

Modeling Spatial Knowledge for Generating Verbal and Visual Route Directions

Stephanie Schuldes¹, Katarina Boland², Michael Roth³, Michael Strube¹, Susanne Krömker², and Anette Frank³

¹ NLP Group, Heidelberg Institute for Theoretical Studies gGmbH, Germany

<http://www.h-its.org>

² NGG Group, Interdisciplinary Center for Scientific Computing, Heidelberg University, Germany

<http://www.iwr.uni-heidelberg.de>

³ Department of Computational Linguistics, Heidelberg University, Germany

<http://www.cl.uni-heidelberg.de>

Abstract. Challenges in the development of state-of-the-art navigation tools involve the synchronous generation of user-tailored natural language directions and accompanying graphical representations. Of particular interest is the combination of indoor and outdoor scenarios. We describe an architecture for simultaneously generating route directions and suitable visualizations of the navigational context for walking directions. We present a representation model for 3D environments, the *spatial model*, and a route representation format based on OpenGIS standards, aiming at an elegant and efficient way to support users with and without knowledge about the environment.

Keywords: multimodal generation, route directions, spatial and conceptual ontology, 3D visualization

1 Introduction

The aim of our work is to synchronously generate natural language route directions and corresponding 3D scenes. This requires knowledge not only about correspondences between visual information and its linguistic expression, but also regarding spatial information about the navigational context which has to be made accessible to the language and 3D scene generation algorithms. In a spatially situated context, the representation of geographic space serves as an anchor to link the linguistic and visual modalities of the scene.

We capture this in an architecture that relates the representation of a 3D environment, our *spatial model*, with a concept ontology that associates linguistic expressions with graphical representations. The aligned knowledge bases serve as a backbone for simultaneous generation of linguistic and 3D instructions for indoor and outdoor navigation. We further propose the use of several constraints to control the interaction of both modalities in an integrated constraint-based generation architecture.

This paper first describes our *spatial model*, an extension of the *Schematic Geometry* by Rüetschi [22], and its linking to a large concept ontology, to yield an aligned knowledge base for generating route directions. Since manual creation of spatial models becomes infeasible in larger contexts, we describe our method for automatically generating a spatial model from existing architectural and geographic information sources (Section 3), and show how our concept ontology is derived (Section 4). We introduce a route representation format based on OpenGIS standards (Section 5), and motivate how the aforementioned components can be used to generate route descriptions and visualizations tailored for users with and without knowledge about the environment (Section 6).

Our system builds upon *aligned knowledge bases* which combine the hierarchical representation of the spatial model with a (structurally distinct) concept ontology (cf. Figure 1). Each concept has associated realizations and prototypes for both the linguistic and the graphical modality. For example, the concept for an interior door can be referred to as “door” in the instructions and rendered in the 3D visualization using the prototypical 3D object associated with that concept node. These concepts are aligned with the nodes of the spatial model.

The spatial model represents a *hierarchical* organization of space. This will enable the generation of verbal instructions and accompanying visualizations on *different levels of granularity*. We consider this as an important prerequisite for generating cognitively ergonomic route directions, which should be as fine-grained as necessary, yet as coarse as possible [11]. *Landmarks* are another important component for producing natural route directions. Every object or feature that stands out

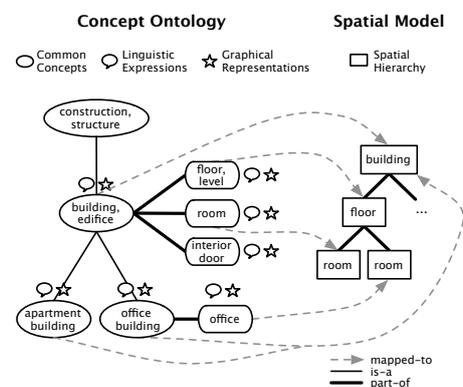


Fig. 1. Aligned knowledge bases

from its surroundings may be used as a landmark. Accordingly, any node of the spatial model must be considered a potential landmark. Thus, in a route instruction based on the spatial model, we use the links to the concept ontology to access object properties of specific landmarks. The graphical prototypes stored with these concepts allow us to generate visualizations. We will apply *salience measures and constraints* to determine which properties will be used in generating referring expressions and the corresponding visualizations.

