can build up on and extend this work. First, given the narrow domain of the 'Text + Berg' corpus and following the principles of corpus-based development of spatial markup languages (Pustejovsky and Moszkowicz 2012), further examination of FM should include the exploration of its pervasiveness in different domains. Related to this is the necessity to run the pipeline on different corpora to check the cross-domain

transferability of rules. While some of the linguistic features we rely on in our approach are corpus-specific (e.g. the particular set of nouns representing spatial entities in mountainous landscapes), others are likely to be transferable to other contexts and types of spatial discourse. Finally, it will be useful to see if syntactic parsing (De Marneffe and Manning 2008), semantic role labelling (Palmer *et al*. 2010) or temporal tagging (Strötgen and Gertz 2010), as well as further fine-tuning of the order and priority of the rules (for concepts that are more probable in one type of FM) can bring an added value.

Given the richness and complexity of spatial language, the importance of exploring the use of structures like FM and the role of corpora of naturally occurring spatial discourse are being increasingly recognized within the GIScience community (Stock *et al*. 2013, Wallgrün *et al*. 2018). On the one hand, it is an important step towards the spatial parsing of text, which will greatly enrich existing toolboxes, allowing capture of spatial information in a variety of textual data. On the other hand, it enhances our understanding of the way spatial information is represented linguistically in various types of spatial discourses, providing a window into the spatial thinking involved. Overall, our approach provides both conceptual and practical pathways to the development of semantic and database representations of geographical scenes described in natural language at different scales. As stated by Goodchild (2009), this could bridge the gap between the informal, loose world of human cognition and discourse and the rigorous, formal and precise representation of space in computerized Geographic Information Systems.

Notes

1. We present and refer to examples throughout the text as numbered examples.
2. http://unitexgramlab.org/.
3. http://www.tei-c.org/index.xml.
4. Further examples of similar initiatives include the Automatic Content Extraction (ACE) (Doddington *et al*. 2004) and the Quaero (Grouin *et al*. 2011) programmes.
5. *Figure* is a moving entity in a *motion event* (Talmy 2000), we will use this term throughout the article.
6. Examples (3a)–(4d) are borrowed from Matsumoto (1996b).
7. Or, if an evolution exists, it is realized on a much larger time scale.
8. By 'FM structure' we mean an atomic sequence of a noun or a noun phrase representing a spatial entity and a verb, and indicate it by bold italics, e.g. ***ridge ran***.
9. This does not mean that present tense is incompatible with a vista or actual motion FM. Rather, these are discourse conventions found in our corpus and elsewhere, where present tense is usually used for describing permanent state of things, while past tense is used for the narration about the events in the past.
10. These are errors resulting from the digitization process. Thus, "the iine went" was recognized by the human annotator as "the line went" and annotated as FM. However, it was not captured through the automated process, since "iine" is not included into the lexicon.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Ludovic Moncla http://orcid.org/0000-0002-1590-9546

References

Abney, S., 1996. Partial parsing via finite-state cascades. *Natural Language Engineering*, 2 (4), 337–344. doi:10.1017/S1351324997001599

Bateman, J.A., *et al*. 2010. A linguistic ontology of space for natural language processing. *Artificial Intelligence*, 174 (14), 1027–1071. doi:10.1016/j.artint.2010.05.008

Bruggmann, A. and Fabrikant, S.I., 2016. How does GIScience support spatio-temporal information search in the humanities? *Spatial Cognition & Computation*, 16 (4), 255–271. doi:10.1080/13875868.2016.1157881

Bubenhofer, N., *et al*. 2015. *Text+Berg Korpus*. Zurich: Institut fuer Computerlinguistik, Universitaet Zuerich.

Bunt, H. and Pustejovsky, J., 2010. Annotating temporal and event quantification. *In*: *Proceedings of the Fifth Joint ISO-ACL/SIGSEM Workshop on Interoperable Semantic Annotation (ISA-5)*. Hong Kong: City University, 15–22.

