From Language towards Formal Spatial Calculi

Parisa Kordjamshidi, Martijn Van Otterlo, and Marie-Francine Moens

Katholieke Universiteit Leuven, Departement Computerwetenschappen
{parisa.kordjamshidi,martijn.vanotterlo,sien.moens}@cs.kuleuven.be

Abstract. We consider mapping unrestricted natural language to formal spatial representations. We describe ongoing work on a two-level machine learning approach. The first level is linguistic, and deals with the extraction of spatial information from natural language sentences, and is called *spatial role labeling*. The second level is ontological in nature, and deals with mapping this linguistic, spatial information to formal spatial calculi. Our main obstacles are the lack of available annotated data for training machine learning algorithms for these tasks, and the difficulty of selecting an appropriate abstraction level for the spatial information. For the linguistic part, we approach the problem in a gradual way. We make use of existing resources such as The Preposition Project (TPP) and the validation data of General Upper Model (GUM) ontology, and we show some computational results. For the ontological part, we describe machine learning challenges and discuss our proposed approach.

1 Introduction

An essential function of language is to convey *spatial relationships* between objects and their relative locations in a space. It is a challenging problem in robotics, navigation, query answering systems, etc. [19]. Our research considers the extraction of spatial information in a multimodal environment. We want to represent spatial information using formal representations that allow spatial reasoning. An example of an interesting multimodal environment is the domain of navigation where we expect a robot to follow navigation instructions. By placing a camera on the robot, it should be able to recognize visible objects and their location. In this context, mapping natural language to a formal spatial representation [4] has several advantages. First, generating language from vision and vice versa visualizing the language is more feasible if a formal intermediate layer is employed [16]. Second, applying the same representation for extraction from image/video data allows combining multimodal features for better recognition and disambiguation in each modality. Finally, a unified representation for various modalities enables spatial reasoning based on multimodal information.
In our work we identify two main layers of information (see also [2]):
1) a **linguistic** layer, in which (unrestricted) natural language is mapped onto ontological structures that convey spatial information, and 2) a **formal** layer, in which the ontological information is mapped onto a specific spatial calculus such as *region connection calculus* (RCC) (cf. [4]). For example, in the sentence the

book is on the table the first step should identify that there is a spatial relation (on) between book and table, after which this could be mapped to a specific, formal relation AboveExternallyConnected(book, table) between two tokens book and table that denote two physical objects in some Euclidean space. For both transformations we propose *machine learning* techniques to deal with the many sources of ambiguity in this task. This has not been done systematically before; most often a restricted language is used to extract highly specific and application-dependent relations and usually one focuses on phrases of which it is known that spatial information is present [8, 6, 19, 11].

To apply machine learning effectively, a clear task definition as well as annotated data are needed. Semantic hand-labeling of natural language is an ambiguous, complex and expensive task and in our two-level view we have to cope with the lack of available data two times. In our recently proposed semantic labeling scheme [10], we tag sentences with the spatial roles according to *holistic spatial semantic* (HSS) theory [21] and also formal spatial relation(s). For mapping between language and spatial information, we defined *spatial role labeling* and performed experiments on the (small amount of) available annotated corpora. The Preposition Project (TPP) data is employed for spatial preposition recognition in the context of learning the main spatial roles *trajector* and *landmark* from data. We have conducted initial experiments on the small corpus of the GUM [1] spatial ontology, and the results indicate that machine learning based on linguistic features can indeed be employed for this task.

The second layer of our methodology consists of mapping the extracted spatial information onto formal spatial systems capable of spatial reasoning. Here we propose to annotate data with spatial calculi relations and use machine learning to obtain a *probabilistic logical* model [3] of spatial relations for this mapping. Such models can deal with both the structural aspects of spatial relations, as well as the intrinsic ambiguity and vagueness in such mappings (see also [5]). In the following sections we will describe both the linguistic and the formal steps, and results of our initial machine learning experiments.

2 Linguistic Level and Spatial Role Labeling

To be able to map natural language to spatial calculi we should first extract the components of spatial information. We call this task **spatial role labeling**. It has not been well-defined before and has not been considered as a stand-alone linguistic task. We define it analogous to *semantic role labeling* (SRL) [15], targeting semantic information associated with specific phrases (usually verbs), but as a stand-alone linguistic task utilizing specific (data) resources.