2 Related Work

Spatial representations. A new line of approaches distinguishes the inherent semantics of linguistic expressions and their spatial interpretation in a given situation [1]. This is necessary because the interpretation involves a non-trivial mapping between ontologies and concepts from spatial language. Approaches following [6] propose to link underspecified linguistic and rich non-linguistic levels of representation [8, inter alia]. In contrast, we follow Bateman [1] who argues that linguistic and non-linguistic information needs to be kept in separate ontological modules and needs to relate individual concepts as required by the context.

Hierarchical spatial models extending to indoor navigation. Rüetschi [22] distinguishes *network space*, which depends on infrastructure (e.g. the street network), and *scene space*, which is determined by open spaces and visibility. As our scenario concentrates on walking directions, a model based on scene space seems to be most appropriate. Recent work [20] extends this approach to model indoor spaces, using three dimensions: structural, functional, and organizational. Starting from a base structural representation derived from architectural floor plans, they build three dimensions. A model consisting of hierarchical graphs with typed edges and nodes is proposed in [17]. Depending on the level of granularity currently examined, the graph and its elements represent different parts of the modeled space (e.g. room, floor, building). Large areas of open space are divided into disjoint convex cells.

Granularity of route directions. *Destination descriptions*, a particular kind of route directions for people with some knowledge of the environment, are used in [25]. These directions guide people to the destination using a hierarchically ordered sequence of references to prominent spatial features. The first reference is most distinguishing, setting the scene by referring to the most prominent referents in the wider vicinity of the destination. It is followed by detailed references to less prominent features closer to the destination. This is supported by a hierarchical model based on the experience of space and measures of prominence for all types of elements of a city. This is in contrast with previous approaches that model hierarchies based on the street network [2, inter alia]. We will discuss this aspect in more detail in Section 3.

Route representations and landmarks. Empirical studies demonstrated that landmarks facilitate navigation and that people prefer to use landmarks over street names when giving route directions themselves [24]. A formal representation language for route elements called *wayfinding choremes* is proposed and extended with context-specific landmarks [10, 12, 19]. Given the importance of landmarks for routes and route directions, the development of an adequate classification scheme for landmarks has since become a separate research task [23, 3]. An extension of the OpenGIS Location Services is proposed in [5]. It includes cognitively relevant information, such as landmarks, in a route representation. We will discuss this work in more detail in Section 5.

Multimodal route directions. The system *Vitra Guide* [18] generates multimodal route directions for vehicle navigation. Route representations are derived from a 3D geometric model of the domain. *Vitra Guide* accepts natural language queries and outputs verbal route directions along with a 2D map or a 3D view of the environment.

A measure for the applicability of spatial relations is used to generate appropriate locating expressions for spatial configurations. Other work [4] develops an algorithm for generating route directions and proposes the combination of moving-dot maps and spoken route instructions as a use case. The idea of combining verbal directions with scenes from a virtual 3D environment has led to a new framework for evaluating NLG systems, the *Challenge on Generating Instructions in Virtual Environments (GIVE)* [13], a regular shared task in the NLG community.

3 Spatial Model

The *spatial model* is a graph representing a 3D environment. Nodes represent spatial objects with their respective properties. Edges represent relationships between the objects. The hierarchical structure of the model allows the aggregation of nodes, thereby enabling the representation of spatial objects at various levels of granularity.

Our spatial model builds on [22, 20], which use image schema instances and relations between them to represent spatial configurations. *Image schemata* [7] are recurrent patterns that structure our empirical knowledge of the world. They incorporate meaning and function and pose constraints on inferences that can be drawn.

Schema	Components	Relations	Class
AGGREGATE	none	part-of (element-of)	collecting
CONTAINER	inside, outside, boundary	part-of (contains)	collecting
REGION	a bounded surface	part-of (on/off)	collecting
GATEWAY	none	connect	linking
ULINK	none	connect	linking
OBJECT	the object	(not considered by [22])	other
PASSAGE	inside, outside, boundary	connect/part-of (contains)	linking/collecting

Table 1. Cognitive spatial schemata from [22]; PASSAGE element from [20].

