Spatial Role Labeling Annotation Scheme

Parisa Kordjamshidi[1], Martijn van Otterlo[2], and Marie-Francine Moens[1]

[1] Katholieke Universiteit Leuven, Departement Computerwetenschappen
{parisa.kordjamshidi,sien.moens}@cs.kuleuven.be
[2] Radboud University Nijmegen m.vanotterlo@donders.ru.nl

1 Introduction

Given the large body of the past research on various aspects of spatial information, the main obstacles for employing machine learning for extraction of this type of information from natural language have been: a) the lack of an agreement on a unique semantic model for spatial information; b) the diversity of formal spatial representation models ; c) the gap between the expressiveness of natural language and formal spatial representation models and consequently; d) the lack of annotated data on which machine learning can be employed to learn and extract the spatial relations.

In this chapter we introduce a spatial annotation scheme for natural language that supports various aspects of spatial semantics, including static and dynamic spatial relations. The annotation scheme is based on the ideas of *holistic spatial semantics* as well as *qualitative spatial reasoning* models. Spatial roles, their relations and indicators along with their multiple formal meanings are tagged using the annotation scheme producing a spatial language corpus. The goal of building such a corpus is to produce a resource for training the machine learning methods for mapping the language to formal spatial representation models, and to use it as ground-truth data for evaluation.

We describe the foundations and the motivations for the concepts used in designing the proposed spatial annotation scheme in Section 2. We illustrate the scheme and its XML and relational representation by means of examples in Section 3. The investigated corpora, annotated data and the annotation challenges are described in Section 4. A review on the related works is provided in Section 5. We conclude in Section 6.

2 Annotation Scheme: Motivation and Foundation

In the proposed annotation scheme two main aspects of spatial information are considered. The first aspect concerns cognitive-linguistic models and the way that spatial concepts are expressed in the language, and the second is about formal models that are designed for spatial knowledge representation and reasoning independent of natural language. A scheme which covers these aspects will be able to connect natural language to formal models and make spatial reasoning based on text feasible. In the following sections we first point to the challenges in making a flexible connection between these two sides of spatial information and after that we describe the main elements that form the basis of the proposed scheme.

2.1 Two Layers of Semantics

Spatial language can convey complex spatial relations along with polysemy and ambiguity present in natural language [8]. Linguistic constructs can express highly complex, relational structures of objects, spatial relations between them, and patterns of motion through space relative to some reference point.

In contrast to natural language, formal spatial models focus on one particular spatial aspect such as orientation, topology or distance and specify its underlying spatial logic in detail [15]. These formal models enable automatic spatial reasoning that is difficult to perform given solely natural language expressions.

However, there is a gap between the level of expressivity and specification of natural language and spatial calculi models [4]. Huge spatial ontologies are needed to be able to represent the spatial semantics expressed in the linguistic expressions. Hois and Kutz investigate the alignment between the linguistic and logical formalizations [14]. 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. Generalized Upper Model (GUM) is the state-of-the-art example of such an ontology [3,44]. The GUM-Space ontology is a linguistically motivated ontology that draws on findings from empirical cognitive and psycholinguistic research as well as on results from theoretical language science [5]. However, for a machine learning practice, mapping to an intermediate linguistic ontology with a fairly large and fine-grained division of concepts is to some extent difficult because first it implies the need for a huge labeled corpus if a supervised setting is considered, second the semantic overlap between the included relations in the large ontologies makes the learning model more complex.

In addition, although the logical reasoning is computationally possible using an ontology such as GUM, the kind of spatial reasoning which is provided by calculi models is not feasible. Hence to perform actual spatial reasoning another layer of bridging between the GUM representation and calculi models is required [14]. Therefore, we use a layer of formal representation models in our proposed scheme besides the linguistically motivated ontologies. However, to alleviate the gap explained above we propose to map the linguistic expressions to multiple calculi. This issue is reflected in our annotation scheme and will be discussed in the following sections. For the sake of conceptual modularity and computational feasibility our spatial scheme is divided into two abstraction layers of cognitive-linguistic and formal models [4,26,23]:

1. A layer of **linguistic conceptual representation** called spatial role labeling (SpRL), which predicts the existence of spatial information at the sentence level by identifying the words that play a particular spatial role as well as their spatial relationship [27];

2. A layer of **formal semantic representation** called spatial qualitative labeling (SpQL), in which the spatial relation is described with semantic attribute values based on qualitative spatial representation models (QSR) [12,24].

In our conceptual model we argue that mapping the language to multiple spatial representation models could help the problem of the existing gap to some extent. Because various formal representations capture the semantics from different angles, their combination covers various aspects of spatial semantics needed for locating the objects in the physical space. Hence, the SpQL has to contain multiple calculi models with a practically acceptable level of generality. Moreover, mapping to spatial calculi forms the most direct approach for automatic spatial reasoning compared to mapping to more flexible intermediated ontologies. However, we believe that this two layered model which can be considered as a *lightweight ontology* does not yield sufficient flexibility for ideal spatial language understanding. As in any other semantic tasks in natural language additional layers of *discourse* and *pragmatics* must be worked out, which is not the focus of this work.

2.2 Holistic Spatial Semantics

One part of our proposed scheme is based on the *holistic spatial semantics* theory. 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. The main spatial concepts considered in this theory are the following.

Trajector: The entity whose location or position is described. It can be static or dynamic; persons, objects, or events. Alternative common terms include local/figure object, locatum, referent, or target.

Landmark: The reference entity in relation to which the location or the motion of the trajector is specified. Alternate terms are reference object, ground, or relatum.

Region: This concept denotes a region of space which is defined in relation to a landmark. By specifying a value such as interior or exterior for this category, the trajector is related more specifically and more precisely with respect to the landmark.

Path: It is a most schematic characterization of the trajector of actual or virtual motion in relation to a region defined by the landmark. In cognitive semantics this concept is used in two different ways, that is rich path or minimal path. The minimal path is represented by its *beginning*, *middle* and *end*, similar to the distinction *source/medium/goal*. The minimal path is enriched when its information is combined with region or place.

Motion: This concept also can be characterized in a rich or minimal way. In its minimal way, motion is treated as a binary component indicating whether there is perceived motion or not. The minimal representation of motion allows a clear separation from the path and direction, while the rich one conflates it with these.

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

Frame of reference: In general, a frame of reference defines one or more *reference points*, and possibly a coordinate system based on axes and angles. Three reference types can typically be grammaticalized or lexicalized in English: intrinsic, relative, and absolute [29]. Recently, more detailed distinctions were presented in [49], where spatial reference frames are represented and systematically specified by the spatial roles locatum, relatum, and (optional) vantage together with a directional system.

However, how these theoretical concepts are applied to linguistic descriptions, is a controversial question. The answer to this question has many challenges such as dealing with polysemy and characterizing the semantic and phonological poles of the language [53]. In the holistic approach a many-to-many mapping between semantic concepts and form classes is allowed [52]. For example, in general a specific word can contribute to expressing the concept of landmark as well as region or even path.

2.3 Qualitative Spatial Representation

The second part of the suggested scheme is based on qualitative spatial reasoning (QSR) models. QSR models are designed based on logical, geometrical or algebraic spatial semantics independent from natural language. However the *cognitive adequacy* of these models has been an important concern. *Cognitive adequacy* refers to the degree in which a set of concepts and relationships, and the computational inference over them is consistent with the mental conceptualization of humans and the way that a human reasons about those concepts and their relationships [42]. Two important reasons for paying attention to the qualitative approach are a) this model is closer to how humans represent and reason about commonsense knowledge; b) it is flexible in dealing with incomplete knowledge [41].

