

Spatial Role Labeling: Task Definition and Annotation Scheme

Parisa Kordjamshidi, Martijn Van Otterlo, Marie-Francine Moens

Katholieke Universiteit Leuven

Departement Computerwetenschappen Celestijnenlaan 200 A B-3001 Heverlee Belgium

parisa.kordjamshidi,martijn.vanotterlo,sien.moens@cs.kuleuven.be

Abstract

One of the essential functions of natural language is to talk about *spatial relationships* between objects. Linguistic constructs can express highly complex, relational structures of objects, spatial relations between them, and patterns of motion through spaces relative to some reference point. *Learning* how to map this information onto a formal representation from a text is a challenging problem. At present no well-defined framework for automatic spatial information extraction exists that can handle all of these issues. In this paper we introduce the task of *spatial role labeling* and propose an annotation scheme that is language-independent and facilitates the application of machine learning techniques. Our framework consists of a set of spatial roles based on the theory of *holistic spatial semantics* with the intent of covering all aspects of spatial concepts, including both static and dynamic spatial relations. We illustrate our annotation scheme with many examples throughout the paper, and in addition we highlight how to connect to spatial calculi such as *region connection calculus* and also how our approach fits into related work.

1. Introduction

We define *spatial role labeling* as the task of identifying and classifying the spatial arguments of the spatial expressions mentioned in a sentence. For example, in the following sentence: “Give me that book on the table.”, the phrase headed by the token “book” is referring to a *trajector* object, the phrase headed by the token “table” is referring to the role of a *landmark* and these are related by the *spatial expression* “on” denoted as *spatial indicator*. Spatial role labeling is a key task for applications that are required to answer questions about, or have to reason about, spatial relationships. Examples include systems that perform text-to-scene conversion, robot navigation tasks, traffic management systems, geographical information systems (GIS) and many others.

Analogous to *semantic role labeling* (Màrquez et al., 2008), a spatial role labeling system needs to make several decisions about i) *spatial indicator identification* (i.e., which indicator features to use); ii) *argument identification* (i.e., which tokens have spatial roles with respect to these spatial indicators); iii) *argument classification* (i.e., which roles these tokens play); and iv) *spatial sense disambiguation*. So far there have not been any machine learning approaches that tackle this problem directly, which explains why there are virtually no corpora available today, but see (Mani et al., 2008). Here we aim at alleviating this situation by annotating a data set and formalizing the necessary spatial roles for spatial role labeling tasks. Recent spatial annotation schemes like SpatialML (Mani et al., 2008), STM (Pustejovsky and Moszkowicz, 2009) and also that of (Shen et al., 2009) are more limited than the kind of annotation scheme that would be needed for tackling the full problem of spatial role labeling, mainly because they were developed for particular tasks (e.g. GIS), or are limited in the number of spatial concepts they can handle.

One of our main technical contributions is a proposal for an annotation scheme for tagging the tokens that participate in expressing a spatial concept based on *holistic spatial semantic theory* (HSS) (Zlatev, 2003). In the proposed scheme we try to cover all aspects of concepts in spatial lan-

guage semantics (both static and dynamic) and map them to formal models. One of our immediate tasks is the preparation of a corpus for learning spatial relations based on this scheme, which we intend to make publicly available.

Extracting spatial information, through the mapping of natural language to a formal representation of spatial relations has several advantages for our long term research goal of working in a multi-modal environment. First of all, formal representations of spatial knowledge facilitate the visualization of spatial relations such that learning the connection between language and perception will be more feasible. Second, applying the same representation model for extraction from image (and video) data enables the combination of multi-modal features for better recognition and disambiguation in each modality. Third, a unified spatial representation for various modalities would enhance the interpretation of multi-modal information considerably, since information from different sources could be combined, and reasoning about this joint information would be possible.

Outline: The rest of this paper is organized in the following way: In Section 2, we introduce the spatial roles based on HSS. Section 3 presents our scheme for annotating the data, in Section 4 mapping spatial terms to spatial ontologies are described and in section 5 we conclude.

2. Holistic spatial semantics

An approach to spatial semantics that has the utterance (itself embedded in discourse and a background of practices) as its main unit of analysis, rather than the isolated word is characterized as *holistic*. Such an approach aims at determining the semantic contribution of each and every element of the spatial utterance in relation to the meaning of the whole utterance. One major advantage of such an approach is that it does not limit the analysis to a particular linguistic form, form class (e.g. prepositions), or theoretically biased grammatical notion (Zlatev, 2003). Our annotation scheme is based on this holistic theory and it helps to map the language to a formal spatial representation. In addition it is easily applicable for annotating spatial roles in image data (Hollink et al., 2004).

TRAJECTOR(id, token)
LANDMARK(id, token, path)
SPATIAL-INDICATOR(id, token, general-type, specific-type, spatial-value)
MOTION-INDICATOR(id, token)
SR(id, trajectory, landmark, spatial-indicator, frame-of-reference, motion-indicator)

Table 1: Relational representation of the annotation scheme

The semantic spatial components in HSS theory are *trajectory*, *landmark*, *frame of reference*, *path*, *region*, *direction* and *motion* (Zlatev., 2007). Each spatial relation (SR) consists of these components: (Zlatev, 2003; Zlatev., 2007).