Table 1 displays the schemata used by [22] and [20]. CONTAINER instances, e.g. rooms, are areas surrounded by physical barriers. AGGREGATE instances are a collection of functionally or spatially related elements, e.g. a building is a collection of floors, rooms and all objects inside the building. A REGION is an area that is perceived as a unit although it is not surrounded by barriers, such as a square. CONTAINER, AGGREGATES and REGIONS belong to the class *collecting*, i.e., they can contain or can be composed of other objects. Areas are connected by links, which can be GATEWAYS or ULINKS. GATEWAYS are perceived when travelled through, ULINKS are not. Doors and escalators are GATEWAYS, the boundaries between REGION elements are ULINKS. PASSAGES are CONTAINERS that primarily serve as links such as corridors. PASSAGES are both collecting and linking.

Our spatial model extends previous work by fixed and moveable objects, additional edge types, attributes relevant to natural language generation and visualization and outdoor concepts. Outdoor/City space is represented by aggregation levels above the building level such as blocks, districts and cities. An additional image schema, PATH, is used to represent streets. An example for the indoor scenario is shown in Figure 2.

Following [22], image schemata are assigned to each node in the model. objects are represented by the OBJECT schema. Both fixed objects (*components*) and moveable objects (*furniture*) may be used as reference objects in directions and are thus essential parts of our model. Along with an image schema, each node has an identification number, a name (e.g. a room number or the proper name of a location), absolute coordinates, an orientation (rotation around the specified axes, e.g. CONT_room.5 in Figure 2 is rotated 1.57 rad around the x-axis), a position (fixed or moveable, needed for the computation of salience) and a type. The type attribute contains a link to a corresponding concept in the concept ontology. This way, we align the concept ontology with our spatial model as shown in Figure 1. Each object represented in the spatial model is supplemented with visual and linguistic data while knowledge about objects and spatial configurations remain distinct.

We distinguish the following types of edges. Aggregate edges represent mereological relationships. Physical edges denote transitions between spaces, i.e., they hold between links and regions. Also, they denote mereological relationships as links are always part of the spaces they link. ULink edges represent transitions between spaces that are not perceived by the traveller. Physical and ULink edges correspond to the network structure of the environment and are called *passable* edges. Passable edges are propagated to higher level elements, creating inferred edges indicated by dashed lines in Figure 2. Together with aggregate edges, they are called *abstract* edges. They are particularly useful for determining the granularity of route directions (cf. Section 6).

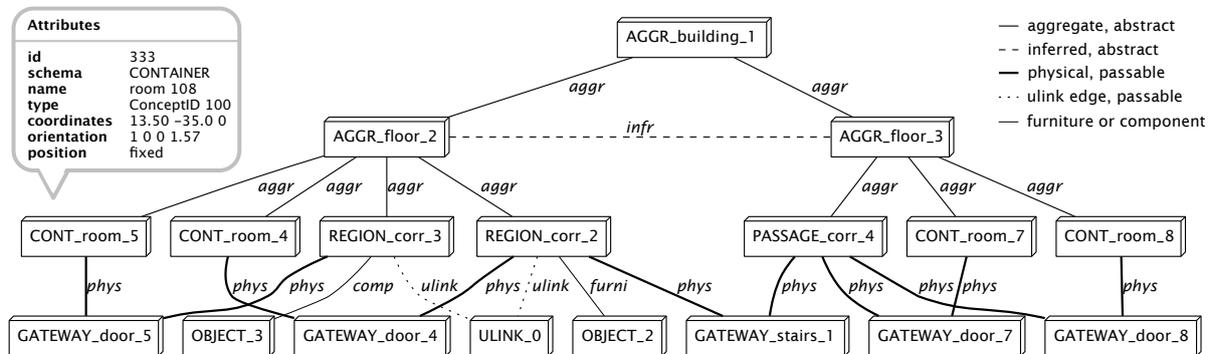


Fig. 2. Simplified example of a spatial model representing a two-floor building. Edge types: aggregate (*aggr*), physical (*phys*), ULink (*ulink*), inferred from physical and ULink (*infr*), furniture (*furni*), component (*comp*).