Three main aspects of spatial information are topological, directional and distal information which are somehow complementary information that could specify the location of the objects under consideration. Other aspects are size, shape, morphology, and spatial change (motion). Most of the qualitative spatial calculi focus on a single aspect, e.g., topology, direction, distance but recently there are combinatory models and tools that are able to reason based on multiple calculi models [41,51]. Here we briefly describe the main aspects of the spatial information that are the basis of the spatial meaning representation in the proposed scheme and the qualitative calculi models that are available for them.

Topological Relations Distinguishing topological relationships between spatial entities is a fundamental aspect of spatial knowledge. Topological relations are inherently qualitative and hence suitable for qualitative spatial reasoning. In reasoning models based on topological relations, the spatial entities are assumed to be regions rather than points, and regions are subspaces of some topological space [41]. A set of jointly exhaustive and pairwise disjoint relations, which can be defined in all topological models based on *parthood* and *connectedness* relations, are DC, EC, PO, EQ, TPP, NTPP, TPP^{-1}, NTPP^{-1}.

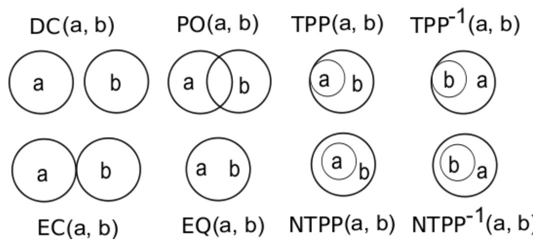

Fig. 1. The RCC-8 relations.

The best known approach in this domain is the Region Connection Calculus by Randell et al. [39] known as the RCC-8 model that we use to represent the topological relationships expressed in the language. RCC is heavily used in qualitative spatial representation and reasoning. The above relation symbols are abbreviations of their meanings (see Fig. 1): disconnected $\mathsf{DC}(\mathsf{a},\mathsf{b})$, externally connected $\mathsf{EC}(\mathsf{a},\mathsf{b})$, partial overlap $\mathsf{PO}(\mathsf{a},\mathsf{b})$, equal $\mathsf{EQ}(\mathsf{a},\mathsf{b})$, tangential proper-part $\mathsf{TPP}(\mathsf{a},\mathsf{b})$, non-tangential proper-part $\mathsf{NTPP}(\mathsf{a},\mathsf{b})$, tangential proper-part inverse $\mathsf{TPP}^{-1}(\mathsf{a},\mathsf{b})$, and non-tangential proper-part inverse $\mathsf{NTPP}^{-1}(\mathsf{a},\mathsf{b})$, which describe mutually exclusive and exhaustive overlap and touching relationships between two (well-behaved) regions in the space. The cognitive adequacy of this model is discussed in [42]. There are other topological models such as 9-intersection given by Egenhofer [9] which is based on interior, exterior, and boundary of regions.

Directional Relations Direction or orientation is also frequently used in linguistic descriptions about spatial relations between objects in qualitative terms, for example the expressions such as *to the left* or *in the north* are more often used than *45 degrees*. The frame of reference discussed in the previous section is an important feature to characterize directional relations. Absolute directions are in the form of {S(south), W(west), N(north), E(east), NE(northeast), SE(southeast), NW(northwest), SW(southwest)} in a geographical space. Relative directions are {Left, Right, Front, Behind, Above, Below} and used in a local space. These are only different in terminology compared to the former set of relations and can be adapted and used in qualitative direction calculus such as the cone-base, projection-based and double-cross models [41] (see Fig. 2). The double cross model (Fig. 2.c) assumes an additional axis and considers a perspective point in addition to the reference point.

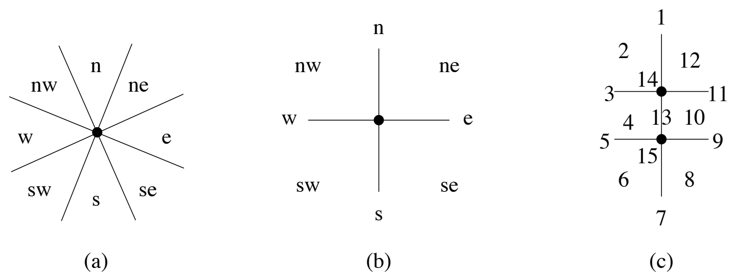

Fig. 2. Directional relations between points: (a) Cone-based model; (b) Projection-based model; (c) Double-cross model [41].

```
(1) TRAJECTOR(idT,token)
(2) LANDMARK(idL,token,path)
(3) SPATIAL_INDICATOR(idI,token)
(4) MOTION_INDICATOR(idM,token)
(5) SR(idS,idI,idT,idL,idM)
(6) SRType(idS,id_gtype,gtype,stype,sp_value,f_o_ref)
```

Table 1. Relational representation of the annotation scheme.

Distal Relations Along with the topology and direction, distance is one of the most important aspects of the space. Distance is a scalar entity and can be represented *qualitatively* such as *close*, *far* or *quantitatively* such as *two meters far*. Distances are also categorized as being either *absolute* or *relative*. The absolute distance describes the distance between two entities and the relative distance describes the distance between two entities compared to a third one. The computational models for distances often consider spatial entities as points. For more information about the various models for distal reasoning see [41,51].

3 Annotation Scheme: Relational Representation

We design an annotation scheme for tagging natural language with spatial roles, relations and their meaning. We take into account the cognitive-linguistic spatial primitives according to the theory of holistic spatial semantics as well as spatial relations according to the well-known qualitative spatial representation models described in Section 2.3. Table 1 shows the relational representation of the proposed spatial scheme. We describe these relations and the used terminology in the following.

In all these relations a *token* can be a word or a set of words. Each token that identifies a spatial role is assigned a unique *key*. Each token can play multiple roles as trajector or landmark in the sentence, thereby participating in various spatial relations. Each token is assigned a new identifier for each role that it plays. As it is shown in table 1,

In relation `(1)`, `idT` is an identifier that identifies a *token* that plays the role of *trajector*.

In relation `(2)`, `idL` is an identifier that identifies a *token* that plays the role of *landmark*. Each landmark is related to a *path* which characterizes a path or a complex landmark with a value in `{BEGIN,MIDDLE,END,ZERO}`. `ZERO` value is assigned when the path is not relevant.

In relation `(3)`, `idI` is an identifier that identifies a token that indicates the existence of a spatial relation and is called *spatial indicator*. According to the HSS theory [52], the relationship between trajector and landmark is not expressed directly but mostly via the region or direction concepts. We abstract from the semantics of these bridging concepts and tag the tokens which define constraints on the spatial properties- such as the location of the trajector with respect to the landmark- as a *spatial indicator* (e.g. *in*, *on*). A spatial indicator signals the existence of a spatial relation independent from its semantics.

In relation `(4)`, `idM` is an identifier that identifies a token (a word here) that indicates the existence of any kind of motion with a spatial influence in the sentence.

In relation `(5)`, we present a complex relation which links all the elements that are a part of a whole *spatial configuration* containing the identifiers of the above mentioned relations. This relation, which is named as `SR`, is identified by the identifier `idS` to be used in describing its semantic properties in relation `(6)`. We later refer to this relation as *spatial relation*.