Trajectory : the entity (object, person or event) whose location or motion is of relevance.

Landmark : the reference entity in relation to which the location or motion of the trajectory is determined.

Region : a region of space which is always defined in relation to a landmark, e.g. the interior or exterior.

Path : a most schematic characterization of the trajectory of actual or virtual motion in relation to a region defined by the landmark in terms of its *beginning*, *middle* and *end*, similar to the distinction *source/medium/goal*. HSS distinguishes between path and region.

Motion : a binary component indicating whether there is perceived motion or not.

Direction : denotes a direction along the axes provided by the different frames of reference, in case the trajectory of motion is not characterized in terms of its relation to the region of a landmark.

Frame of reference : one of three types of frame of reference; *intrinsic*, *relative* or *absolute*.

In spatial information theory the relations and properties are usually grouped into the domains of *topological*, *directional*, and *distance* relations and also shape (Stock, 1997). Hence we label the sentences in a way that enables us to identify the HSS components on the one hand, and classify or map those components onto the spatial relation categories and their more finely-grained types on the other.

3. Spatial relations annotation approach

We have designed an annotation scheme for tagging natural language with spatial roles that takes into account aforementioned definitions and the concepts described in HSS theory. In Table 1 each token is a unique key related to a word or a set of words. The related tokens for trajectories and landmarks in each sentence are identified and annotated. Each token can have different roles, thereby participating in various spatial relations.

Each landmark is related to a path which characterizes a path or a complex landmark in terms of its beginning, middle and end points. If these parts are not relevant, then a zero value is assigned. Hence, the attribute values for path

are {BEGIN, MIDDLE, END, ZERO}. This helps presenting relations like “in between” or “in the middle” which describe the location of one object referring to more than one reference object and also when there is motion.

In our scheme we tag the words which define constraints on the spatial properties – such as the location of the trajectory with respect to the landmark – as a spatial indicator (e.g. *in*, *on*). In fact, spatial indicators explain the type of the spatial relations and can express the region and direction semantics and even distances. Thus, we assign a general type attribute to indicators which has one of the values {REGION, DIRECTION, DISTANCE}. The specific relation expressed by the indicators is stated in a specific-type attribute. If the general-type is REGION then we map this onto topological relations in a formalism like RCC8 (Stock, 1997) (or any similar, topological model). Using the RCC8 (region connection calculus) model we classify the relations as {EC(externally connected), DC(disconnected), EQ(equal), PO(partially overlapping), TPP(tangential proper part), NTPP(non-tangential proper part), TPPI(tangential proper part inverse), NTPPI(non-tangential proper part inverse)} and assign these to the spatial-value attribute. If an indicator of direction is observed then the specific type can be {ABSOLUTE, RELATIVE}. For absolute directions we use 8 directions {S(south), W(west), N(north), E(east), NE(northeast), SE(southeast), NW(northwest), SW(southwest)} and for relative directions we use the six directions {LEFT, RIGHT, FRONT, BEHIND, ABOVE, BELOW}.

In case the general type of the indicator is DISTANCE then it is classified as {QUALITATIVE, QUANTITATIVE}. For qualitative distances we use a predefined set of terms including *far*, *near*, etc., and for quantitative distances the numbers and values in the text form the key distance information. This information can be mapped onto one formal absolute or relative model after the extraction of the relations. The spatial terms which are indicators of motion are tagged as so (e.g. propositional verbs). Each spatial relation is tagged by its frame of reference because this concept is useful (and often necessary) for visualization of the relations. Furthermore it aids in mapping the relations onto a unified frame of reference in the future. The frame of reference has one of the values {INTRINSIC, RELATIVE, ABSOLUTE}.

Our scheme can be easily extended to include various forms of *temporal* information. One can add *time stamps* to each specific spatial relation, expressing when they did hold. Since we have included the concept of a path, having motion is not problematic. Temporal aspects could represent the time of being at a particular part of the related path(=landmark) mentioned in the path attribute value. An-

other extension to this scheme could be the addition of *shape* and *size* attributes to the trajector and even landmark objects because it would enable a better connection to perception and related images/videos data in multi-modal setting. Note that in this paper we focus solely on the spatial aspects of our annotation scheme. The examples in the following subsections illustrate how our annotation scheme works using an XML notation for tags.

3.1. Annotation guidelines

Semantic annotation of a corpus is a challenging, and ambiguous task (Mooney, 2008). We have investigated several kinds of spatial descriptions to facilitate the annotation process, we have defined guidelines to make the task easier and less ambiguous. Below we list a set of questions which annotators should ask themselves while annotating. The general goal of applying machine learning to spatial role labeling is to obtain these answers automatically from text. The annotations are performed at the sentence level. The annotators use their understanding of explicit words and their senses. The questions are:

1. Is there a spatial description in the sentence?
2. Which words are the indicators of the spatial information?
3. Which words are the arguments of those spatial indicators?
4. Which tokens have a role of target for the spatial indicator and of *what* is the spatial description described?
5. Which tokens have the role of landmark for the spatial indicator ? (Is there a landmark?)
6. Is there a "motion"? and if so, which tokens are the motion indicator?
7. What is the frame of reference?
8. How can we map this spatial relation to a formal spatial relation in our predefined set of various spatial information.

