From Descriptions to Depictions: A Conceptual Framework

Maria Vasardani[1], Sabine Timpf[1,3], Stephan Winter[1], Martin Tomko[2]

[1] Department of Infrastructure Engineering, [2] Faculty of Architecture, Building and Planning
University of Melbourne, Parkville, VIC, 3010, AU
{mvasardani, winter, tomkom}@unimelb.edu.au
[3] Department of Geography, University of Augsburg, 86135, Augsburg, DE
sabine.timpf@geo.uni-augsburg.de

Abstract. People use verbal descriptions and graphical depictions to communicate spatial information, thus externalizing their spatial mental representations. In many situations, such as in emergency response, the ability to translate the content of verbal descriptions into a conflict-free sketch map could greatly assist with the interpretation of the message. In this paper, we present an outline of a semi-automatic framework enabling seamless transition between verbal descriptions and graphical sketches of precinct-scale urban environments. The proposed framework relies on a three-step approach: NL parsing, with spatial named entity and spatial relation recognition in natural language text; the construction of the spatial Property Graph capturing the spatial relationships between pairs of entities; and the sketch drawing step where the identified entities are placed on a canvas in a manner that minimizes conflicts between the verbalized spatial relationships, thus providing a plausible representation of the described environment. The approach is manually demonstrated on a natural language description of a university campus, and the opportunities and challenges of the suggested framework are discussed. The paper concludes by highlighting the contributions of the framework and by providing some insights for its actual implementation.

Keywords: NL place descriptions, spatial information extraction, dynamic sketching

1 Introduction

People often use verbal and graphical (sketch) language to communicate information about spatial scenes in a way that externalizes the mental images they have created through direct or indirect experience with the environment. An intuitive spatial human-computer interaction should accommodate both modes of communication. For example, in emergency response, fast automatic sketching of the affected environment from a caller's verbal input can be critical, in cases where direct georeferencing is not possible. By asking for additional natural language (NL) input from the caller, thus initiating a dialog, the service could refine or disambiguate the sketch. In an ef-

adfa, p. 1, 2013.

fort to enable seamless switching between the two modes of communication, and since different groups already work on generating spatial NL expressions from map sketches, this research focuses on generating spatial sketches from unrestricted NL.

An important aspect of both verbal and graphical spatial language is the extended use of spatial relations to communicate the various absolute and relative locations of features in the descriptions. In verbal language the expression of spatial relations is done with abstract, yet functional linguistic terms. Such terms enable the understanding of the infinite variability of the world, while revealing aspects of the environment with some functional relevance [1]. And sketch maps are themselves visually simple, two-dimensional representations of features in an environment and the spatial relations among them. Sketch maps can serve as an intuitive user interface for geospatial computer applications (e.g., [2]). In this paper, we demonstrate the conceptual outline of an algorithm for the (semi) automatic construction of plausible, two-dimensional sketch maps, based on the extraction of spatial relations and features from text-based natural language (NL) descriptions.

Both verbal descriptions and sketches differ substantially from how computers and geospatial services generally process and present geographic information. While work from various communities is being done separately on NL processing for extracting geospatial information (e.g., [3]), on comparing NL descriptions and sketches that express the same cognitive image (e.g., [4]), and on assessing the information presented on a man-made sketch map (e.g., [5]), there is no effort so far—to the best of our knowledge—on assessing whether the information extracted from NL processing can result in a plausible, two-dimensional sketch-like representation of the described environment. This transformation forms a substantial challenge, acknowledging that NL descriptions can be notoriously underspecified and ambiguous.

This paper is a first take on this challenge of going from descriptions to depictions. Our hypothesis is that spatial information extracted from NL place descriptions can be (semi) automatically represented in a topological conflict-free sketch map. Such a sketch can reveal the way that the person who expressed the NL description conceptualizes the environment. Positive evidence would support Tversky's suggestion that the meanings of linguistic elements can be mapped straight onto the meanings of depicted elements and vice versa [6].

The conceptual algorithm we are proposing in this work actually comprises three steps: 1) a NL parsing step, 2) a graph-producing step, and 3) and a sketch-drawing step. We are expecting unrestricted NL descriptions of a specific, urban environment. Thus, implementation of the parsing step depends on ongoing work on part-of-speech tagging, spatial named entity recognition, and toponym resolution for a parser that would be able to identify the geographic features in a description and the spatial relations among them (cf. [7–13]). For a plausible automated placement of geographic features on the plane, spatial relations are categorized according to formal models of topological, cardinal, projection and orientation relations ([14–17]). During the second phase, the identified spatial features and relations are placed on a spatial Property Graph (sPG). The sPG strips away the cognitive details of the NL place descriptions and provides the basic spatial information in the form of reference objects (RO), locata (L)—objects to be located in relation to the RO—and the identified spatial relations (*r*) between them. The ROs and Ls are the nodes in such graphs, while the *r*s are labeled edges. During the final phase, all ROs and Ls are arranged on the plane, in a sketch that represents a plausible representation of the described environment, according to the extracted spatial information. By plausible here, we mean a topologically conflict-free representation that is most probably one of many such plausible representations that one can derive from the descriptions. The conceptual design of the proposed algorithm is based on a case study of NL descriptions from memory of the layout of a university campus, from volunteer graduate students.