Burnard, L., 2007. New tricks from an old dog: an overview of TEI P5. *In*: *Dagstuhl Seminar Proceedings*. Schloss Dagstuhl, Germany: Internationales Begegnungs- und Forschungszentrum fuer Informatik (IBFI).

Chinchor, N. and Robinson, P., 1997. MUC-7 named entity task definition. *In*: *Proceedings of the 7th Conference on Message Understanding*, 29 April–1 May, Fairfax, Virginia. vol. 29.

Criminisi, Antonio, Jamie Shotton, and Ender Konukoglu. "Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi–supervised learning." *Foundations and Trends® in Computer Graphics and Vision 7.2–3* 2012: 81–227.

Davies, C., 2013. Reading geography between the lines: extracting local place knowledge from text. *In*: T. Tenbrink, *et al*., eds. *International Conference on Spatial Information Theory (COSIT 2013)*, 2–6 September, Scarborough, UK. Cham: Springer, 320–337.

De Marneffe, M.C. and Manning, C.D., 2008. *Stanford typed dependencies manual*. Technical report, Stanford: Stanford University.

Denis, M., *et al*. 1999. Spatial discourse and navigation: an analysis of route directions in the city of Venice. *Applied Cognitive Psychology*, 13 (2), 145–174. doi:10.1002/(SICI)1099-0720(199904)13:2<145::AID-ACP550>3.0.CO;2-4

Derungs, C. and Purves, R.S., 2016. Characterising landscape variation through spatial folksonomies. *Applied Geography*, 75, 60–70. doi:10.1016/j.apgeog.2016.08.005

Doddington, G.R., *et al*., 2004. The automatic content extraction (ACE) program-tasks, data, and evaluation. *In*: *Fourth International Conference on Language Resources and Evaluation (LREC 2004)* 26–28 May, Lisbon, Portugal. European Language Resources Association, 837–840.

Egorova, E., Boo, G., and Purves, R.S., 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*, 27–30 September, Montreal, Canada. 92–95.

Egorova, E., Tenbrink, T., and Purves, R.S., 2018. Fictive motion in the context of mountaineering. *Spatial Cognition & Computation*, 1–26. doi:10.1080/13875868.2018.1431646

Friburger, N. and Maurel, D., 2004. Finite-state transducer cascades to extract named entities in texts. *Theoretical Computer Science*, 313 (1), 93–104. doi:10.1016/j.tcs.2003.10.007

Goodchild, M.F., 2009. Geographic information systems and science: today and tomorrow. *Annals of GIS*, 15 (1), 3–9. doi:10.1080/19475680903250715

Grishman, R. and Sundheim, B., 1996. Message understanding conference-6: a brief history. *In*: *COLING 1996 Volume 1: The 16th International Conference on Computational Linguistics*, 5–9 August, Copenhagen, Denmark. vol. 1, 466–471.

Gritta, M., et al., 2018. What's missing ingeographical parsing?. *Language Resources and Evaluation*, 52 (2),603–623.

Grouin, C., *et al*., 2011. Proposal for an extension of traditional named entities: from guidelines to evaluation, an overview. *In*: *Proceedings of the 5th Linguistic Annotation Workshop*, 23–24 June, Portland, Oregon. Association for Computational Linguistics, 92–100.

Herskovits, A., 1987. *Language and Spatial Cognition*. Cambridge: Cambridge University Press.

Khan, A., Vasardani, M., and Winter, S., 2013. Extracting spatial information from place descriptions. *In*: S. Scheider, *et al*., eds, *Proceedings of the First ACM SIGSPATIAL International Workshop on Computational Models of Place*. 5–8 November Orlando, FL. New York, NY: ACM, 62–69.

Kim, J., Vasardani, M., and Winter, S., 2017. Similarity matching for integrating spatial information extracted from place descriptions. *International Journal of Geographical Information Science*, 31 (1), 56–80. doi:10.1080/13658816.2016.1188930

Klippel, A., 2003. Wayfinding choremes. *In*: W. Kuhn, M. Worboys, and S. Timpf, eds. *Spatial Information Theory. Foundations of Geographic Information Science (COSIT 2003)*, 24–28 September, Kartause Ittingen, Switzerland. Berlin, Heidelberg: Springer, 301–315.