To aid dealing with ambiguities in the annotation task we categorize the spatial descriptions into *complex* and *simple* descriptions. The annotation guidelines and examples are described first in the simple case and later extended to complex cases. The answer to question 8 requires the selection of a formal spatial representation or a spatial ontology.

3.2. Simple descriptions

We define a *simple description* as a spatial description which includes one target, at most one landmark and at most one indicator. For answering the first question mentioned in the previous section we consider the conventional specifications of the location or change of location (i.e. translocation) of an entity in space as a spatial description such that conversational implications are excluded. For example, the answer *He is washing the dishes* to the question *Where is he?* could – with some inference on this conversation – imply *He is in the kitchen*, but we do not consider that here. Examples of simple descriptions are:

EXAMPLE 1.

- a. There is a meeting on Monday.
- b. There is a book on the table.

Sentence 1 has the same structure of a spatial description with the preposition "on" which can be a spatial indicator but "on Monday" is a temporal expression, so there is no

spatial description, but in sentence 2, there is a spatial description about the location of a book. In case there is a spatial description in the sentence, its components are tagged according to the aforementioned definitions.

Trajector

The following sentences show the way *trajector* should be annotated.

EXAMPLE 2.

- a. She is at school.
<TRAJECTOR id='1'> She </TRAJECTOR>
- b. She went to school.
<TRAJECTOR id='1'> She </TRAJECTOR>
- c. The book is on the table.
<TRAJECTOR id='1'> The book </TRAJECTOR>
- d. She is playing in her room.
<TRAJECTOR id='1'> She </TRAJECTOR>
- e. Go left!
<TRAJECTOR id='1'> you </TRAJECTOR>

For instructions as in example 2e. the trajector is implicit and "you" is added as trajector. This problem is mostly with the instructions grammar and can be solved in the preprocessing beforehand.

Landmark

A *landmark* is tagged according to its aforementioned definition. The source of ambiguity here is that sometimes an explicit landmark is not always needed, for example in the case of directions. The second more difficult case is when the landmark is deleted by ellipsis and it is implicit. In such cases we annotate the landmark by NIL such that handling the ellipsis of it becomes easier later on.

EXAMPLE 3.

- a. The balloon passed over the house.
<LANDMARK id='1' path='ZERO'>the house</LANDMARK>
- b. The balloon passed over.
<LANDMARK id='1' path='ZERO'>NIL</LANDMARK>
- c. The balloon went up.
<LANDMARK id='1' path='ZERO'>NIL</LANDMARK>
- d. The balloon went over there.
<LANDMARK id='1' path='ZERO'>there</LANDMARK>
- e. John went out of the room. <LANDMARK id='1' path='BEGINNING'> the room </LANDMARK>
- f. John went through the room.
<LANDMARK id='1' path='MIDDLE'>the room</LANDMARK>
- g. John went into the room.
<LANDMARK id='1' path='END'>the room</LANDMARK>
- h. John is in the room.
<LANDMARK id='1' path='ZERO'>the room</LANDMARK>

In example 3c. we have a relative direction, and thus an implicit landmark should be there. In example 3d, "there" should be resolved in preprocessing or post-processing and the annotators should do not concern the reference resolution here. Another special case happens when there is a motion with spatial effect and the landmark is like a path and the indicators indicate a relation in some

part of the path. In that case a path attribute is set; see the examples 3e to 3h.

Spatial indicator

The spatial terms, or spatial indicators, are mostly prepositions but can also be verbs, nouns and adverbs or a combination of them. We annotate each indicator and fill in the spatial attributes of the indicator.

EXAMPLE 4.

a. He is in front of the bush.

```
<SPATIAL-INDICATOR id='1' general-type='DIRECTION'
specific-type='RELATIVE' spatial-value='FRONT'> in front of</SPATIAL-INDICATOR>
```

b. Sit behind the bush.

```
<SPATIAL-INDICATOR id='1' general-type='DIRECTION'
specific-type='RELATIVE' spatial-value='BEHIND'> behind</SPATIAL-INDICATOR>
```

c. John is in the room.

```
<SPATIAL-INDICATOR id='1' general-type='REGION'
specific-type='RCC8' spatial-value='TPP'> in</SPATIAL-INDICATOR>
```

The difficulty of annotation here is how to fill in the indicator attributes. In other words the mapping between linguistic terms and formal relations like RCC is not always clear and easy. We discuss this later.

Motion indicator

These are mostly the prepositional verbs but we leave it open for other semantical categories like adverbs, etc. For the moment we just tag them as indicators but later we can map them to motion verb classes.

EXAMPLE 5.

a. The bird flew to its nest.

```
<MOTION-INDICATOR id='1'> flew to</MOTION-INDICATOR>
```

We tag the token “flew to” as the indicator as the preposition affects the semantics of the motion.