The remainder of the paper is structured as follows: Section 2 discusses background work on concepts and tools necessary for the design of our framework, and our approach is presented in Section 3—the three steps of the conceptual algorithm. Section 4 discusses the case study and an example of (manually) applying the algorithm to produce the sPG and corresponding sketch(es). In Section 5 we study the results and provide an overall discussion of the important aspects of our conceptual design. We conclude this paper and discuss future work in Section 6.

2 Related Work

In this section we discuss previous work that is related to the three steps of our approach, namely the NL parsing step, the property graph production, and the sketch drawing, and point out the differences with comparable approaches.

2.1 Extracting spatial information from NL descriptions

We are interested in an algorithm that can produce a plausible two-dimensional representation of the place described in an unrestricted NL narrative. Some earlier work on NL processing had focused on the extraction of specific and application-dependent spatial information from restricted language [7, 18, 19]. Most computational aspects of spatial information do not consider linguistic issues. The intricate, bendable and uncontrollable linguistic concepts that lead to under- or over-specification when mapping NL to formal representations of spatial information can be regarded responsible for the lack in domain-independent applications. Bateman *et al.* [20] argue extensively about this shortage, especially when it comes to formal models of spatial relations which are based on the logic of human spatial cognition. Such models do not take into consideration the linguistic constructs and the way that people actually verbalize spatial relations. In [21] the need for semantics and a linguistically-oriented spatial ontology that can facilitate mapping between NL and spatial calculi, such as the Generalized Upper Model (GUM) ontology [22], is supported. Recently, however, work on parsers that can process unrestricted NL for spatial information and thus come closer to the requirements of our algorithm has surfaced, and is described here.

Kordjamshidi *et al.* [23] introduce *spatial role labeling* (SpRL), a method for assigned the roles of ROs, locata and *spatial relations* roles to terms. NL sentences are tagged according to the *holistic spatial semantic* theory in [24], and later mapped onto formal spatial relations frameworks, to facilitate subsequent spatial reasoning. The authors propose machine learning methods to cope with the various sources of ambiguity of spatial information in NL. In [8] they employ the proposition project (TPP) used in SemEval-2007 [25] to disambiguate the prepositions' spatial meanings and enhance their SpRL technique. Evaluation of the spatial role labeling task is done using the General Upper Model (GUM) spatial ontology [22].

Another group, interested in the generation of rendered 3D images from NL scene descriptions, created a dependency parser, which labels all terms in a sentence according to their direct or indirect relation with the (finite) verb [10] . The first version of the system (Words-Eye) handled 200 verbs in an *ad hoc* manner with no systematic semantic modeling, while their current system utilizes the Scenario-Based Lexical Resource (SBLR) [11]. SBLR consists of an ontology and lexical semantic information extracted from WordNet [26] and FrameNet [27]. With such parsers, the goal of labeling terms of unrestricted NL descriptions with specific spatial roles becomes much more feasible, and they are considered a necessary prerequisite for the conceptual design of the proposed framework.

2.2 Property Graphs

In our conceptual framework, there is an intermediate step between the input NL processing and the two-dimensional representation of the layout—the creation of a spatial Property Graph (sPG). Property graphs [28–31] are a special type of graph data models introduced whenever the data complexity exceeds the capabilities of the relational model as in e.g., transport networks [32] or spatially embedded networks [33]. Property graphs are directed, labeled, edge- and node-attributed multi-graphs, with nodes representing entities, and edges the relations between them [31]. We call the property graph constructed in the suggested framework a *spatial* property graph, emphasizing the fact that the labeled directed edges between nodes represent spatial relations with (possible) properties, frequently belonging to formal models.

2.3 Mental maps

Producing a sketch map from text descriptions cannot be a direct action, but requires first understanding of the process by which people built mental models [34]. This process, called cognitive mapping, comprises a series of psychological transformations by which an individual acquires, codes, stores, recalls, and decodes information about the relative locations and attributes of phenomena in their everyday spatial environment" (p.9, [35]). Sketching the spatial relations from a mental model is called mental mapping and the end results, i.e. mental maps, are a prominent source for researching human's understanding and feelings about geographic places [36, 37].