In relation `(6)`, the type of the semantics of the *spatial configuration* is determined regarding the involved components. Since all of these components (trajector, landmark, etc.) contribute to the semantics of the relation, the fine-grained semantics are assigned to the whole *spatial configuration* which was identified by `idS`. We allow multiple semantics to be assigned to one spatial configuration, hence the additional identifier `id_gtype` is used to identify each related type. All the above mentioned elements are related to the cognitive elements of the spatial configuration but this relation is about the *formal representation* of the semantics which we now clarify in detail.

Formal semantics. As discussed in Section 2.1, to cover all possible semantic aspects of a linguistic expression about a spatial configuration, we allow multiple semantics to be assigned to it. For each spatial relation/configuration, we assign one or more general types which have one of the values `{REGION,DIRECTION,DISTANCE}`. With respect to each general type a specific type is established. The specific type of a relation that is expressed by the configuration is stated in the `stype` attribute. If the `gtype` is `REGION` then we set `stype` with topological relations in a formalism like RCC8 [48] (any other topological model might be used here). If an indicator of direction is observed then the `stype` can be `{ABSOLUTE,RELATIVE}`. The absolute and relative direction values are discussed in Section 2.3. In case the `gtype` of the spatial relation 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. Finally, each spatial relation given its general type identifier is tagged by a frame of reference f_o_ref with a value in {INTRINSIC,RELATIVE,ABSOLUTE}. The chosen relational representation can be easily represented in an XML format or stored in a relational database which makes the use of annotated data for machine learning models, retrieval systems, or even as a resource for the semantic web very convenient.

3.1 Annotation Approach

Semantic annotation of a corpus is a challenging, and ambiguous task [36]. We have investigated several kinds of spatial descriptions to find an appropriate corpus for annotation, and we have defined guidelines to make the task easier and less ambiguous. The list below is a set of questions which annotators should ask themselves while annotating. 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 any direct (without commonsense implications) spatial description in the sentence?
2. Which words are the indicators (that is trigger or signal) of the spatial information?
3. Which words are the arguments of those spatial indicators (semantically connected: see the following detailed questions)?
4. Which tokens have the role of trajector for the spatial indicator and *what* is the spatial entity (e.g. object, person) described?
5. Which tokens have the role of landmark for the spatial indicator? (how the trajector location is described and is there any landmark?)
6. Link the above three spatial concepts as one spatial relation.
7. If the trajector/landmark are conjunctive phrases, annotate all the components separately and generate all possible spatial relations.
8. If you can not complete the spatial relation (implicit roles in the sentence) annotate those roles as a null/undefined role but finding the spatial indicator is always required.
9. Is there a complex landmark? if so, can we describe it in terms of a point in a path (beginning, middle, end)?
10. Is there any motion with spatial effect? if so, which tokens trigger it and are the motion indicator?
11. What is the frame of reference? Indicate maximum one frame for each relation.
12. Given a predefined set of formal spatial relations, imagine the trajector and landmark as two regions: which formal relation describes the spatial semantics the best?
13. Does the spatial relation imply directional semantics?
14. Does the spatial relation imply regional semantics?
15. Does the spatial relation provide any information about the distance?
16. Is one formal semantic type enough for a rough visualization/schematization of the meaning of the spatial relation, and locating the objects in the space?

17. Do we need multiple annotations to capture the semantics of the relation, and to be able to draw a rough sketch? Annotate with as many as possible semantics that are covered by the relation.
18. When annotating multiple semantics, choose only one fine-grained type for each general category of {direction, region, distance}.

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 answers to questions $12 - 18$ require the selection of a formal spatial representation which can involve multiple choices.

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 spatial 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 – 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.

Example 1.*a*. has the same structure of a spatial description with the preposition "on" but "on Monday" is a temporal expression, so there is no spatial description, but in Example 1.*b*., 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='0'> NIL </TRAJECTOR>

When the trajector is implicit as in example 2.*e*. "NIL"is added as trajector.

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.

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 3.*c.* we have a relative direction, and thus an implicit landmark should be there. In example 3.*d.* "there" should be resolved in preprocessing or postprocessing and the annotators should not be concerned about 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 3.*e.* to 3.*h*.

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 signal of the existence of the spatial information in the sentence as spatial indicator. EXAMPLE 4.

a. He is in front of the bush.
<SPATIAL-INDICATOR id='1' > in front of</SPATIAL-INDICATOR>
b. Sit behind the bush.
<SPATIAL-INDICATOR id='1' > behind </SPATIAL-INDICATOR>
c. John is in the room.
<SPATIAL-INDICATOR id='1' > in </SPATIAL-INDICATOR>

Motion Indicator These are mostly the prepositional verbs but we leave it open for other semantical categories like adverbs, etc. In this scheme we just tag them as indicators but a further extension is to 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 because the preposition affects the semantics of the motion.

Spatial Relation and Formal Semantics The spatial configuration’s components recognized by the annotators should be put in relations called *spatial relations* (SR). In a simple description it is often easy because we have maximum one trajector, maximum one landmark and only one spatial indicator, so these constitute at least one clear coarse spatial relation to be tagged. If a motion indicator is present which is related to the spatial relation and the location of the trajector then the identifier of the motion also is added to the spatial relation. Each spatial relation is associated with a number of formal semantics, for example, when it implies both topological and directional information. The difficulty when annotating is how to fill in the semantic attributes. In other words the mapping between linguistic terms and formal relations like RCC is not always clear and easy. We discuss this later in this chapter. For each type of relation we add a new frame of reference as an attribute. For example, the frame of reference is more relevant for the directional relationships compared to topological relationships. Hence, it makes more sense to assign this concept according to each specific annotated type of semantics.

Example 6.

a. She is at school.
<TRAJECTOR id=‘1’ > She</TRAJECTOR>
<LANDMARK id=‘1’ path=’ZERO’>school</LANDMARK>
<SPATIAL-INDICATOR id=‘1’ > at </SPATIAL-INDICATOR>
<SR id=‘1’ trajector=‘1’ landmark=‘1’ spatial-indicator=‘1’ motion-indicator=‘NIL’/>
<SR id=‘1’ SRtype id=‘1’ general-type=‘REGION’ specific-type=‘RCC8’ spatial-value=‘TPP’ frame-of-reference=‘INTRINSIC’/>
b. She went to school.
<TRAJECTOR id=‘1’> She</TRAJECTOR>
<LANDMARK id=‘1’ path=’END’> school </LANDMARK>
<SPATIAL-INDICATOR id=‘1> 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’/>
<SR id=‘1’ SRtype id=‘1’ general-type=‘REGION’ specific-type= ‘RCC8’ spatial-value=‘TPP’ frame-of-reference=‘INTRINSIC’/>
c. The book is on the table.
<TRAJECTOR id=‘1’> The book </TRAJECTOR>
<LANDMARK id=‘1’ path=‘ZERO’> table </LANDMARK>
<SPATIAL-INDICATOR id=‘1’> on </SPATIAL-INDICATOR>