Spatial relation

The components recognized by the annotators should be put in relations called *spatial relations* (SR). In a simple description it is not that difficult because we have one trajector, one/zero landmark and one spatial indicator, so these constitute at least one clear coarse spatial relation which should be tagged. For each relation we add the frame of reference as an attribute. If a motion indicator is present which is related to the spatial relation and the location of the trajector then this is mentioned in the attributes of the spatial relation.

EXAMPLE 6.

a. She is at school.

```
<TRAJECTOR id='1'> She</TRAJECTOR>
<LANDMARK id='1' path='ZERO'> school</LANDMARK>
<SPATIAL-INDICATOR id='1' general-type='REGION'
specific-type='RCC8' spatial-value='TPP'> at</SPATIAL-INDICATOR>
<SR id='1' trajector='1' landmark='1' spatial-indicator='1'
frame-of-reference='INTRINSIC' motion-indicator='NIL'/>
```

b. She went to school.

```
<TRAJECTOR id='1'> She</TRAJECTOR>
<LANDMARK id='1' path='END'> school</LANDMARK>
<SPATIAL-INDICATOR id='1' general-type='REGION'
specific-type='RCC8' spatial-value='TPP'> to</SPATIAL-INDICATOR>
<MOTION-INDICATOR id='1'> went to</MOTION-INDICATOR>
<SR id='1' trajector='1' landmark='1' spatial-indicator='1'
frame-of-reference='INTRINSIC' motion-indicator='1'/>
```

c. The book is on the table.

```
<TRAJECTOR id='1'> The book</TRAJECTOR>
<LANDMARK id='1' path='ZERO'> table</LANDMARK>
<SPATIAL-INDICATOR id='1' general-type='REGION'
specific-type='RCC8' spatial-value='EC'> on</SPATIAL-INDICATOR>
<SR id='1' trajector='1' landmark='1' spatial-indicator='1'
frame-of-reference='INTRINSIC' motion-indicator='NIL'/>
```

d. She is playing in her room.

```
<TRAJECTOR id='1'> She</TRAJECTOR>
<LANDMARK id='1' path='ZERO'> her room</LANDMARK>
<SPATIAL-INDICATOR id='1' general-type='REGION'
specific-type='RCC8' spatial-value='TPP'> in</SPATIAL-INDICATOR>
<MOTION-INDICATOR id='1'> playing</MOTION-INDICATOR>
<SR id='1' trajector='1' landmark='1' spatial-indicator='1'
frame-of-reference='INTRINSIC' motion-indicator='1'/>
```

3.3. Complex Descriptions

In this section we illustrate how our scheme is able to handle complex spatial descriptions. In (Barclay and Galton, 2008) three classes of complex description forms are identified and we give examples of them here:

I: Complex locative statements are locative phrases with more than one reference or as we call here landmarks. The explanations are about one target, meanwhile some relations can be inferred between landmarks, but for the annotation – annotators should not do additional reasoning steps – only what is explicitly expressed in the sentence should be tagged. Therefore the annotation in example 7, is a straightforward annotation of various possible spatial relations.

EXAMPLE 7.

The vase is in the living room, on the table under the window.

```
<TRAJECTOR id='1'> The vase</TRAJECTOR>
<LANDMARK id='1' path='ZERO'> the living room</LANDMARK>
<LANDMARK id='2' path='ZERO'> the table</LANDMARK>
<LANDMARK id='3' path='ZERO'> the window</LANDMARK>
<SPATIAL-INDICATOR id='1' general-type='REGION'
specific-type='RCC8' spatial-value='NTPP'> in</SPATIAL-INDICATOR>
<SPATIAL-INDICATOR id='2' general-type='REGION'
specific-type='RCC8' spatial-value='EC'> on</SPATIAL-INDICATOR>
<SPATIAL-INDICATOR id='3' general-type='DIRECTION'
specific-type='RELATIVE' spatial-value='BELOW'> under</SPATIAL-INDICATOR>
```

```

<SR id='1' trajector='1' landmark='1' spatial-indicator='1'
frame-of-reference='INTRINSIC' motion-indicator='NIL'/>
<SR id='2' trajector='1' landmark='2' spatial-indicator='2'
frame-of-reference='INTRINSIC' motion-indicator='NIL'/>
<SR id='3' trajector='1' landmark='3' spatial-indicator='3'
frame-of-reference='INTRINSIC' motion-indicator='NIL'/>

```

II: Path and route descriptions are possibly the most important for multi-modal systems. In this kind of descriptions a *focus shift* can happen. It means the speaker explains one target referring to some landmarks but at some point explains another object or landmark, i.e. the focus shifts to another entity as target. Annotators should recognize this focus shift and annotate the rest of the phrases by the new target. The following example shows such an expression but here we only tagged the spatial indicators and not the motion indicators to simplify its representation.

EXAMPLE 8.