Construction of spatial models. Our indoor model is automatically derived from CAD architectural drawings. These plans are organized in layers grouping components by type, i.e., all symbols denoting doors are stored on a door layer, all strings are stored on a text layer, room boundaries are stored on a room layer etc. Components on different layers are not associated with each other, i.e., a room is not connected with its description, its doors or its walls. However, humans draw inferences from the arrangement of elements in graphics [26, inter alia]. For example, strings placed inside the boundaries of a room are interpreted as a description of the room, walls and windows on the boundary of a room as constituting its barriers, doors between two rooms as connecting these rooms. Therefore, relations between components can be reconstructed from their coordinates and their assignment to a layer. Fixed objects, such as walls and windows, are directly extracted from architectural floor plans and inserted into the spatial model. Furniture usually is not included in floor plans. We therefore investigate the use of interior design plans to enrich our representation with this kind of objects.

For the outdoor extension of our spatial model, we use a spatial representation of outdoor environments provided by collaborating researchers from Geoinformatics. In the outdoor data, objects are located using absolute coordinates, latitude and longitude. Objects in architectural floor plans are located in a coordinate system whose origin typically lies in the upper left of the drawing. In order to combine indoor and outdoor environments in our spatial model, we use absolute coordinates to store outdoor objects while buildings define local coordinate systems in which fixed and moveable objects are located.

Locating objects in space. Objects are located relative to other objects with known locations rather than in an absolute way, e.g. by using coordinates [27, 26]. The representation of the location of an object therefore requires the object itself, one or more reference objects and their relationships [14]. Spatial relationships can be expressed using one of three different frames of reference: absolute, relative and intrinsic frame [16]. In the absolute frame, objects are located in terms of coordinates based on cardinal directions or fixed bearings centered on the reference object (e.g. *north of the chair*). In the relative and intrinsic frames, objects are located using a local coordinate system. In the relative frame, the coordinate system is originated in the viewer’s position and his body axes serve as coordinate axes (e.g. *in front of the chair*). In the intrinsic frame, the origin of the coordinate system is located at the position of the reference object and the coordinate axes are aligned with the intrinsic axes of the reference object (e.g. *at the front of the chair*).

Preference for reference frames depends on scale. The absolute frame is used for locating objects on a large scale, for example countries or cities, but it is uncommon to use it for locating objects inside of a room. The viewer’s position, as needed for the relative frame, is supplied by an underlying route graph (see section 5). The fixed bearings needed for the absolute frame can also easily be defined. However, determining the intrinsic axes of an object as needed for the intrinsic frame is not trivial. Objects may or may not have inherent axes that distinguish front and back, top and bottom, or different sides. The segmentation of objects with respect to their axes depends not only on shape, but also on the function of an object or its parts and other factors [9].

Since our model comprises a limited set of objects, we will manually assign axial structures to the objects contained in our 3D prototype ontology. The actual orientation of an object in the environment is stored in the orientation attribute in the spatial model. The objects’ coordinates and orientations and the viewer’s position and orientation can then be used to perform reasoning on directions to generate comprehensible locating expressions using qualitative direction concepts like “left”, “under” or “in front of”. In addition, image schemata are used to determine the applicability of certain locative expressions. For example, the preposition *in* may be used with a CONTAINER instance as reference object but not with a REGION.

4 Concept Ontology

In addition to the knowledge about the structure and configuration of the environment which is represented in the spatial model, we need information about how to refer to parts of this model. This information is of two kinds: On the language side, lexemes and object attributes need to be selected in order to generate referring expressions. On the visualization side, prototypic 3D representations need to be stored and adapted according to the object attributes. Both realization modalities use the same spatial objects. Thus, we build a concept ontology in which each concept has associated information for both modalities, and align it with the nodes in the spatial model via their type attributes.