<SR id=‘1’ trajector=‘1’ landmark=‘1’ spatial-indicator=‘1’ motion-indicator=’NIL’/>
<SR id=‘1’ SRtype id=‘1’ general-type=‘REGION’ specific-type=‘RCC8’ spatial-value=‘EC’ ‘frame-of-reference=‘INTRINSIC’/>
d. She is playing in her room.
<TRAJECTOR id=‘1’> She </TRAJECTOR>
<LANDMARK id=‘1’ path=’ZERO’> her room </LANDMARK>
<SPATIAL-INDICATOR id=‘1’> in </SPATIAL-INDICATOR>
<MOTION-INDICATOR id=‘1’> playing </MOTION-INDICATOR>
<SR id=‘1’ trajector=‘1’ landmark=‘1’ spatial-indicator=‘1’ motion-indicator=‘1’/ >
<SR id=‘1’ SRtype id=‘1’ general-type=‘REGION’ specific-type=‘RCC8’ spatial-value=‘TPP’ frame-of-reference=‘INTRINSIC’/>

3.3 Complex Descriptions

In this section we illustrate how our scheme is able to handle complex spatial descriptions. In [1] three classes of complex description forms are identified to which we point here:
I: Complex locative statements are locative phrases with more than one landmark. 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’> in </SPATIAL-INDICATOR >
<SPATIAL-INDICATOR id=‘2’> on </SPATIAL-INDICATOR >
<SPATIAL-INDICATOR id=‘3’> under </SPATIAL-INDICATOR> <SR id=‘1’ trajector=‘1’ landmark=‘1’ spatial-indicator=‘1’ motion-indicator=‘NIL’ />
<SR id=‘1’ SRtype=‘1’ general-type=‘REGION’ specific-type=‘RCC8’ spatial-value=‘NTPP’ frame-of-reference=‘INTRINSIC’ />
<SR id=‘2’ trajector=‘1’ landmark=‘2’ spatial-indicator=‘2’ motion-indicator=‘NIL’/>
<SR id=‘2’ SRtype=‘1’ general-type=‘REGION’ specific-type=‘RCC8’ spatial-value=‘EC’ frame-of-reference=‘INTRINSIC’/>
<SR id=‘3’ trajector=‘1’ landmark=‘3’ spatial-indicator=‘3’ motion-indicator=‘NIL’/>
<SR id=‘3’ SRtype=‘1’ general-type=‘DIRECTION’ specific-type=‘RELATIVE’ spatial-value=‘BELOW’ frame-of-reference=‘INTRINSIC’ />

II: Path and route descriptions are possibly the most important when dealing with multimodal systems. In this kind of descriptions a *focus shift* can happen. It means that the speaker explains one target referring to some landmarks, but at some point explains another object or landmark, i.e. the focus shift to another entity as trajector. Annotators should recognize this focus shift and annotate the rest of the phrases by the new trajector. 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='3' path='END'> the alley <LANDMARK/>
<TRAJECTOR id='2'> the alley </TRAJECTOR >
<LANDMARK id='4' path='ZERO'> the church </LANDMARK>
<SPATIAL-INDICATOR id='1'> between </SPATIAL-INDICATOR >
<SPATIAL-INDICATOR id='2'> along </SPATIAL-INDICATOR>
<SPATIAL-INDICATOR id='3'> down </SPATIAL-INDICATOR>
<SPATIAL-INDICATOR id='4'> by </SPATIAL-INDICATOR>

<SR id='1' trajector='1' landmark='1' spatial-indicator='1' motion-indicator='NIL'/>
<SR id='1' SRtype='1' general-type='Region' specific-type='RCC8' spatial-value='IN' frame-of-reference='INTRINSIC' motion-indicator='NIL'/>
<SR id='2' trajector='1' landmark='2' spatial-indicator='2' motion-indicator='NIL'/>

<SR id='2' SRtype id='1' general-type='Region' specific-type='RCC8' spatial-value='EC' frame-of-reference='INTRINSIC'/>
<SR id='3' trajector='1' landmark='3' spatial-indicator='3' frame-of-reference='RELATIVE' motion-indicator='NIL'/>
<SR id='3' SRtype id='1' general-type='Direction' specific-type='Relative' spatial-value='Below' frame-of-reference='RELATIVE'/>
<SR id='4' trajector='2' landmark='4' spatial-indicator='4' frame-of-reference='INTRINSIC' motion-indicator='NIL'/>
<SR id='4' SRtype='1' general-type='Region' specific-type='RCC8' spatial-value='DC' frame-of-reference='INTRINSIC'/>

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

In addition to the complex descriptions mentioned in [1], 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’> for </SPATIAL-INDICATOR>
<MOTION-INDICATOR id=‘1’> left </MOTION-INDICATOR>
<SR id=‘1’ trajector=‘1’ landmark=‘1’ spatial-indicator=‘NIL’ motion-indicator=‘1’/>
<SR id=‘1’ SRtype id=‘1’ general-type=‘Direction’ specific-type=‘Relative’ spatial-value=‘NTPP’ frame-of-reference=‘ABSOLUTE’/>
<SR id=‘2’ trajector=‘1’ landmark=‘2’ spatial-indicator=‘1’ motion-indicator=‘1’/
>
<SR id=’2’ SRtype=’1’ general-type=’Direction’ specific-type=’Relative’ spatial-value=’NTPP’ 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 multimodal 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 we suggest that the true relations are added. For 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. The exact meaning can be constrained when additional situational information are provided from external resources.
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’> behind </SPATIAL-INDICATOR>
<SPATIAL-INDICATOR id=‘2’> on the left </SPATIAL-INDICATOR>
<SR id=‘1’ trajector=‘1’ landmark=‘1’ spatial-indicator=‘1’ motion-indicator=‘NIL’/>
<SR id=‘1’ SRtype=‘1’ general-type=‘Direction’ specific-type=‘Relative’ spatial-value=‘BEHIND’ frame-of-reference=‘RELATIVE’ motion-indicator=‘NIL’/>
<SR id=‘2’ trajector=‘1’ landmark=‘NIL’ spatial-indicator=‘2’ frame-of-reference=‘RELATIVE’ motion-indicator=‘NIL’/>
<SR id=‘2’ SRtype=‘1’ general-type=‘Direction’ specific-type=‘Relative’ spatial-value=‘LEFT’ frame-of-reference=‘RELATIVE’/>
<TRAJECTOR id=‘2’>The door </TRAJECTOR>
<SR id=‘3’ trajector=‘2’ landmark=‘NIL’ spatial-indicator=‘2’ frame-of-reference=‘RELATIVE’ motion-indicator=‘NIL’/>
<SR id=‘3’ SRtype=‘1’ general-type=‘Direction’ specific-type=‘Relative’ spatial-value=‘LEFT’ frame-of-reference=‘RELATIVE’/>

In example 12, there are one trajector, three landmarks and three indicators. The landmarks are geographically related, but the 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’> within </SPATIAL-INDICATOR>
<SPATIAL-INDICATOR id=‘2’> from </SPATIAL-INDICATOR>
<SPATIAL-INDICATOR id=‘3’> to </SPATIAL-INDICATOR>
<MOTION-INDICATOR id=‘1’> drives </MOTION-INDICATOR>
<SR id=‘1’ trajector=‘1’ landmark=‘1’ spatial-indicator=‘1’ motion-indicator=‘1’/ >
<SR id=‘1’ SRtype=‘1’ general-type=‘Region’ specific-type=‘RCC8’ spatial-value=‘NTPP’ frame-of-reference=‘ABSOLUTE/>
<SR id=‘2’ trajector=‘1’ landmark=‘2’ spatial-indicator=‘2’ motion-indicator=‘1’/>

<SR id=‘2’ SRtype=‘1’ general-type=‘Region’ specific-type=‘RCC8’ spatial-value=‘NTPP’ frame-of-reference=‘ABSOLUTE’/>
<SR id=‘3’ trajector=‘1’ landmark=‘2’ spatial-indicator=‘3’ motion-indicator=’1’/ >
<SR id=‘3’ SRtype=‘1’ general-type=‘Region’ specific-type=‘RCC8’ spatial-value=‘NTPP’ frame-of-reference=‘ABSOLUTE’/>

Another possibility is having one indicator but with various roles. 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 about the 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 attribute is added to related spatial relations.