Task definition. We define spatial role labeling (SpRL) as *the automatic labeling of natural language with a set of spatial roles.* The sentence-level spatial analysis of text deals with characterizing spatial descriptions, denoting the spatial properties of objects and their location (e.g. to answer "what/who/where"-questions). A spatial term (typically a preposition) establishes the type of spatial relation and other constituents express the participants of the spatial relation (e.g. a location). The *roles* are drawn from a pre-specified list of possible spatial roles and the role-bearing constituents in a spatial expression must be identified and their correct spatial role labels assigned.

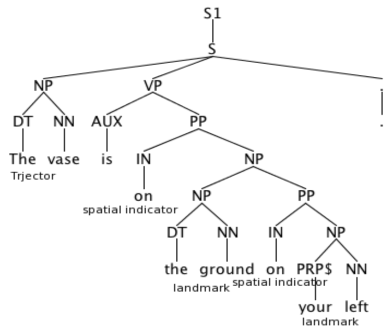

Fig. 1. Parse tree with spatial roles

Representation based on spatial semantics. The spatial role set we employ contains the core roles of *trajector, landmark, spatial indicator*, and *motion indicator* [6, 21], as well as the features *path* and *frame of reference*. Our set of spatial roles are motivated by the theory of holistic spatial semantics upon which we have defined an annotation scheme in [10]. We describe these terms briefly. A **trajector** is the entity whose (trans)location is of relevance. It can be static or dynamic; a person or an object. It can also be expressed as a whole event. Other terms often used for this concept are the *local object*, *locatum*, *figure object*, *referent* and *target*. A **landmark** is the reference entity in relation to which the location or the trajectory of motion of the trajector is specified. Alternative terms include *reference object*, *ground* and *relatum*. A **spatial indicator** is a token which defines constraints on the spatial properties, such as the location of the trajector with respect to the landmark (e.g. in, on). It explains the type of the spatial relation and usually is a preposition, but can also be a verb, a noun, etc. It is the pivot of a spatial relation, and in terms of GUM ontology it is called a spatial modality. A **motion indicator** is a spatial term which is an indicator of motion, e.g. motion *verbs*. We also consider other conceptual aspects like **frame of reference** and the **path** of a motion that are important for spatial semantics and roles [21].

Linguistic challenges. Given a sentence, SpRL should answer: **Q1.** Does the sentence contain spatial information? **Q2.** What is the **pivot** of the spatial information? (spatial indicator) **Q3.** Starting from the pivot how can we identify/classify the related arguments with respect to predefined set of spatial roles? Spatial relations in English are mostly expressed using prepositions [7], but verbs and even other lexical categories can be central spatial terms. Hence SpRL consists of identifying the boundaries of the arguments of the identified spatial term and then labeling them with spatial roles (argument classification). However, there are very sparse and limited resources for learning spatial roles. Other work typically uses a limited set of words, often based on a set of spatial prepositions and specific grammatical patterns in a specific domain [13, 8].

General extraction of spatial relations is hindered by several things. First, there is not always a regular mapping between a sentence's parse tree and its spatial semantic structure. This is more challenging in complex expressions which convey several spatial relations [4]; see the following sentence (and Fig. 1).

The vase is **on** the ground **on** your left.

Here a dependency parser relates the first "on" to "vase" and "ground". This will produce a valid spatial relation. But the second "on" is related to "ground" and "left", producing a meaningless spatial relation (ground on your left). For more complex relations and nested noun phrases, deriving spatially valid relations is not straightforward and depends on the *lexical* meaning of the words. Other linguistic phenomena such as *spatial-focus-shift* and *ellipsis of trajector and landmark* [11] make the extraction more difficult. Recognizing the right PP-attachment (i.e. whether the preposition is attached to the verb phrase or noun phrase) could help the identification of spatial arguments when the verb in the sentence conveys spatial meaning. *Spatial motion detection* and *recognition of the frame of reference* are additional challenges but will not be dealt with here.
Approach. We aim to tackle the problem using machine learning, in a way similar to SRL, but with important differences. The first difference is that the main focus of SRL is on the *predicate*, its related arguments and their roles [15]. On the other hand, in SpRL the spatial indicator plays the main role and should be identified and disambiguated beforehand. Second, the set of main roles is quite different in SpRL and a large enough English corpus is not available from which spatial roles can be learned directly. Hence new data resources are needed. The main point is that we aim at domain-independent and unrestricted language analysis. This prohibits using very limited data or a small set of extraction rules. However, utilizing existing linguistic resources which can partially or indirectly help to set up a (relational) joint learning framework will be of great advantage. It can relinquish the necessity of expensive labeling of one huge corpus. Our results for preliminary experiments are briefly described in Section 4.