Construction of the concept ontology. Since we work with a corpus of route directions in German, the base for our concept ontology is GermaNet [15]. We map nouns extracted from the corpus to word senses in GermaNet. Table 2 displays information about coverage and ambiguity of terms in GermaNet. We disambiguate ambiguous terms using the image schemata of nodes in the spatial model and the toplevel concepts as constraints. E.g., for

```

--Sense--> Bank (1)
           Bank (1) | Sitzbank (1)
--Hyper--> Sitzmoebel (1)
--Hyper--> Moebel (1) | Moebelstueck (1) | Einrichtungsgegenstand (1)
--Hyper--> Artefakt (1) | Werk (1)

--Sense--> Bank (1)
           Bank (1) | Geldinstitut (1) | Kasse (1) | Geldhaus (1)
--Hyper--> wirtschaftliche Institution (1) | wirtschaftliche Einrichtung (1)
--Hyper--> Institution (1) | Einrichtung (1)
--Hyper--> Gebaeude (1) | Bau (1)
--Hyper--> Bauwerk (1)
--Hyper--> Artefakt (1) | Werk (1)
    
```

Fig. 3. GermaNet senses and hyperonyms for the German word *Bank* ('bank', 'bench')

	<i>base</i>		<i>cs</i>	
# corpus terms total	545	100%	545	100%
- match	343	63%	483	89%
- unambiguous	211	39%	254	47%
- ambiguous	132	24%	229	42%
- no match	202	37%	62	11%

Table 2. Coverage of lemmatized corpus terms in GermaNet 5.2. Conditions: direct string match of the terms (*base*) and roots after compound splitting (*cs*). For the matched terms, we report the proportion of ambiguous vs. unambiguous terms.

the term *Bank*, a node with the OBJECT schema will be assigned the sense *Bank, Sitzbank* ('bench'). In contrast, a node with the AGGREGATE schema will be assigned the sense *Bank, Geldinstitut, ...* ('financial institute').

In contrast to verbal concepts, visualizations always need to be fully specified. Some concepts mentioned in route directions are too abstract to be reasonably mapped to any 3D representation, e.g. there is no way to visually represent the concept *bend* without specifying its angle and direction. Using an arbitrary visualization in those cases could be confusing and misleading for the user, therefore abstract concepts are not assigned a 3D prototype. 3D prototypes for fixed and moveable objects can be constructed semi-automatically. Their 2D representations are parsed from digitized architectural drawings and mapped into 3D. Since prototypical appearances vary across domains (e.g. buildings in industrial areas look different from buildings in rural areas), the drawings should represent the same domain for which the route directions shall be generated. Objects not contained in architectural drawings can be taken from existing repositories, e.g. Google 3D Warehouse⁴ or have to be created manually. For outdoor scenes, we additionally use existing 3D representations created by collaborating researchers from Geoinformatics. In order to enable our system to refer to and, if desired, also highlight separate parts of objects, we divide the prototypes into parts so that each part can be mapped to a concept individually. Thus, the 3D prototype ontology itself exhibits a hierarchical structure.

5 An Extended Route Representation Format: XOLS

The routes for direction giving are computed on a route graph representing only the spatial structure of the scenario using standard graph-search algorithms (e.g., Dijkstra or A*). The route graph can be deduced from the network structure represented in the spatial model. GATEWAYS, ULINKS and the spaces they link correspond to the nodes of the route graph, called decision points. Passable edges in the spatial model correspond to the edges of the graph. Thus, route graph and spatial model are naturally aligned. In the following, we describe a format for both the input route request and the output of the planning steps that includes information about the sequence of actions to be taken, the information required for generating instructions at each decision point (regarding both actions and landmarks), and the data needed for creating visualisations at decision points.

Format Specification: eXtended OpenLS We build upon the XML-based specifications of the OpenGIS Location Services⁵ (OpenLS), which are provided by the Open Geospatial Consortium. Cognitive OpenLS [5] extends the standard specification of one of these services, the Route Service, to include more cognitively relevant information, e.g. landmarks and chunking of similar instructions into one. For our format, we adopt some of the concepts defined there to suit the needs of our application. Our extension is based on the OpenLS Route Service. The

⁴ <http://sketchup.google.com/3dwarehouse/>

⁵ <http://www.opengeospatial.org/standards/ols>

connections between the different extensions of OpenLS are shown in Figure 4. The adapted format, eXtended OpenLS (XOLS), has been developed in cooperation with researchers from Geoinformatics and will be integrated into their implementation of the OpenLS Route Service.

Since our architecture combines indoor and outdoor scenarios, a major requirement for our format is that concepts necessary to represent indoor routes are integrated into the originally street-based Route Service. This concerns both landmark structures and types of junctions and turns. Below we give details for the elements that we modified or added and that are particularly important for our task.