4 Data Resources

We performed a broad investigation to find possible data resources to be used as training data by supervised machine learning models for the extraction of spatial information. As, to our knowledge, such data were not publicly available so far, we have built a corpus, based on the aforementioned annotation scheme we refer to it as CLEF which is used as a benchmark for the **SemEval-2012 shared task**. Several machine learning models and experiments have been performed over editions of this corpus [27,22,24,25,43,2]. In addition to the main corpora we annotated very small datasets from different domains and used these in cross domain evaluations in [27]. We also point to a few datasets which were indirectly relevant for the targeted concepts in the proposed scheme. The detailed information is given in the following sections and the relevant statistics are provided in Tables 2 and 3.

4.1 Corpus Collection

The main annotated corpus for the whole scheme is a subset of the **IAPR TC-12 image Benchmark** [13] referred to as **CLEF**. It contains 613 text files that include 1213 sentences in total. The original corpus was available without copyright restrictions. The corpus contains $20,000$ images taken by tourists with textual descriptions in up to three languages (English, German and Spanish). The texts describe objects, and their absolute and relative positions in the image. This makes the corpus a rich resource for spatial information. However the descriptions are not always limited to spatial information. Therefore they are less domain-specific and contain free explanations about the images. An essential property of this corpus is not only that it contains a large enough number of spatial language texts for learning, but also that it has additional (non-linguistic) spatial information, i.e. images, from which a qualitative spatial model can be built that can be related to the textual information. Hence, an additional advantage of this dataset is providing the possibility for further research on combining spatial information from vision and language.

The first column in table 2 shows the detailed statistics about the spatial roles in this data. The average length of the sentences in this data is about 15 words including punctuation marks with a standard deviation of 8. The textual descriptions have been indexed and annotated with the spatial roles of trajector, landmark, and their corresponding spatial indicator. At the starting point two annotators, one of the authors and a non-expert (but with some linguistics background) annotated 325 sentences for the spatial roles and relations. The goal was to realize the disagreement points and prepare a set of instructions in a way to achieve highest-possible agreement. From the first effort an inter-annotator agreement of 0.89 for Cohen's kappa was obtained [7]. This very first version of annotations is used in the experiments in [27]. We refer to it as **SemEval-0** version.

We continued with a third annotator for the remaining 888 sentences. None of the annotators were native English speakers. The third, non-expert, annotator received an explanatory session and a set of instructions and previously annotated examples as a guidance to obtain consistent annotations. This version is referred to as **SemEval-2012** and is used as a benchmark in the workshop with this name.

The roles are assigned to phrases and the head words of the phrases. The verbs and their dependents (i.e. compound verbs and possibly dependent prepositions) are annotated, only when they participate in forming the spatial configurations. This is mostly the case for dynamic spatial relations and for motion verbs. Each sentence with a spatial relation is additionally annotated as DYNAMIC or STATIC, and each spatial relation is annotated with a GUM-Space modality which are used in some experiments in [24] . The spatial relations are annotated with the formal QSR semantics in SpQL layer described in Section 3. For annotating this first corpus we simply used spreadsheet tables. We used a tokenizer and the position of each word in the sentence attached to the words as their index. The annotators used the indexes of the words to fill in the columns for each role. Afterwards, this annotated data was parsed and converted into XML format to be used by the SpRL shared task participants. The whole annotation process was manually done, except for the tokenization and word indexing. Possible mismatches between the annotations and the original sentences (e.g. in terms of incorrect indexes) were corrected semi-automatically, i.e. spotted by a parser and then checked and corrected manually.

The data has a minor revision in its latest edition and is enriched with the QSR annotations. This version is referred to as **SemEval-1**. In SemEval-1 for the directional relations such as *on the left*, the landmark is assumed to be implicit while the word *left* was annotated as landmark in the previous versions. Such expressions, in fact, express *left* of some implicit object depending on the frame of reference. This edition is used in [25].

	CLEF	**GUM** (Maptask)	**Fables**	**DCP**
#Sentences	1213	100	289	250
#Spatial relations	1706	112	121	222
#Trajectors	1593	65	106	199
#Landmarks	1462	69	95	188
#Spatial indicators	1468	112	121	222
#nonSpatial prepositions	695	10	743	587

Table 2. Data statistics on the occurrence of spatial components in different corpora; The CLEF corpus is used for SemEval-2012.

The statistics about formal spatial semantics of the relations are shown in table 3. In the current corpus only 50 examples are annotated with more than one general spatial type. For example, *"next to"* is annotated as the topological relation DC in terms of RCC-8 and as the distance relation CLOSE in terms of a relative distance:

(1) *Two people are sitting next to her.*

```
trajector: people
landmark: her
spatial-indicator: next to
```

Spatial relations 1706

Topological	EQ	DC	EC	PO	PP
1040	6	142	462	15	417

Directional	BELOW	LEFT	RIGHT	BEHIND	FRONT	ABOVE
639	18	159	103	101	185	71

Distal
82

Table 3. Data statistics of the QSR additional annotations on SemEval-2012, referred to as SemEval-1.

```
general-type: region/distance
specific-type: RCC-8 / relative-distance
spatial-value: DC / close
path: none
frame-of-reference: none
```

2D vs. 3D annotations. Although the textual data used is accompanied by images, the qualitative spatial annotation for CLEF was based on the text itself. This was done to focus on information that can actually be extracted from the language itself. Nevertheless, human imagination about a described scene can interfere with the textual description, which has resulted in some variations. As an example, take the following sentence and its annotation:

(2) *Bushes and small trees (are) on the hill.*

```
trajector: bushes
landmark: the hill
spatial-indicator: on
general-type: region
specific-type: RCC-8
spatial-value: EC
path: none
frame-of-reference: none
```

This 3-D projection of the description of a 2-D image is annotated as externally connected. In the 2-D image, however, a partial overlap may also be adequate. In contrast, a 2-D map (with an allocentric perspective) of the described scene would lead to a non-tangential proper part annotation. This example illustrates the necessity of the situational information for capturing the semantics and also the necessity of clarifying the issues such as perspective and dimensions in the annotated data to be able to broaden the usage of such a corpus [49].

Dynamic vs. static annotations. In the CLEF data set 25 of the relations are annotated as DYNAMIC, the others as STATIC. If a dynamic situation is annotated with a (static) RCC-8 relation, the qualitative relation can be regarded as a snapshot of the situation. This is shown in the following example:

(3) *People are crossing the street.*

```
trajector: people
landmark: road
spatial-indicator: crossing
general-type: region / direction
specific-type: RCC-8 / undefined
spatial-value: EC / undefined
path: middle
frame-of-reference: none
```

Hence, the annotations refer to time slices for the (linguistic) explanation of the (static) image. This allows a mapping from dynamic descriptions to (static) RCC-8 relations mainly by including the path feature and the relative situation of the trajector with respect to an imaginary path related to the landmark. Allowing RCC-8 annotations for dynamic descriptions is also supported by the conceptual neighborhood graphs [11]. Every topological change, i.e. movements of regions with respect to each other and their changing relations, can be split into a sequence of adjacent RCC-8 relations according to the neighborhood graph [19]. The annotated RCC-8 relation thus reflects one relation out of this sequence, i.e. one moment in time of the topological change (also see [37]). However, we may not predict if the annotations refer to a time slice that reflects the start, intermediate, or end point of the path or the motion process. For instance, it is shown that linguistic expressions seem to focus primarily on the end point of the motion [40].