3 Towards Spatial Calculi and Spatial Formalizing

Mapping the spatial information in a sentence onto spatial calculi is the second step in our framework. We denote this as *spatial formalizing task*.
Task definition. We define spatial formalizing as *the automatic mapping of the output of SpRL to formal relations in spatial calculi.* In the previous section we have assumed that our spatial role representation covers all the spatial semantic aspects according to HSS. For the target representation of spatial formalizing we also require that it can express various kinds of spatial relations.
Spatial challenges. Ambiguity and under-specification of spatial information conveyed in the language, but also overspecification of spatial calculi models, make a direct mapping between the two sides difficult [2]. Most of the qualitative spatial models focus on a single aspect, e.g. topology, direction, or shape [12]. This is a drawback, particularly from a linguistic point of view and with respect to the pervasiveness of the language. Hence spatial formalizing should cover multiple aspects with practically acceptable level of generality. In the work of [5] the alignment between the linguistic and logical formalizations is discussed. Since these two aspects are rather different and provide descriptions of the environment from different viewpoints, constructing an intermediate, linguistically motivated ontology is proposed to establish a flexible connection between them.

GUM (*Generalized Upper Model*) is the state-of-the-art example of such an ontology [1, 17]. Moreover, in [5] $S-$connections are suggested as a similarity-based model to make a connection between various formal spatial systems and mapping GUM to various spatial calculi. However, obtaining an annotated corpus is the main challenge of machine learning for mapping to the target relations/ontology. In this respect using an intermediate level with a fairly large and fine-grained division of concepts is to some extent difficult and implies the need to have a huge labeled corpus. In addition, the semantic overlap between the included relations in the large ontologies makes the learning model more complex.

Moreover, mapping to spatial calculi is an inevitable step for spatial reasoning. Hence even if a corpus is constructed by annotating with a linguistically motivated ontology, mapping to spatial calculi still should be handled as a separate and difficult step. Even at this level, it is not feasible to define a deterministic mapping by formulating rules because bridging models to each other is not straightforward and external factors, context and all the involved spatial components, discourse features, etc influence this final mapping. Therefore the relationships between instances in different domains are not deterministic and they are often ambiguous and uncertain [5]. Given that for each learning step, a corpus should be available, we argue that it seems most efficient to learn a mapping from SpRL to (one or several) spatial calculi directly.

Representation based on spatial calculi. To deal with these challenges we proposed an annotation framework [10] inspired by the works of SpatialML [14] and a related scheme in [18]. We suggest to map the extracted spatial indicators and the related arguments onto the general type of the related spatial relation Region, Direction, Distance because these relations cover all coarse-grained aspects of space (except shape). The specific relation expressed by the indicators is stated in the suggested scheme with an attribute named *specific-type*. If the general-type is REGION then we map this onto topological relations in a qualitative spatial reasoning formalism, so the specific-type will be RCC8 which is a popular formal model. For directions the specific type gets a value in {ABSOLUTE, RELATIVE}. For absolute directions we use {S(south), W(west), N(north), E(east), NE(northeast), SE(southeast), NW(northwest), SW(southwest)} and for relative directions {LEFT, RIGHT, FRONT, BEHIND, ABOVE, BELOW} which can be used in qualitative direction calculi. Distances are tagged with {QUALITATIVE, QUANTITATIVE} (cf. [10]). To provide sufficient flexibility in expressing all possible spatial relations our idea is to allow more than one formal relation to be connected to one linguistic relation, helped by a (probabilistic) logical representation. The following examples illustrate this.

a)...and next to that left of that is my computer, perhaps a meter away.

Let $\mathtt{X}$=my computer, $\mathtt{Y}$=that, then a SpRL gives $\mathtt{nextTo(X,Y)}$, $\mathtt{leftOf(X,Y)}$, and a resulting spatial formalization is $\mathtt{DC(X,Y)}$, $\mathtt{LEFT(X,Y)}$, $\mathtt{Distance(X,Y,'value')}$ which in GUM corresponds to `leftprojectionexternal`.

b) The car is between two houses.