Basics

The representation of a route is organized as a list of *maneuvers*, which represent actions leading from one decision point to the next. Each maneuver is associated with information about action type, location, landmarks, junction type and direction of turn.

Extension of the *Maneuver* Element

The two following attributes of the *Maneuver* element were adopted from Cognitive OpenLS and extended:

junctionType. This attribute holds one of the possible values for defining junctions. This includes street junction types (e.g. *Intersection* and *Roundabout*), as well as ways of entering buildings or vehicles (e.g. *EntranceRamp*, *ExitRamp*, and *BoardingRamp*). We added connections between parts of a building (e.g. *Door*, *Stairs* and *Elevator*).

directionOfTurn. This attribute holds one of the possible values for describing the direction of turns, which amount to ten possible values: *Straight*, *KeepLeft*, *KeepRight*, *SlightLeft*, *Left*, *SharpLeft*, *SlightRight*, *Right*, *SharpRight*, *UTurn*. We included the values *Up* and *Down* for (indoor) level changes.

XOLS adds new elements to the basic definition of a maneuver: *JunctionCategory*, which is adopted from Cognitive OpenLS, and *UsedLandMark*, a new element defined in XOLS.

JunctionCategory. This element indicates the category of the intersection and categorizes the branches as to where the route continues and where it does not [5].

UsedLandMark. This element indicates a landmark located at the current route segment. The elements are represented as children of *UsedLandMarkList*.

Representing Landmarks

LandMark. This element provides information about a single landmark. Since we do not need all details from the landmark taxonomy defined in Cognitive OpenLS, XOLS only uses a subset of this information.

6 Applying Aligned Knowledge Bases and XOLS for Generating Natural Directions

In the following we describe how the aligned knowledge bases, the concept ontology and the spatial model, and the route representation format XOLS will be applied in different components of our system.

Hierarchical representation and levels of granularity. The hierarchical organization of space in the spatial model enables the generation of verbal instructions and accompanying visualizations on arbitrary levels of granularity. For a given route, we locate the nodes representing start and destination points on the CONTAINER level and walk up the hierarchy, trying to find a connecting path at the highest level of aggregation possible. This level is determined by weights for the edges in the spatial model, which are provided by a user model that formalizes the user's knowledge of the local environment. For example, users with good knowledge of the area around the start point but less confidence in the destination area will receive directions on a higher level for the beginning steps and a more detailed description of actions towards the destination.

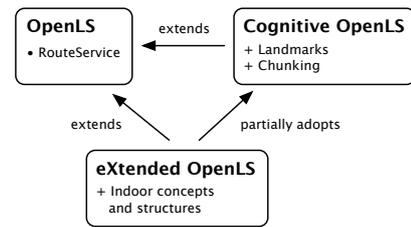


Fig. 4. Extension structure for XOLS.

Since granularity plays an important role in our system, it is important that the format provides transparent representations of higher-level elements, e.g., chunked actions such as *Make a left at the third intersection*, which implies going straight at two intersections. In case the user needs more detailed information, it can easily be derived from the embedded elements, to generate instructions such as: *Walk straight at the intersection with the gas station. Continue straight at the intersection with the tram station. Make a left at the following intersection.* Chunking is already integrated into Cognitive OpenLS [5]. Including it into XOLS will be part of our future work.

Landmark and attribute selection. As mentioned above, landmarks are an important component of route directions. Once the path through the spatial model has been determined, we access its nodes and use the link to the concept ontology to access object properties. We will apply salience measures and constraints to determine which properties will be used in generating referring expressions and the corresponding visualizations. Salience measures include object attributes represented in the spatial model and concept ontology (e.g. type, size, color), the visibility of a possible landmarks and whether the object is moveable or fixed. Here, we build on the work on visual salience by [9] as well as the approach for formalizing structural salience by [12].

Connecting the spatial model and XOLS. The conceptualization of spatial structures crucially depends on the perspective from which they are being approached. For example, a junction with three branches is called T-junction if approached from the branch forcing a turn (because no straight continuation is possible), but not if it is reached from a different branch which allows for straight continuation. Furthermore, for deciding which structures and objects are good landmark candidates, both perspective and visibility have to be taken into account. Therefore, in order to determine appropriate referring expressions, static information from the spatial model will be combined with directional information from the route under consideration.