The man came from between the shops, ran along the road and disappeared down the alley by the church.

```

<TRAJECTOR id='1' > the man </TRAJECTOR>
<LANDMARK id='1' path='BEGINNING'> the shops
</LANDMARK>
<LANDMARK id='2' path='MIDDEL'> the road </
LANDMARK>
<LANDMARK id='3' path='END'> the alley <LAND-
MARK/>
<TRAJECTOR id='2' >the alley </TRAJECTOR >
<LANDMARK id='4' path='ZERO'> the church </
LANDMARK>
<SPATIAL-INDICATOR id='1' general-type='Region'
specific-type='RCC8' spatial-value='IN'> between </
SPATIAL-INDICATOR >
<SPATIAL-INDICATOR id='2' general-type='Region'
specific-type='RCC8' spatial-value='EC'> along </
SPATIAL-INDICATOR >
<SPATIAL-INDICATOR id='3' general-type='Direction'
specific-type='Relative' spatial-value='Below'> down </
SPATIAL-INDICATOR >
<SPATIAL-INDICATOR id='4' general-type='Region'
specific-type='RCC8' spatial-value='DC'> by </SPATIAL-
INDICATOR >
<SR id='1' trajector='1' landmark='1' spatial-indicator='1'
frame-of-reference='INTRINSIC' motion-indicator='NIL'/>
<SR id='2' trajector='1' landmark='2' spatial-indicator='2'
frame-of-reference='INTRINSIC' motion-indicator='NIL'/>
<SR id='3' trajector='1' landmark='3' spatial-indicator='3'
frame-of-reference='RELATIVE' motion-indicator='NIL'/>
<SR id='4' trajector='2' landmark='4' spatial-indicator='4'
frame-of-reference='INTRINSIC' motion-indicator='NIL'/>

```

III: Sequential scene descriptions are linked descriptive phrases. After each description usually an *object focus shift* happens.

EXAMPLE 9.

Behind the shops is a church, to the left of the church is the town hall, in front of the town hall is a fountain.

```

<TRAJECTOR id='1'> church </TRAJECTOR>
<LANDMARK id='1' path='ZERO'> shops </
LANDMARK>
<SPATIAL-INDICATOR id='1' general-type='Direction'
specific-type='Relative' spatial-value='Behind'> behind
</SPATIAL-INDICATOR >

```

```

<TRAJECTOR id='2' > town hall </TRAJECTOR>
<LANDMARK id='2' path='ZERO'> church </
LANDMARK>
<SPATIAL-INDICATOR id='2' general-type='Direction'
specific-type='Relative' spatial-value='Left'> to the left of
</SPATIAL-INDICATOR >
<TRAJECTOR id='1'> fountain </TRAJECTOR>
<LANDMARK id='2' path='ZERO'> town hall </
LANDMARK>
<SPATIAL-INDICATOR id='3' general-type='Direction'
specific-type='Relative' spatial-value='Front'> in front of
</SPATIAL-INDICATOR >
<SR id='1' trajector='1' landmark='1' spatial-indicator='1'
frame-of-reference='INTRINSIC' motion-indicator='NIL'/>
<SR id='2' trajector='2' landmark='2' spatial-indicator='2'
frame-of-reference='INTRINSIC' motion-indicator='NIL'/>
<SR id='3' trajector='3' landmark='3' spatial-indicator='3'
frame-of-reference='RELATIVE' motion-indicator='NIL'/>

```

In addition to the complex descriptions mentioned in (Barclay and Galton, 2008), the following examples show some additional special characteristics. The next example contains one indicator *for* for two relations.

EXAMPLE 10.

John left Boston for New York.

```

<TRAJECTOR id='1'> John </TRAJECTOR>
<LANDMARK id='1' path='BEGIN'>Boston </
LANDMARK >
<LANDMARK id='2' path='END'> New York </
LANDMARK >
<SPATIAL-INDICATOR id='1' general-type='Direction'
specific-type='Relative' spatial-value='NTPP'> for </
SPATIAL-INDICATOR >
<MOTION-INDICATOR id='1'> left </MOTION-
INDICATOR >
<SR id='1' trajector='1' landmark='1' spatial-indicator='1'
motion-indicator='1' frame-of-reference='ABSOLUTE' />
<SR id='2' trajector='1' landmark='2' spatial-indicator='1'
motion-indicator='1' frame-of-reference='ABSOLUTE'/>

```

In example 11 the focus shift is ambiguous. The phrase *on the left* can refer to the door or to the table. If more information is available (for example in a multi-modal context other information could come from video input) then we could estimate the likeliness of each alternative. In general, if an annotator is not sure about the reference then all the true relations are added. For the machine learning purposes, this is still a correct annotation because no additional inference is performed and both meanings can be extracted for the same sentence.

EXAMPLE 11.

The table is behind the door on the left.

```

<TRAJECTOR id='1'>The table </TRAJECTOR >
<LANDMARK id='1' path='ZERO'>the door </
LANDMARK >
<SPATIAL-INDICATOR id='1' general-type='Direction'
specific-type='Relative' spatial-value='BEHIND'> behind
</SPATIAL-INDICATOR >
<SPATIAL-INDICATOR id='2' general-type='Direction'
specific-type='Relative' spatial-value='LEFT'> on the left
</SPATIAL-INDICATOR >
<SR id='1' trajector='1' landmark='1' spatial-indicator='1'
frame-of-reference='RELATIVE' motion-indicator='NIL'/>

```

```
<SR id='2' trajector='1' landmark='NIL' spatial-indicator='2'
frame-of-reference='RELATIVE' motion-indicator='NIL'/>
<TRAJECTOR id='2' >The door </TRAJECTOR >
<SR id='3' trajector='2' landmark='NIL' spatial-indicator='2'
frame-of-reference='RELATIVE' motion-indicator='NIL' />
```

In example 12, there are one target, three landmarks and three indicators. Landmarks are geographically related but annotators should not use their background about this geographical information.