SpRL: $\mathtt{between(car,houses)}$, spatial relations: $\mathtt{left(car,houses)}$ AND $\mathtt{right(car,houses)}$ which corresponds to GUM's `Distribution`.

c) The wheatfield is in line with crane bay.
SpRL: `inline(wheatfield, cranebay)`, spatial relations: `behind(wheatfield, cranebay)` XOR `front(wheatfield, cranebay)` GUM: `RelativeNonProjectionAxial`

Approach. The above mentioned examples show that a logical combination of basic relations can provide the required level of expressivity in the language. These annotations will enable learning probabilistic logical models relating linguistic spatial information to relations in multiple spatial calculi. Afterwards qualitative (or even probabilistic) spatial reasoning will be feasible over the produced output. The learned relations could be considered as probabilistic constraints about most probable locations of the entities in the text. Probabilistic logical learning [3] provides a tool in which considerable amounts of (structured) background knowledge can be used in the presence of uncertainty. The available linguistic background knowledge and features includes i) the features of the first step of spatial role labeling (syntactic, lexical and semantical information from the text) and ii) linguistic resources such as WordNet, FrameNet, language models and word co-occurences [20]. These could be combined with visual features extracted from visual resources in a multimodal environment for more specification of spatial relations. Structured outputs (i.e. the mapping to formal relations) could be learned in a joint manner. By exploiting a joint learning platform, annotating a corpus by aforementioned spatial semantics in addition to annotating by the final spatial relations (derived from spatial calculi) is less expensive than annotating and learning the two levels independently. Implementing such a learning setting is ongoing work.

4 Current Experiments

To start with empirical studies, we have performed experiments on the first SpRL learning phase. We learn to identify spatial indicators and their arguments trajector and landmark. We do not treat *motion*, *path* and *frame of reference* in this paper, and focus solely on prepositions as spatial indicators here.
Spatial preposition. For unrestricted language it seems valuable to first *recognize whether* there is any spatial indicator in the text. Since prepositions mostly play key roles for the spatial information in the first step we examine whether an existing preposition in a sentence conveys a spatial sense. Here we use linguistically motivated features, such as parse and dependency trees and semantic roles. We extracted these features from the training and test data of the TPP data set and tested several classifiers. The current results are a promising starting point for the spatial sense recognition and the extraction of spatial relations. The selected features were evaluated experimentally and our final coarse-grained MaxEntropy sense classifier outperformed the best system of the SemEval-2007 challenge by providing an F1 measure of about 0.874. We achieved an accuracy of about 0.88 for the task of recognizing whether a preposition has a spatial meaning in a given context.
Extraction of trajector and landmark. In the second SpRL step we extract the trajector and landmark arguments. Our features are inspired by those

in SRL. The main difference is that the pivot of the semantic relations here is the preposition, and not the predicate. The features from the parse/dependency tree and semantic role labeler are extracted from GUM examples. We labeled the nodes in the parse tree with GUM labels `trajector`(`locatum`), `landmark`(`relatum`) and spatial indicator (`spatialModality`).

We assume the spatial indicator (preposition) is correctly disambiguated and given, i.e. we perform a multi-class classification of parse tree nodes by `trajector`, `landmark` and `none`, for which we employed standard classifiers (naive Bayesian (NB), and maximum entropy (MaxEnt)). In addition, we tagged the sentences as sequences using the same features and applied a simple sequence tagger based on conditional random fields (CRF). The spatial annotations of GUM were altered in some instances to be able to obtain more regular patterns for machine learning. We labeled the continuous words (prepositions) and their modifiers as one spatial modality even if they had been tagged as individual relations in GUM, and we do not tag implicit trajectors/landmarks. In ongoing experiments we classify the headwords instead of whole constituents [9]. Table 1 presents the preliminary results for "trajector" (T) and "landmark" (L) recognition including overall accuracy evaluated by 10-fold cross validation. The simple multi-class classification ignores the global correlations between classes and as Table 1 indicates, more sophisticated CRF models can improve the results in particular for landmarks. Since the main sources of errors are a lack of data and the dependency of spatial semantics on lexical information, we will employ additional (lexical) features and ideally will use a larger corpus in our future experiments. However the current results show the first step of applying machine learning for SpRL and indicate a promising start towards achieving the entire automatic mapping from language to spatial calculi.