Kordjamshidi, P., Van Otterlo, M., and Moens, M.F., 2011. Spatial role labeling: towards extraction of spatial relations from natural language. *ACM Transactions on Speech and Language Processing (TSLP)*, 8 (3), 4.

Kosar, T., *et al*. 2010. Comparing general-purpose and domain-specific languages: an empirical study. *Computer Science and Information Systems*, 7 (2), 247–264. doi:10.2298/CSIS1002247K

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

Landau, B. and Jackendoff, R., 1993. Whence and whither in spatial language and spatial cognition? *Behavioral and Brain Sciences*, 16 (2), 255–265. doi:10.1017/S0140525X00029927

Langacker, R.W., 1987. *Foundations of Cognitive Grammar: Theoretical Prerequisites*. Vol. 1, Stanford: Stanford university press.

Langacker, R.W., 2005. Dynamicity, fictivity, and scanning: the imaginative basis of logic and linguistic meaning. *In*: D. Pecher and R.A. Zwaan, eds. *Grounding Cognition: The Role of Perception and Action in Memory, Language, and Thinking*. Cambridge: Cambridge University Press, 164–198.

Leveling, J. and Hartrumpf, S., 2008. On metonymy recognition for geographic information retrieval. *International Journal of Geographical Information Science*, 22 (3), 289–299. doi:10.1080/13658810701626244

Levinson, S.C., 1999. Frames of reference and Molyneux's question: crosslinguistic evidence. *In*: P. Bloom, *et al*., eds. *Language and space*. Cambridge, Massachusetts: MIT press, 109–169.

Levinson, S.C., 2003. *Space in Language and Cognition: Explorations in Cognitive Diversity*. Vol. 5, Cambridge: Cambridge University Press.

Mani, I., *et al*. 2010. SpatialML: annotation scheme, resources, and evaluation. *Language Resources and Evaluation*, 44 (3), 263–280. doi:10.1007/s10579-010-9121-0

Marcus, M.P., Marcinkiewicz, M.A., and Santorini, B., 1993. Building a large annotated corpus of English: the Penn Treebank. *Computational linguistics*, 19 (2), 313–330.

Matsumoto, Y., 1996a. Subjective motion and English and Japanese verbs. *Cognitive Linguistics*, 7 (2), 183–226. doi:10.1515/cogl.1996.7.2.183

Matsumoto, Y., 1996b. How abstract is subjective motion? A comparison of coverage path expressions and access path expressions. *In*: A. Golderg, ed. *Conceptual structure, discourse and language*. Stanford: CSLI Publications, 359–373.

Mikheev, A., Moens, M., and Grover, C., 1999. Named entity recognition without gazetteers. *In*: *Proceedings of the Ninth Conference on European Chapter of the Association for Computational Linguistics*, 8–12 June, Bergen, Norway. Association for Computational Linguistics, 1–8.

Moncla, L., *et al*. 2016. Reconstruction of itineraries from annotated text with an informed spanning tree algorithm. *International Journal of Geographical Information Science*, 30 (6), 1137–1160. doi:10.1080/13658816.2015.1108422

Montello, D.R., 1993. Scale and multiple psychologies of space. *In*: A. Frank and I. Campari, eds, *Spatial Information Theory: A Theoretical Basis for GIS (COSIT 1993)*. 19–22 September, Elba Island, Italy. Berlin: Springer, 312–321.

Palmer, M., Gildea, D., and Xue, N., 2010. Semantic role labeling. *Synthesis Lectures on Human Language Technologies*, 3 (1), 1–103. doi:10.2200/S00239ED1V01Y200912HLT006

Piotrowski, M., 2012. Natural language processing for historical texts. *Synthesis Lectures on Human Language Technologies*, 5 (2), 1–157. doi:10.2200/S00436ED1V01Y201207HLT017

Pustejovsky, J. and Moszkowicz, J., 2012. The role of model testing in standards development: the case of ISO-space. *In*: N. Calzolari, *et al*., eds. *Eighth International Conference on Language Resources and Evaluation (LREC 2012)*, 21–27 May. European Language Resources Association, 3060–3063.