The properties of mental maps exhibit all of the processes humans go through by first cognitive mapping and then mental mapping: they are highly simplified, schematized, selective and distorted [38]. They are entirely qualitative, thus reflecting humans' thinking about geographic space [39]. Topological relations are usually correctly preserved, as are order and sequence relations [40, 41]. Previous research indicates that information represented in a mental map is preferentially North-Up oriented

[42]. From this discussion, it should be clear that mental maps have specific properties that this research intends to exploit in the sketching process.

2.4 Comparable approaches

With a focus on spatial relations inference to extract new information, Wiebrock *et al.* [12] produce a visualization from descriptions of spatial layouts. They also produce a graph of objects and spatial relations between them, which resembles the sPG suggested here, in that their graph is also a labeled, directed graph where the edges are spatial relations. The authors then use multiplication of transformation matrices, constraint propagation, and verification to realize inference. Apart from the focus on inference of new spatial information alongside the visualization of the spatial layout, their approach differs from ours in that input to their system is not unrestricted NL descriptions, but a sequence of propositions from instructed agents. They also consider only table-top objects and room environment descriptions, with a main interest in human-robot communication for indoor robot movements, instead of relations between features in urban environments, which are what we are mainly focusing on. Finally, while for robot navigation a representation of indoor environments must be consistent with reality at every location, we require our representations to be plausible (i.e., free of topological conflict), but they need not be consistent with reality.

In another approach, a route description from search engines is used to produce a map of the route called LineDrive [43]. In this work, spatial cognition insights are used to schematize the route map, i.e. linearizing the streets, straightening angles between route segments, emphasizing the route and de-emphasizing (up to disappearance) the context information, resulting in a very clear and simple spatial representation similar to sketch maps. Among other fundamental differences, input information for the sketch comes from an already interpreted set of route instructions that were produced by a database search, whereas we focus on the generation of plausible sketches from unrestricted NL descriptions of a spatial layout.

Finally, in an another effort to generate sketch maps from route descriptions, Fraczak [44] attempts to directly map NL expressions to a so called graphic code, which in turn is supposed to generate a map. From the literature it is not clear if this mapping was ever completely implemented. Again, the author focuses on NL descriptions of routes rather than spatial layouts. In addition, Fraczak's goal is to derive an interpretation of the route description that is consistent with reality, whereas we stress the dynamic nature of the sketching process in this framework that allows for several plausible solutions.

3 Approach

The general approach we are proposing consists of three steps:

1. An unrestricted NL parser processes the descriptions of a specific environment's layout and labels the terms in each sentence according to whether they carry a spatial role or not.

2. Algorithm 1 takes the parser's output and produces a spatial Property Graph (sPG).
3. Algorithm 2 processes the sPG graph and produces a plausible 2D representation of the described environment's layout, in the form of a sketch.

3.1 Parser output

The first step in the conceptual framework is feeding the parser with an unrestricted NL description of the spatial layout of an environment. The parser labels the terms in each sentence according to a spatial role set, consisting of some of the basic spatial semantic concepts that are defined as 'universals' in the spatial semantics literature [24], also used in the annotation scheme developed in [8], and contains the core roles: *reference object* (RO), *locatum* (L), *spatial relation* (*r*), and *none*—if the word provides no additional spatial information. The *spatial relation* can be a *static spatial indicator* or a *motion indicator*. We decided to employ this already defined spatial role set because it is commonly used in the current spatial semantics parsers' literature (cf. [9, 10, 12]), with some of the roles having alternative names such as *trajectory* for the L, or *landmark* and *relatum* for the RO (cf. [24]). The roles of RO and L are identified for features on level 3-building and level 4-street of the granularity schema developed in [45], as we are focusing on spatial relations and features of an urban environment.

The RO is the reference object in relation to which the location or the trajectory of motion of the locatum L is indicated. The L is the entity whose location or trajectory is of relevance, and it can be static or dynamic. The *r* explains the type of spatial relation between RO and L and is usually a preposition (e.g., "house *on* the lake"), but can also be a verb (e.g., "buildings *surrounding* the lawn"), or a noun (e.g., "the road to *the east*), or even implicit (e.g., "Paris, France", implying *in*). In GUM ontology [22] it is defined as *spatial modality*, and is the main axis of the spatial relation. It can also be an indicator of motion, e.g., in the form of *motion verbs* [8].