4.2 Other Linguistic Resources

In this part we briefly point to other relevant resources for spatial information extraction from language, which we used in our research.

- **TPP dataset** Since the spatial indicators are mostly prepositions, the preposition sense disambiguation is an important relevant task to our problem. Fortunately, for this specific task, there is standard test and training data provided by the SemEval-2007 challenge [31]. It contains 34 separate XML files, one for each preposition, totaling over 25,000 instances with 16,557 training and 8,096 test example sentences; each sentence contains one example of the respective preposition.
- **GUM-evaluation (Maptask) dataset** Another relevant small corpus is the general upper model (GUM) evaluation data [3], comprising a subset of a well-known Maptask corpus for spatial language. It has been used to validate the expressivity of spatial annotations in the GUM ontology. Currently, the dataset contains more than 300 English and 300 German examples. We used 100 English sample sentences in the GUM (Maptask) corpus in some machine learning models described in [27]. The following example shows the GUM-annotation for one sentence represented with GUMs predicate formalism for representation:

(4) *The destination is beneath the start.*

```
SpatialLocating(locatum:destination,
process:being,placement: GL1
(relatum:start,
hasSpatialModality:UnderProjectionExternal)).
```

Here, *relatum* and *locatum* are alternative terms for landmark and trajector. *Spatial modality* is the spatial relation mentioned in the specific spatial ontology. Although complete phrases are annotated in this dataset, we only use a phrase's headword with trajector (**tr**) and landmark (**lm**) labels and their spatial indicator (**sp**). Using this small corpus to evaluate our approach for a very domain-specific corpus, including only instructions and guidance for finding the way on a map, is beneficial.

– **DCP dataset** The dataset contains a random selection from the website of *The Degree Confluence Project*.[3] This project seeks to map all possible latitude-longitude intersections on earth, and people who visit these intersections provide written narratives of the visit. The main textual parts of randomly selected pages are manually copied, and up to 250 sentences are annotated. Approximately 30% of the prepositions are spatial. This percentage represents the proportion of spatial clauses in the text. The webpages of this dataset are similar to travelers' weblogs but include more precise geographical information. The richness of this data enables broader applicability for future applications. Compared to CLEF, this dataset includes less spatial information, and the type of text is narrative rather than descriptive. It also contains more free (unrestricted) text. Moreover, the spatio-temporal information contained in this data has recently been used to extract discourse relations [16].
– **Fables dataset** This dataset contains 59 randomly selected fable stories[4], which have been used for data-driven story generation [35]. The dataset contains a wide scope of vocabulary and only 15% of the prepositions have a spatial meaning, making it a difficult corpus for automatic annotation. We annotated 289 sentences of this corpus.

There is another small dataset about *Room descriptions* prepared by Tenbrick et al. in [47]. This data is not publicly available. We had a limited access to 124 sentences of this corpus that contains directional and topological descriptions for an automatic wheelchair about the objects in a room. The full dataset which contains pictures of the room can help preparing multimodal analyses.

[3] http://confluence.org/
[4] http://homepages.inf.ed.ac.uk/s0233364/McIntyreLapata09/

5 Related Work

In recent cognitive and linguistic research on spatial information and natural language, several annotation schemes have been proposed such as ACE[5], GUM[6], GML[7], KML[8], TRML[9] which are described and compared to the SpatialML scheme in [32]. The most systematic pioneer work on spatial annotation is the SpatialML scheme which focuses on geographical information [33]. SpatialML 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 are also connected to RCC8 relations. However, SpatialML considers static spatial relations and focuses on geographical domains. The corpus which is provided along with the SpatialML scheme contains rich annotations for toponymy but does not provide many examples about spatial relations and especially not about relations between arbitrary objects.

GUM also aims at organizing spatial concepts that appear in natural language from an ontological point of view. The formulated concepts are very expressive, but the ontology is large and more fine-grained than what could be effectively learnable from a rather small corpus. An XML scheme based on SpatialML and GUM was proposed in [46], targeting spatial relations in the Chinese language. It also deals with geographical information and defines two main tags, that relate to geographical entity and spatial expression. In [37], a spatio-temporal markup language for the annotation of motion predicates in text informed by a lexical semantic classification of motion verbs, is proposed. The noticeable point is that the proposed scheme seems suitable for tagging dynamic spatial relations, based on motions in space and time. However, the focus is on motion verbs and their spatial effects and not on spatial language in general. There is another spatial annotation scheme proposed in [37] in which the pivot of the spatial information is the spatial verb.

The most recent and active research work regards the ISO-Space scheme [38] which is based on the last mentioned scheme and SpatialML. The ISO-Space considers detailed and fine-grained spatial and linguistic elements, particularly motion verb frames. The detailed semantic granularity considered there makes the preparation of the data for machine learning more expensive and there is no available data for machine learning annotated according to that scheme yet. A thorough investigation of motion in spatial language, its formal representation and computational practices is given in [34]. Our proposed scheme is closely related to the SpatialML scheme, but is more domain independent considering more universal spatial primitives and cognitive aspects. It is relevant to the ISO-Space scheme but the pivot of the relation is not necessarily the verb, and

[5] Automatic content extraction
[6] General upper model
[7] Geography markup language
[8] Keyhole markup language
[9] Toponym resolution markup language

a general notion of spatial indicator is used as the pivot of each spatial configuration.

Spatial information is directly related to the part of language that can be visualized. Thus, the extraction of spatial information is useful for multimodal environments. One advantage of our proposed scheme is that it considers this dimension. Because it abstracts the spatial elements that could be aligned with the objects in images/videos, it can be used for annotation of audio-visual descriptions as shown in [6]. Our scheme is also useful in other multimodal environments where, for example, natural language instructions are given to a robot for finding the way or objects.

There are a few sparse efforts towards creating annotated data sets for extraction of some limited elements of our scheme. For example in [30] 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. It has been applied in a very limited machine learning setting; only a binary classifier was used so far for the extraction of the trajector. In [46] 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 be mentioned that from the linguistic point of view, FrameNet frames [10] 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 the frame-based semantic annotation. There are various semantic frames which are related to spatial roles and semantics. Frames like LOCATIVE RELATION, SELFMOTION, 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 [28]. With this view a corpus is annotated (in German) for walking directions [45]. 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. However, the available resources and corpora are very limited for broad machine learning research on this area, hence we provide an annotated dataset [10] [11] according to the proposed scheme which we described in this chapter and which has been used as the first benchmark for spatial information extraction from natural language in SemEval2012.