Towards generating multimodal directions. As a first step towards a corpus-based NLG approach, we automatically align the route representations provided by XOLS with an annotated corpus of NL directions given by humans for the respective routes [21]. We will enhance these alignments using the spatial model in order to capture aspects of the route that are not inferrable from XOLS. In a second step, the induced alignments will serve to identify which elements of a route to select for verbalization, and to deliver correspondences between route segments and linguistic expressions that will be used as a basis for statistical NL generation. For graphical output, we will select relevant route segments and visualize objects using the coordinates and orientation attributes stored in the spatial model and the associated 3D representations in the concept ontology. We will use a constraint-based architecture to model the interaction between both modalities and synchronously generate language and accompanying 3D scenes. Examples of potential interacting constraints include but are not limited to the following:

1. If an object is mentioned verbally, it should also be included in the visualization.
Vice versa: If an object is not visible from the user's current position, it should not be mentioned verbally, unless it is salient and its location is known by the user.
2. If reference to one object is needed to describe another, at least the former should be visible.
3. If a referring expression mentions visual attributes of an object, the actual appearance has to match the description.
Vice versa: If no matching texture or material is available, a different attribute should be used for the referring expression.
4. If specific actions are mentioned in an instruction, they should be embedded in the visualized context. For example, a turning instruction has to be visualized using arrows or by appropriate camera movement.
5. If a situational parameter of an action is visible and can be left implicit on linguistic grounds, it may be omitted for reasons of economy.
6. The presentation of visualizations should be synchronized with the verbal realizations of events.

7 Conclusions

We presented an architecture for representing knowledge required to generate natural language route directions and corresponding 3D scenes. Geographic knowledge serves as an anchor to link the linguistic and visual modalities of a scene. It is represented in a hierarchical spatial model extending the Schematic Geometry by [22]. We align this model with a concept ontology that associates linguistic expressions with graphical representations, providing object properties needed for generating natural language directions and 3D scenes. In addition to these aligned knowledge bases, we introduced a route representation format based on OpenGIS standards. These components provide the building blocks for our future work in generating linguistic and visual route directions.