EXAMPLE 12.

```
He drives within New England from Boston to New York.
<TRAJECTOR id='1' > He </TRAJECTOR >
<LANDMARK id='1' path='ZERO'> New England
<LANDMARK >
<LANDMARK id='2' path='BEGIN'> Boston </
LANDMARK >
<LANDMARK id='3' path='END'> New York </
LANDMARK >
<SPATIAL-INDICATOR id='1' general-type='Region'
specific-type='RCC8' spatial-value='NTPP'> within </
SPATIAL-INDICATOR>
<SPATIAL-INDICATOR id='2' general-type='Region'
specific-type='RCC8' spatial-value='NTPP'> from </
SPATIAL-INDICATOR>
<SPATIAL-INDICATOR id='3' general-type='Region'
specific-type='RCC8' spatial-value='NTPP' > to </
SPATIAL-INDICATOR >
<MOTION-INDICATOR id='1'> drives </MOTION-
INDICATOR>
<SR id='1' trajector='1' landmark='1' spatial-indicator='1'
motion-indicator='1' frame-of-reference='ABSOLUTE'/>
<SR id='2' trajector='1' landmark='2' spatial-indicator='2'
motion-indicator='1' frame-of-reference='ABSOLUTE'/>
<SR id='3' trajector='1' landmark='2' spatial-indicator='3'
motion-indicator='1' frame-of-reference='ABSOLUTE'/>
```

Another possibility of having one indicator but with different meanings. In example 13, "cross" is a motion indicator and also spatial indicator.

EXAMPLE 13.

The car crosses the street.

To map the relations to formal representations, the ontology of the objects and also shape information of objects are necessary for the machine to learn from. We do not discuss these issues here further, but just show two examples.

EXAMPLE 14.

The room is at the back of the school.
The tree is at the back of the school.

In the first sentence the semantics of the spatial indicator *at the back of* is about an interior region of the school whereas in the second sentence it is about an exterior region.

3.4. Adding a temporal dimension

In the suggested scheme for each relation a time dimension can be easily added. Temporal analysis of sentences can be combined with spatial analysis to assign a value to the temporal dimension of each relation and the interpretation is the time instant at which the spatial relation holds.

Looking back to example 10, in the first spatial relation, the temporal dimension is related to *yesterday*.

EXAMPLE 16.

```
John left Boston for New York yesterday.
<TIME-INDICATOR id='1'> yesterday </TIME-
INDICATOR >
<SR id='1' trajector='1' landmark='1' spatial-indicator='1'
motion-indicator='1' frame-of-reference='ABSOLUTE'
time-indicator='1'/>
```

The analysis of temporal expressions could be done separately and only the time-indicator is added to related spatial relations.

4. Mapping to spatial ontologies

As already mentioned in the previous section, we need to connect the spatial relations in the natural language to formal spatial relations if we want to perform spatial reasoning about the information conveyed in a text. Our goal is to define a framework in such a way that we obtain the mapping through machine learning. The main challenge is that linguistic and logical formalizations of space are present at different levels within spatially aware information systems that interact with natural language dialogue systems. Therefore relations between both types of representations – linguistic and logical – that provide descriptions of the environment from different viewpoints, have to be aligned and integrated with each other (Hois and Kutz, 2008a).

Connections between viewpoints are strongly influenced by external factors, and so the relationship between instances in different domains is not deterministic and often ambiguous. On the language side, the state of the art linguistic categorization designed specifically for spatial descriptions has been developed in the *Generalized Upper Model* (GUM) (Bateman et al., 2007). It has been successfully applied in a natural language system (Ross et al., 2005) and its expressiveness has been evaluated by a linguistic corpora with more than 600 entries for both English and German.

Although the GUM covers all the spatial aspects in terms of semantics, it is not concerned with the pragmatic principles and distinctions associated with particular lexicogrammatical items and structures. However, a very fine-grained ontology that would actually incorporate all of them would render much less of a useful formal logical representation that could be used for general, formal reasoning about the spatial information. Therefore, we can see it as an expressive intermediate level between linguistic and logical representation, but the mapping to a formal logic is inevitable for spatial reasoning (Hois and Kutz, 2008b). In our annotation framework we can map each spatial indicator to a formal representation in RCC8, employing regions, six absolute and relative directions, and in addition, distances. This results in a general representation for spatial relations extracted from the text, which is useful for reasoning about them.

This being said, which specific model corresponds to the linguistic description and vice versa depends on external aspects. Whether one or more connections between language and space are necessary, and to what degree they hold, has to be determined based on indications from these external aspects. A very important aspect will presumably also be

the incorporation of discourse background knowledge to be able to perform this automatically (Hois and Kutz, 2008a).