Apart from the related research prior to this work there is follow up research that has used the annotation scheme proposed here. There are several machine

[10] http://www.cs.york.ac.uk/semeval-2012/task3/
[11] http://www.cs.kuleuven.be/groups/liir/sprl/sprl.php

learning practices using a part of the annotated data, mostly related to recognition of spatial relations, that is SpRL layer [27,22]. The overall experimental results show that machine learning models can learn from this annotated data to extract the spatial roles and relations, outperforming standard semantic role labelers when we look specifically at spatial semantics in the language. The annotations of the SpRL layer, in addition to the general types of the formal semantics of the relations (region, direction and distance), were the subject of the SemEval-2012 shared task [21,43] on the CLEF corpus. The annotated data was extended for the SemEval-2013 shared task [20,2] with 1789 additional sentences from the DCP corpus. The annotations also were extended to distinguish between *path* in the dynamic relations compared to basic *landmarks* in the static relations of the CLEF corpus. Moreover, the prior practices on SpRL were on the word-level and concerned labeling the headwords of the phrases while this was extended to the phrase boundaries predictions in SemEval-2013.

Machine learning efforts have been performed on the SpQL layer too and show promising results for recognition of the formal semantics of the spatial relations in terms of qualitative spatial representation and reasoning models [23,24,25]. The same elements as in the proposed scheme have been used for recognizing discourse relations in [17]; the experimental results show the advantage of using spatial information such as trajectors and landmarks in discourse relation extraction. The annotation scheme proposed in this chapter has been exploited for annotating audio-visual scene descriptions in [6]. The spatial relation which is composed of the roles of trajector, landmark and spatial indicator is augmented with the descriptive modifiers in the sentences and the same structure has been used for extraction of spatial information from place descriptions [18]. A complementary work uses the basics in the proposed scheme and extracts the spatial relations and their attributes in terms of formal relations and makes depictions from the textual descriptions [50].

6 Conclusion

The first contribution of this chapter is proposing a spatial annotation scheme on the basis of the existing research. The advantages of the proposed scheme compared to other existing schemes are: a) it is based on the concepts of two layers of cognitive spatial semantics and formal spatial representation models; b) it is domain-independent and useful for real world applications and it is rather flexible to be extended in its two layers to cover all aspects of spatial information; c) it is easily applicable for annotating spatial concepts in image data and multimodal settings; d) it supports static as well as dynamic spatial relations; e) by using multiple formal semantic assignments, it bridges the gap between the natural language spatial semantics and formal spatial representation models. For each of the cognitive and formal semantic aspects, we exploit the most commonly accepted concepts and their formalizations to establish an agreeable setting for spatial information extraction. Extraction of the spatial information accruing to this scheme facilitates automatic spatial reasoning based on linguistic information.

The second contribution of this chapter regards corpora preparation according to the proposed scheme and assessing the available resources for spatial information extraction from natural language based on machine learning techniques. The noticeable points about the selected data are: a) the data contains free text about various topics, including spatial and non spatial information; b) the textual descriptions in the corpus are related to images implying that they contain rich spatial information; c) they create opportunities to learn in multimodal contexts if texts are accompanied by images carrying the same information as in the text, where language elements can be grounded in the images.

A part of the annotated data has been used as a benchmark in the SemEval-2012 shared task on *spatial role labeling* [21] and an extension of it is used in SemEval-2013 [20]. Both versions are publicly available for follow-up research in this field [12]. Providing such a benchmark is an important step towards persuasion to work on, and thus progress in, spatial information extraction as a formal computational linguistic task. In addition, it generates and understanding of the practical side when working on enriching both the corpora and the proposed task. These things are hard to achieve without working on practical systems as well.

References

1. M. Barclay and A. Galton. A scene corpus for training and testing spatial communications. In *Proceedings of the AISB Convention (Communication, Interaction, and Social Intelligence)*, 2008.
2. E. Bastianelli, D. Croce, R. Basili, and D. Nardi. UNITOR-HMM-TK: Structured kernel-based learning for spatial role labeling. In *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, pages 573–579, Atlanta, Georgia, USA, June 2013. Association for Computational Linguistics.
3. J. Bateman, T. Tenbrink, and S. Farrar. The role of conceptual and linguistic ontologies in discourse. *Discourse Processes*, 44(3):175–213, 2007.
4. J. A. Bateman. Language and space: A two-level semantic approach based on principles of ontological engineering. *International Journal of Speech Technology*, 13(1):29–48, 2010.
5. J. A. Bateman, J. Hois, R. Ross, and T. Tenbrink. A linguistic ontology of space for natural language processing. *Artificial Intelligence*, 174(14):1027–1071, 2010.
6. T. Butko, C. Nadeu, and A. Moreno. A multilingual corpus for rich audio-visual scene description in a meeting-room environment. In *ICMI Workshop on Multimodal Corpora for Machine Learning: Taking Stock and Roadmapping the Future*, pages 1–6. ACM Press, 2011.
7. J. Carletta. Assessing agreement on classification tasks: the Kappa statistic. *Computational Linguistics*, 22(2):249–254, 1996.
8. L. A. Carlson and S. R. Van Deman. The space in spatial language. *Journal of Memory and Language*, 51:418–436, 2004.

[12] http://www.cs.kuleuven.be/groups/liir/sprl/sprl.php

9. M. Egenhofer. Reasoning about binary topological relations. In Oliver Gunther and Hans Jorg Schek, editors, *Advances in Spatial Databases*, volume 525 of *Lecture Notes in Computer Science*, pages 141–160. Springer Berlin Heidelberg, 1991.
10. T. Fontenelle. FrameNet and frame semantics: a special issue. *International Journal of Lexicography*, 16(3), 2003.
11. C. Freksa. Qualitative spatial reasoning. In D. M. Mark and A. U. Frank, editors, *Cognitive and Linguistic Aspects of Geographic Space*, pages 361–372. Kluwer Academic Publishers, 1991.
12. A. Galton. Spatial and temporal knowledge representation. *Journal of Earth Science Informatics*, 2(3):169–187, 2009.
13. M. Grubinger, P. Clough, H. Müller, and T. Deselaers. The IAPR benchmark: a new evaluation resource for visual information systems. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*, pages 13–23, 2006.
14. 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: Proceedings of the Fifth International Conference (FOIS)*, pages 266–279, 2008.
15. J. Hois and O. Kutz. Natural language meets spatial calculi. In C. Freksa, N. S. Newcombe, P. Gärdenfors, and S. Wölfl, editors, *Spatial Cognition VI. Learning, Reasoning, and Talking about Space*, volume 5248 of *LNCS*, pages 266–282. Springer, 2008.
16. B. Howald and E. Katz. On the explicit and implicit spatiotemporal architecture of narratives of personal experience. In Max Egenhofer, Nicholas Giudice, Reinhard Moratz, and Michael Worboys, editors, *Spatial Information Theory*, volume 6899 of *Lecture Notes in Computer Science*, pages 434–454. Springer Berlin / Heidelberg, 2011.
17. B. S. Howald and E. G. Katz. The exploitation of spatial information in narrative discourse. In *Proceedings of the Ninth International Conference on Computational Semantics*, IWCS '11, pages 175–184, Stroudsburg, PA, USA, 2011. Association for Computational Linguistics.
18. A. Khan, M. Vasardani, and S. Winter. Extracting spatial information from place descriptions. In *Proceedings of 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2013.
19. A. Klippel and R. Li. The endpoint hypothesis: a topological-cognitive assessment of geographic scale movement patterns. In *Spatial Information Theory, COSIT'09*, pages 177–194, 2009.
20. O. Kolomiyets, P. Kordjamshidi, M.F. Moens, and S. Bethard. Semeval-2013 task 3: Spatial role labeling. In *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, pages 255–262, Atlanta, Georgia, USA, June 2013. Association for Computational Linguistics.
21. P. Kordjamshidi, S. Bethard, and M. F. Moens. SemEval-2012 task 3: Spatial role labeling. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval)*, volume 2, pages 365–373. ACL, 2012.
22. P. Kordjamshidi, P. Frasconi, M. van Otterlo, M. F. Moens, and L. De Raedt. Relational learning for spatial relation extraction from natural language. In *The Proceedings of ILP 2011, Lecture Notes in Artificial Intelligence*, volume 7207, pages 204–220. Springer, 2012.