Pustejovsky, J., 2017. Iso-space: annotating static and dynamic spatial information. *In*: N. Ide and J. Pustejovsky, eds. *Handbook of Linguistic Annotation*. Dordrecht: Springer, 989–1024.

Pustejovsky, J., Moszkowicz, J., and Verhagen, M., 2012. A linguistically grounded annotation language for spatial information. *TAL*, 53 (2), 87–113.

Pustejovsky, J. and Yocum, Z., 2013. Capturing motion in ISO-spacebank. *In*: *Workshop on Interoperable Semantic Annotation*. 19–20 March. Potsdam, Germany: Association for Computational Linguistics, 25.

Ravin, Y. and Leacock, C., 2000. *Polysemy: Theoretical and Computational approaches*. Oxford: OUP Oxford.

Rojo, A. and Valenzuela, J., 2009. Fictive motion in Spanish: travellable, non-travellable and path-related manner information. *In*: J. Valenzuela, A. Rojo, and C. Soriano, eds. *Trends in Cognitive Linguistics: Theoretical and Applied Models*. Hamburg, Germany: Peter Lang, 244–260.

Stock, K., *et al*., 2013. Creating a corpus of geospatial natural language. *In*: T. Tenbrink, *et al*., eds. *International Conference on Spatial Information Theory (COSIT 2013)*, 2–6 September, Scarborough, UK. Cham: Springer, 279–298.

Strötgen, J. and Gertz, M., 2010. Heideltime: high quality rule-based extraction and normalization of temporal expressions. *In*: *Proceedings of the 5th International Workshop on Semantic Evaluation (SemEval 2010)*, 15–16 July, Uppsala, Sweden. Association for Computational Linguistics, 321–324.

Talmy, L.,2000. Toward a cognitive semantics. 1st, *Language, Speech, and Communication*. Vol. 1, Cambridge, Massachusetts: MIT Press.

Taylor, H.A. and Tversky, B., 1992. Spatial mental models derived from survey and route descriptions. *Journal of Memory and Language*, 31 (2), 261–292. doi:10.1016/0749-596X(92)90014-O

Tenbrink, T., Wiener, J.M., and Claramunt, C., eds., 2013. *Representing Space in Cognition: Interrelations of Behaviour, Language, and Formal Models*. Oxford: Oxford University Press.

Van Deursen, A. and Klint, P., 2002. Domain-specific language design requires feature descriptions. *Journal of Computing and Information Technology*, 10 (1), 1–17. doi:10.2498/cit.2002.01.01

Van Deursen, A., Klint, P., and Visser, J., 2000. Domain-specific languages: an annotated bibliography. *ACM Sigplan Notices*, 35 (6), 26–36. doi:10.1145/352029

Vasardani, M., *et al*., 2013. From descriptions to depictions: a conceptual framework. *In*: T. Tenbrink, *et al*., eds. *International Conference on Spatial Information Theory (COSIT 2013*), 2–6 September, Scarborough, UK. Cham: Springer, 299–319.

Wallgrün, J.O., *et al*., 2018. Geocorpora: building a corpus to test and train microblog geoparsers. *International Journal of Geographical Information Science*, 32 (1), 1–29.

Wang, W. and Stewart, K., 2015. Spatiotemporal and semantic information extraction from Web news reports about natural hazards. *Computers, Environment and Urban Systems*, 50, 30–40. doi:10.1016/j.compenvurbsys.2014.11.001

Xu, S., *et al*. 2014. Exploring regional variation in spatial language using spatially stratified web-sampled route direction documents. *Spatial Cognition & Computation*, 14 (4), 255–283. doi:10.1080/13875868.2014.943904

Zlatev, J., 2007. Spatial semantics. *In*: D. Geeraerts and H. Cuyckens, eds. *The Oxford Handbook of Cognitive Linguistics*. Oxford: Oxford University Press, 318–350.