5. Related work

Extraction of spatial relations from natural language through machine learning has not been investigated in a domain-independent way. Hence, usually limited languages with specific words and structure are considered. For example, several systems exist that extract information directly from text to determine spatial relationships between objects in a 3D scene to generate 3D scenes from these textual descriptions. In these systems the semantic models of spatial relations and their computational implementation is considered. However, they are restricted to simple narratives, typically invented by the authors themselves and do not consider a real corpus. A recent overview of such vision and language systems can be found in (Keller, 2003). More generally, dealing with spatial relations for reasoning and inference has been investigated in the literature in many different contexts (Stock, 1997) and many formalisms exist for the representation of, and reasoning with, spatial relations (Galton, 2009). However, in the context of machine learning usually a limited number of relations are defined and considered to keep the problem tractable. This is one of the reasons the problem has not been defined in a uniform way and no large enough corpus exists for training.

There are some research works focusing on annotating the spatial descriptions in natural languages and several proposed annotation schemes and terminologies exist for specifying and formulating the spatial relation components in the textual data. Examples are ACE, GUM, GML, KML, TRML and SpatialML. The concepts and the terminology of these schemes is highly affected by the domain of the data of interest and the related applications. However mostly these schemes could be mapped to each other when their application is the same. The most recently developed markup language for marking spatial relations is SpatialML (Mani et al., 2008). Compared to our scheme it uses PLACE tags to identify geographical features. SIGNAL, RLINK and LINK tags are defined to identify the directional and topological spatial relations between a pair of locations. Topological spatial relations in SpatialML also are connected to RCC8 relations. However SpatialML only considers static spatial relations and focuses on geographic domains. A similar scheme TimeML (Pustejovsky et al., 2005) exists for annotation of temporal relations in natural language. GUM (generalized upper model), as we have mentioned in the previous section, also aims at organizing spatial relations. The formulated relations are very expressive from a linguistic point of view but the ontology is very large and more fine-grained than what could be effectively learnable from a corpus. An interesting new XML scheme based on SpatialML and GUM was proposed in (Shen et al., 2009), targeting spatial relations in the Chinese language. It also deals with geographical information and defines two main tags of geographical entity and spatial expression. In (Pustejovsky and Moszkowicz, 2009), a spatio-temporal markup language for the annotation of motion predicates in text informed by a lexical semantic clas-

sification of these verbs, is proposed. The interesting point is that the proposed scheme seems suitable for tagging dynamic spatial relations, based on motions in the space and time. However the center of attention is the motion verbs and their spatial effects and not general spatial language.

Since our aim is to apply machine learning to learn to identify spatial relations in texts, annotated data is of interest, although we have already mentioned the general lack of such data in spatial domains. There are a few efforts towards creating annotated data sets at the language level of spatial relations. For example in (Li et al., 2006) the Chinese version of Aesops Fables has been labeled in terms of trajector, landmark and spatial expressions and turned into an evaluation database for the extraction of spatial relations. In the experiments however, only a binary classifier was used so far for the extraction of trajector. In (Shen et al., 2009) texts from a Chinese encyclopedia concerning geographical information is annotated using the XML scheme we have mentioned. GUM also is accompanied by an evaluation corpus containing a limited set of 600 sentences in German and English.

It should also be mentioned that FrameNet frames (Fontenelle, 2003) are a useful linguistic resource which can be very helpful for identifying spatial components in the sentence. Spatial relations can be seen, to some extent, as a part of frame-based semantic annotation. However there are various semantic frames which are related to spatial roles and semantics. Frames like LOCATIVE RELATION, SELF MOTION, PERCEPTION, BEING LOCATED seem most related to spatial semantics. Hence, using these semantic frames requires making a connection between the general spatial representation scheme and the specific frames that could be related to each word. Therefore defining a tag set is important to have a unified spatial semantic frame for spatial semantics and to integrate partial annotations that tend to be distributed over different layers (Kuroda et al., 2006). Towards this direction, in (Schuldes et al., 2009) a corpus is annotated (in German) for walking directions. The preprocessed texts are annotated on the following three levels: *pos lemma* (part-of-speech and lemma), *syn dep* (dependency relations) and *sem frame* (frames and semantic roles). For tagging walking directions on the semantic frame level, annotation was carried out using FrameNet frames.

6. Conclusion

In this paper we have introduced an annotation scheme for natural language that supports various kinds of spatial information, including static and dynamic spatial relations. The annotation scheme is based on the ideas of holistic spatial semantics. Spatial roles, their arguments and indicators can be tagged using the annotation scheme with the main goal being that of building an annotated corpus. Such a corpus will be used for training machine learning methods for the recognition of the spatial roles in text, and as ground truth data for evaluation.