23. P. Kordjamshidi, J. Hois, M. van Otterlo, and M. F. Moens. Machine learning for interpretation of spatial natural language in terms of QSR. In *The poster presentation of 10th International Conference on Spatial Information Theory COSIT'11, extended abstract*, pages 1–5, 2011.
24. P. Kordjamshidi, J. Hois, M. van Otterlo, and M. F. Moens. Learning to interpret spatial natural language in terms of qualitative spatial relations. In T. Tenbrink, J. Wiener, and C. Claramunt, editors, *Representing Space in Cognition: Interrelations of Behavior, Language, and Formal Models. Series Explorations in Language and Space, Oxford University Press*, pages 115–146. Oxford University Press, 2013.
25. P. Kordjamshidi and M.F. Moens. Structured learning for mapping natural language to spatial ontologies. *Submitted*, 2013.
26. P. Kordjamshidi, M. van Otterlo, and M. F. Moens. From language towards formal spatial calculi. In Robert J. Ross, Joana Hois, and John Kelleher, editors, *Workshop on Computational Models of Spatial Language Interpretation (CoSLI'10, at Spatial Cognition)*, pages 17–24, 2010.
27. P. Kordjamshidi, M. van Otterlo, and M. F. Moens. Spatial role labeling: towards extraction of spatial relations from natural language. *ACM - Transactions on Speech and Language Processing*, 8:1–36, 2011.
28. K. Kuroda, M. Utiyama, and H. Isahara. Getting deeper semantics than Berkeley FrameNet with MSFA. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC)*, 2006.
29. S. C. Levinson. *Space in Language and Cognition: Explorations in Cognitive Diversity*. Cambridge University Press, 2003.
30. H. Li, T. Zhao, S. Li, and Y. Han. The extraction of spatial relationships from text based on hybrid method. *International Conference on Information Acquisition*, pages 284–289, 2006.
31. K. Litkowski and O. Hargraves. SemEval-2007 Task 06: Word-sense disambiguation of prepositions. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, pages 24–29. ACL, 2007.
32. I. Mani. SpatialML: annotation scheme for marking spatial expression in natural language. Technical Report Version 3.0, The MITRE Corporation, 2009.
33. I. Mani, J. Hitzeman, J. Richer, D. Harris, R. Quimby, and B. Wellner. SpatialML: annotation scheme, corpora, and tools. In N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, and D. Tapias, editors, *Proceedings of the Sixth International Language Resources and Evaluation (LREC)*. European Language Resources Association (ELRA), 2008.
34. I. Mani and J. Pustejovsky. *Interpreting Motion: Grounded Representations for Spatial Language*. Explorations in language and space. Oxford University Press, 2012.
35. N. McIntyre and M. Lapata. Learning to tell tales: a data-driven approach to story generation. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pages 217–225. ACL, 2009.
36. R. J. Mooney. Learning to connect language and perception. In Dieter Fox and Carla P. Gomes, editors, *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence*, pages 1598–1601, 2008.
37. J. Pustejovsky and J. L. Moszkowicz. Integrating motion predicate classes with spatial and temporal annotations. In Donia Scott and Hans Uszkoreit, editors, *COLING 2008: Companion volume D, Posters and Demonstrations*, pages 95–98, 2008.

38. J. Pustejovsky and J. L. Moszkowicz. The role of model testing in standards development: The case of ISO-space. In *Proceedings of LREC'12*, pages 3060–3063. European Language Resources Association (ELRA), 2012.
39. D. A. Randell, Z. Cui, and A.G. Cohn. A spatial logic based on regions and connection. In *Proceedings of the 3rd International Conference on the Principles of Knowledge Representation and Reasoning, KR'92*, pages 165–176, 1992.
40. T. Regier and M. Zheng. Attention to endpoints: a cross-linguistic constraint on spatial meaning. *Cognitive Science*, 31(4):705–719, 2007.
41. J. Renz and B. Nebel. Qualitative spatial reasoning using constraint calculi. In M. Aiello, I. Pratt-Hartmann, and J. van Benthem, editors, *Handbook of Spatial Logics*, pages 161–215. Springer, 2007.
42. J. Renz, R. Rauh, and M. Knauff. Towards cognitive adequacy of topological spatial relations. In *Spatial Cognition II*, pages 184–197, 2000.
43. K. Roberts and S.M. Harabagiu. UTD-SpRL: A joint approach to spatial role labeling. In **SEM 2012: The First Joint Conference on Lexical and Computational Semantics, Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval'12)*, pages 419–424, 2012.
44. R. Ross, H. Shi, T. Vierhuff, B. Krieg-Brückner, and J. Bateman. Towards dialogue based shared control of navigating robots. In C. Freksa, M. Knauff, B. Krieg-Brückner, B. Nebel, and T. Barkowsky, editors, *Proceedings of Spatial Cognition IV: Reasoning, Action, Interaction*, pages 478–499. Springer, Berlin/ Heidelberg, 2005.
45. S. Schuldes, M. Roth, A. Frank, and M. Strube. Creating an annotated corpus for generating walking directions. In *Proceeding of the ACL-IJCNLP 2009 Workshop: Language Generation and Summarization*, pages 72–76, 2009.
46. 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*, volume 3, pages 418 – 421, 2009.
47. H. Shi and T. Tenbrink. Telling Rolland where to go: HRI dialogues on route navigation. In K. Coventry, T. Tenbrink, and J. Bateman, editors, *Spatial Language and Dialogue*, pages 177–189. Oxford Univ. Press, 2009.
48. O. Stock, editor. *Spatial and Temporal Reasoning.* Kluwer Academic Publishers, 1997.
49. T. Tenbrink and W. Kuhn. A model of spatial reference frames in language. In Max Egenhofer, Nicholas Giudice, Reinhard Moratz, and Mike Worboys, editors, *Conference on Spatial Information Theory (COSIT'11)*, pages 371–390. Springer, 2011.
50. M. Vasardani, S. Timpf, S. Winter, and M. Tomko. From descriptions to depictions: A conceptual framework. In A. Galton Z.Wood T. Tenbrink, J.G. Stell, editor, *Conference on Spatial Information Theory. Proceedings of COSIT 2013*, volume 8116, pages 299–319. Heidelberg: Springer Verlag, 2013.
51. J. Wallgrün, L. Frommberger, D. Wolter, F. Dylla, and C. Freksa. Qualitative spatial representation and reasoning in the SparQ-Toolbox. In Thomas Barkowsky, Markus Knauff, Gérard Ligozat, and Daniel R. Montello, editors, *Spatial Cognition V Reasoning, Action, Interaction*, volume 4387, chapter 3, pages 39–58. Springer Berlin/Heidelberg, 2007.
52. J. Zlatev. Holistic spatial semantics of Thai. *Cognitive Linguistics and Non-Indo-European Languages*, pages 305–336, 2003.
53. J. Zlatev. Spatial semantics. In D. Geeraerts and H. Cuyckens, editors, *The Oxford Handbook of Cognitive Linguistics*, pages 318–350. Oxford Univ. Press, 2007.