References

1. Bateman, J.: Situating spatial language and the role of ontology: Issues and outlook. *Language and Linguistics Compass* 4(8), 639–664 (2010)
2. Dale, R., Geldof, S., Prost, J.P.: Using natural language generation in automatic route description. *Journal of Research and Practice in Information Technology* 37(1), 89–106 (2005)
3. Furlan, A., Baldwin, T., Klippel, A.: Landmark classification for route directions. In: *Proceedings of the 4th ACL-SIGSEM Workshop on Prepositions, Prague, Czech Republic, 28 June 2007*. pp. 9–16 (2007)
4. Habel, C.: Incremental generation of multimodal route instructions. In: Freedman, R., Callaway, C. (eds.) *Working Papers of the 2003 AAAI Spring Symposium on Natural Language Generation in Spoken and Written Dialogue*. pp. 44–51. AAAI Press, Menlo Park, California (2003)
5. Hansen, S., Richter, K.F., Klippel, A.: Landmarks in OpenLS: A data structure for cognitive ergonomic route directions. In: *Proceedings of the 4th International Conference on Geographic Information Science, Münster, Germany, 20-23 September 2006*. pp. 128–144 (2006)
6. Jackendoff, R.: The architecture of the linguistic-spatial interface. In: Bloom, P., Peterson, M.A., Nadel, L., Garrett, M.F. (eds.) *Language and Space*, pp. 1–30. MIT Press, Cambridge, MA (1999)
7. Johnson, M.: *The body in the mind: The bodily basis of meaning, imagination, and reason*. University of Chicago Press (1987)
8. Jørgensen, F., Lønning, J.T.: A minimal recursion semantic analysis of locatives. *Computational Linguistics* 35, 229–270 (2009)
9. Kelleher, J.D.: *A Perceptually Based Computational Framework for the Interpretation of Spatial Language*. Ph.D. thesis, Dublin City University (2003)
10. Klippel, A.: *Wayfinding Choremes – Conceptualizing Wayfinding and Route Direction Elements*. Ph.D. thesis, University of Bremen (2003)
11. Klippel, A., Hansen, S., Richter, K.F., Winter, S.: Urban granularities – a data structure for cognitively ergonomic route directions. *GeoInformatica* 13(2), 223–247 (2009)
12. Klippel, A., Winter, S.: Structural salience of landmarks for route directions. In: Cohn, A., Mark, D. (eds.) *Spatial Information Theory. International Conference COSIT*. pp. 347–362. Springer, Berlin (2005)
13. Koller, A., Striegnitz, K., Byron, D., Cassell, J., Dale, R., Moore, J., Oberlander, J.: *The First Challenge on Generating Instructions in Virtual Environments*. In: Krahmer, E., Theune, M. (eds.) *Empirical Methods in Natural Language Generation*. Springer, Berlin (2010)
14. Landau, B., Jackendoff, R.: “What” and “Where” in spatial language and spatial cognition. *Behavioral and Brain Sciences* 16, 217–265 (1993)
15. Lemnitzer, L., Kunze, C.: GermaNet – representation, visualization, application. In: *Proceedings of the 3rd International Conference on Language Resources and Evaluation, Las Palmas, Canary Islands, Spain, 29–31 May 2002*. pp. 1485–1491 (2002)
16. Levinson, S.C.: *Frames of reference and Molyneux’s question: Crosslinguistic evidence*, pp. 109–169. *Language, speech, and communication*, MIT Press, Cambridge, MA (1996)
17. Lorenz, B., Ohlbach, H.J., Stoffel, E.P.: A hybrid spatial model for representing indoor environments. In: *Proceedings of 6th International Symposium on Web and Wireless Geographical Information Systems (W2GIS 2006), Hong Kong, China, 4-5 December 2006*. vol. 4295, pp. 102–112 (2006)
18. Maaß, W., Wazinski, P., Herzog, G.: *Vitra Guide: Multimodal route descriptions for computer assisted vehicle navigation*. In: *Proceedings of the 6th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems (IEA/AIE-93)*. pp. 144–147. Edinburgh, Scotland, UK (1993)
19. Richter, K.F.: A uniform handling of different landmark types in route directions. In: Winter, S., Duckham, M., Kulik, L., Kuipers, B. (eds.) *Spatial Information Theory*. pp. 373–389. Springer, Berlin (2007)
20. Richter, K.F., Winter, S., Rüetschi, U.J.: Constructing hierarchical representations of indoor spaces. In: *Proceedings of the 10th International Conference on Mobile Data Management: Systems, Services and Middleware, Taipei, Taiwan, 18-20 May 2009*. pp. 686–691 (2009)
21. Roth, M., Frank, A.: Computing EM-based alignments of routes and route directions as a basis for natural language generation. In: *Proceedings of the 23rd International Conference on Computational Linguistics, Beijing, China, 23-27 August 2010*. pp. 958–966 (2010)
22. Rüetschi, U.J.: *Wayfinding in Scene Space: Modelling Transfers in Public Transport*. Ph.D. thesis, University of Zurich (2007)
23. Sorrows, M.E., Hirtle, S.C.: The nature of landmarks for real and electronic spaces. In: *Proceedings of the International Conference on Spatial Information Theory: Cognitive and Computational Foundations of Geographic Information Science, Stade, Germany, August 1999*. pp. 37–50. Springer, Berlin (1999)
24. Tom, A., Denis, M.: Language and spatial cognition: Comparing the roles of landmarks and street names in route instructions. *Applied Cognitive Psychology* 18, 1213–1230 (2004)
25. Tomko, M., Winter, S.: Pragmatic construction of destination descriptions for urban environments. *Spatial Cognition and Computation* 9(1), 1–29 (January 2009)
26. Tversky, B.: Structures of mental spaces: How people think about space. *Environment and Behavior* 35(1), 66–80 (2003)
27. Tversky, B., Lee, P.U.: How space structures language. In: Freksa, C., Habel, C., Wender, K.F. (eds.) *Spatial Cognition - An Interdisciplinary Approach to Representing and Processing Spatial Knowledge*. pp. 157–176. Springer, London, UK (1998)