Method	F1(T)	F1(L)	Acc(All)
NBayes	0.86	0.70	0.94
MaxEnt	0.91	0.767	0.965
CRF	0.928	0.901	0.921

Table 1. Extraction of trajector (T) and landmark (L)

5 Conclusion and Future Directions

We have introduced a model for mapping natural language to spatial calculi. Both aspects of *spatial role labeling* and *spatial formalizing* have been described. A number of related problems that cause difficulties and ambiguities were addressed, and we have shown preliminary results for experiments on SpRL and the extraction of trajectors and landmarks. Our main idea for future work is to obtain (i.e. create) a corpus which is labeled by holistic spatial semantics plus a combination of spatial calculi. Each relation in the language can be connected to a *set* of relations belonging to predefined spatial calculi. This gives a logical representation of the language based on spatial calculi. We aim to learn statistical relational models for this. This enables adding probabilistic background knowledge related to structural information and spatial semantic notions, and supports (probabilistic) spatial reasoning over the learned models.

References

1. J. Bateman, T. Tenbrink, and S. Farrar. The role of conceptual and linguistic ontologies in discourse. *Discourse Processes*, 44(3):175–213, 2007.
2. J. A. Bateman. Language and space: a two-level semantic approach based on principles of ontological engineering. *Int. J. of Speech Tech.*, 13(1):29–48, 2010.
3. L. De Raedt, P. Frasconi, K. Kersting, and S. Muggleton, editors. *Probabilistic Inductive Logic Programming*, volume 4911 of *LNCS*. Springer, 2008.
4. Antony Galton. Spatial and temporal knowledge representation. *Journal of Earth Science Informatics*, 2(3):169–187, 2009.
5. J. Hois and O. Kutz. Counterparts in language and space – similarity and s-connection. In *Proceedings of the 2008 Conference on Formal Ontology in Information Systems*, 2008.
6. J. D. Kelleher. *A Perceptually Based Computational Framework for the Interpretation of Spatial Language.* PhD thesis, Dublin City University, 2003.
7. J. D. Kelleher and F. J. Costello. Applying computational models of spatial prepositions to visually situated dialog. *Comput. Linguist.*, 35(2):271–306, 2009.
8. T. Kollar, S. Tellex, D. Roy, and N. Roy. Toward understanding natural language directions. In *HRI*, 2010.
9. P. Kordjamshidi, M van Otterlo, and M. F. Moens. Spatial role labeling: Automatic extraction of spatial relations from natural language. Technical report, Katholieke Universiteit Leuven, 2010.
10. P. Kordjamshidi, M van Otterlo, and M. F. Moens. Spatial role labeling: task definition and annotation scheme. In *LREC*, 2010.
11. H. Li, T. Zhao, S. Li, and J. Zhao. The extraction of trajectories from real texts based on linear classification. In *Proceedings of NODALIDA*, 2007.
12. W. Liu, S. Li, and J. Renz. Combining RCC-8 with qualitative direction calculi: algorithms and complexity. In *IJCAI*, 2009.
13. K. Lockwood, K. Forbus, D. T. Halstead, and J. Usher. Automatic categorization of spatial prepositions. In *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, 2006.
14. I. Mani, J. Hitzeman, J. Richer, D. Harris, R. Quimby, and B. Wellner. SpatialML: Annotation scheme, corpora, and tools. In *LREC*, 2008.
15. L. Màrquez, X. Carreras, K. C. Litkowski, and S. Stevenson. Semantic role labeling: An introduction to the special issue. *Comp. Ling.*, 34(2):145–159, 2008.
16. R. J. Mooney. Learning to connect language and perception. In *AAAI*, 2008.
17. R. Ross, H. Shi, T. Vierhuff, B. Krieg-Brückner, and J. Bateman. Towards dialogue based shared control of navigating robots. In *Proceedings of Spatial Cognition IV: Reasoning, Action, Interaction*, pages 478–499, 2005.
18. Q. Shen, X. Zhang, and W. Jiang. Annotation of spatial relations in natural language. In *Proceedings of the International Conference on Environmental Science and Information Application Technology*, 2009.
19. D. A. Tappan. *Knowledge-Based Spatial Reasoning for Automated Scene Generation from Text Descriptions.* PhD thesis, New Mexico State University Las Cruces, New Mexico, 2004.
20. M. Tenorth, D. Nyga, and M. Beetz. Understanding and Executing Instructions for Everyday Manipulation Tasks from the World Wide Web. In *ICRA*, 2010.
21. J. Zlatev. Spatial semantics. *In Hubert Cuyckens and Dirk Geeraerts (eds.) The Oxford Handbook of Cognitive Linguistics, Chapter 13*, pages 318–350, 2007.