As one of our immediate plans for the application of machine learning we will use advanced learning techniques that combine *relational* and *probabilistic* knowledge representation schemes cf. (De Raedt et al., 2008). The rich

semantic structures that spatial information can induce call for expressive models that can handle relations well (e.g. *the book is on the table*), and the intrinsic ambiguity of natural language introduces a need to deal with uncertainty. In addition, the combination of relational representations and probabilistic information opens up possibilities for approximate reasoning about the spatial knowledge extracted from the text. To facilitate this, we will connect the extracted relations to well-defined spatial reasoning models such as RCC8. Learning such mappings will also be beneficial for the connection of language to perception in a multi-modal setting.

Acknowledgements

This research is funded by the AMASS++ project (SBO-IWT060051) and the DBOF/08/043 grant from Katholieke Universiteit Leuven.

We also thank the anonymous reviewers for their valuable comments.

7. References

- M. Barclay and A. Galton. 2008. A scene corpus for training and testing spatial communications. In *Proceedings of the AISB 2008 Convention (Communication, Interaction, and Social Intelligence)*.
- J. Bateman, T. Tenbrink, and S. Farrar. 2007. The role of conceptual and linguistic ontologies in discourse. *Discourse Processes*, 44(3):175–213. Special Issue on Dialogue Modelling: Computational and Empirical Approaches.
- L. De Raedt, P. Frasconi, K. Kersting, and S. Muggleton, editors. 2008. *Probabilistic Inductive Logic Programming – Theory and Applications*, volume 4911 of *Lecture Notes in Computer Science*. Springer.
- T. Fontenelle. 2003. Framenet and frame semantics. *International Journal of Lexicography*, 16(3):231.
- A. Galton. 2009. Spatial and temporal knowledge representation. *Journal of Earth Science Informatics*, 2(3):169–187.
- J. Hois and O. Kutz. 2008a. Counterparts in language and space - similarity and S -connection. In *Proceedings of the International Conference on Formal Ontology in Information Systems (FOIS-2008)*, pages 266–279.
- J. Hois and O. Kutz. 2008b. Natural language meets spatial calculi. In *Proceedings of Spatial Cognition VI. Learning, Reasoning, and Talking about Space*, volume 5248 of *Lecture Notes in Computer Science*, pages 266–282. Springer.
- L. Hollink, G. Nguyen, G. Schreiber, J. Wielemaker, and M. Worring. 2004. Adding spatial semantics to image annotations. In *Proceedings of the 4th International Workshop on Knowledge Markup and Semantic Annotation at ISWCO*, pages 31–40.
- J. D. Kelleher. 2003. *A Perceptually Based Computational Framework for the Interpretation of Spatial Language*. Ph.D. thesis, School of Computing – Dublin City University.
- K. Kuroda, M. Utiyama, and H. Isahara. 2006. Getting deeper semantics than Berkeley framenet with MSFA. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*.
- H. Li, T. Zhao, S. Li, and Y. Han. 2006. The extraction of spatial relationships from text based on hybrid method. In *Proceedings of the IEEE International Conference on Information Acquisition*.
- I. Mani, J. Hitzeman, J. Richer, D. Harris, R. Quimby, and B. Wellner. 2008. SpatialML: Annotation scheme, corpora, and tools. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*.
- L. Màrquez, X. Carreras, K. C. Litkowski, and S. Stevenson. 2008. Semantic role labeling: An introduction to the special issue. *Computational Linguistics*, 34(2):145–159.
- R.J. Mooney. 2008. Learning to connect language and perception. In *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence, (AAAI)*, pages 1598–1601. AAAI Press.
- J. Pustejovsky and J. L. Moszkowicz. 2009. Integrating motion predicate classes with spatial and temporal annotations. In *Proceedings of CoLing 2008: Companion volume D: Posters and Demonstrations*, pages 95–98.
- J. Pustejovsky, R. Ingria, R. Sauri, J. Castaño, J. Littman, R. Gaizauskas, A. Setzer, G. Katz, and I. Mani. 2005. The specification language TimeML. In I. Mani, J. Pustejovsky, and R. Gaizauskas, editors, *The language of time: a reader*, chapter 27, pages 545–557. Oxford University Press, Oxford.
- R. Ross, H. Shi, T. Vierhuff, B. Krieg-Brückner, and J. Bateman. 2005. Towards dialogue based shared control of navigating Robots. In *Proceedings of Spatial Cognition IV: Reasoning, Action, Interaction*, pages 478–499.
- S. Schuldes, M. Roth, A. Frank, and M. Strube. 2009. Creating an annotated corpus for generating walking directions. In *Proceeding of ACL-IJCNLP Workshop on Language Generation and Summarization*.
- Q. Shen, X. Zhang, and W. Jiang. 2009. Annotation of spatial relations in natural language. In *Proceedings of the International Conference on Environmental Science and Information Application Technology*, pages 418 – 421.
- O. Stock, editor. 1997. *Spatial and Temporal Reasoning*. Kluwer Academic Publishers.
- J. Zlatev. 2003. Holistic spatial semantics of Thai. *Cognitive Linguistics and Non-Indo-European Languages*, pages 305–336.
- J. Zlatev. 2007. Spatial semantics. In H. Cuyckens D. Geeraerts, editor, *The Oxford Handbook of Cognitive Linguistics*, chapter 13, pages 318–350.