The desired and required final output of the parser is a list of ordered triplets of the form in Equation 1,

$$<\mathrm{L}\ r\ \mathrm{RO}>i, \text{ with } i=1..n \tag{1}$$

where n is the number of triplets, and i is the index of each triplet according to the order with which it was identified in the NL description. At this point, words labeled as *none* are not considered any further. It should be made clear here, that the same spatial elements may participate in multiple triplets. For example, from the sentence "You can also find NAB and Commonwealth bank branches near the Union House", the parser should produce the following triplets: <NAB *near* Union House>, <Commonwealth bank *near* Union House>. In other words, 'Union House' is the RO for both Ls, 'NAB' and 'Commonwealth bank'. For this reason, and to assist with the algorithm's design, the parser assigns to each RO a unique numerical identifier RO_ID. This identifier is later used to: a) help in recognizing how many Ls are related to the same RO, and b) to help resolve synonymy issues, i.e., if two ROs with different RO_IDs at first are recognized as the same entity, but with different names (e.g., one official and one shorter for brevity), then they get assigned the same RO_ID. In a similar manner, synonymy is resolved between locatums (Ls).

Apart from relations in the form of prepositions, adjectives, nouns, or implied ones, the parser is required to recognize the more complex, path indicating relations such as *across*, *along*, *down the hill from*, etc. which are identified as *PathRepresentingInternals* in the GUM ontology [22]. Motion verbs, when conveying spatial information should also be detected as such. For example, in "You may continue north, past the bakery to reach the market", by interpreting 'may continue north', 'past' and 'to reach' the parser should produce the triplet <market *north* bakery> recognizing that there is a path that leads from one feature to next, in a sequence. Finally, the parser is able to resolve anaphora and other co-references (cf. [23]). For example, in "…of the old buildings. These buildings are…" the parser is able to identify that 'these buildings' refer to the 'old buildings' previously mentioned, and therefore are the same spatial entity.

Interpretation rules. Certain rules are imposed on the parser for the interpretation of the spatial semantics, for our conceptual framework to work properly.

Rule 1: Each description corpus falls entirely into a single top-level container. It is initially assumed that with input from the analyzer, the described spatial environment would be identified as the main container and reference system of each description. For example in our case study, the term 'campus' was recognized as such, and identified ROs and Ls were positioned in it. Use of superlative in descriptions then indicates the outer bounds of such reference system, and is picked out by the parser. For example, in "The Arts School is furthest north on campus", the output triplet should be <Arts School *furthest north on* campus>, instead of just <Arts School *north on* campus>.

Rule 2: A sequence of Ls with no additional spatial information produces inherited relations to the same RO. In the example "From the intersection by continuing north you can get to the museum, the library and then reach the market.", the expected output triples are <museum *north* intersection>, <library *north* museum> and <market *north* library>. Hence the relation *north* is inherited for all triplets, the L of each triplet becomes the RO in the next, and orientation does not change (see Rule 3).

Rule 3: Unless explicitly stated otherwise, the initial orientation identified does not change – this is a general phenomenon when dealing with the orientation of a description. However, when a change in direction is stated, and when there is no obvious RO after the change, then the notion of a corner-point (CP_i) is introduced. The CP_i acts as the RO for the next bit of the description when the direction has changed, but also as the L for the previous part of the description. So, the output triplets in the example "The second path from the south entrance takes you a little bit to the west and then switches to the north. From this path […] you can reach the University House by continuing north" should be <CP_1 *west* south entrance> and <University House *north* CP_1>.

Rule 4: Verbs that imply a container, such as 'contains' or 'houses' are interpreted as the containment relation *in*, and the container feature is then considered the RO of the relation. In the example "The Union House is containing many cafes and restaurants…" the output triplets are <cafes *in* Union House> and <restaurants *in* Union

House>. In these cases, if the RO is involved in another relation, then the relation also holds for the contained features as well, however, the reverse is not true. In the example "[…] in Baretto's in the Alan Gilbert building, across Grattan St. is the medical building", then apart from 'Alan Gilbert building', also 'Baretto's' is considered as a possible locatum across the street from the medical buildings.

Rule 5: Named streets are assigned a role of either RO or L, whereas paths, which are usually unnamed or inferred, are only expressed through the spatial relations of the spatial elements they connect.

Rule 6: Complex spatial relationships including two or more ROs or Ls are broken into binary relationships. Complicated descriptions require extra inference capabilities from the parser. Prepositions such as *around*, require usually an orientation for locating features. In the example "Around South Lawn you would see (from left to right) Graduate House, the MBS […]" the parser should be able to assign the relation *around* between South Lawn and all the mentioned locatums, and also assign the relation *to the right of* between the locatums in the mentioned sequence, i.e. <Graduate House *around* South Lawn>, <MBS *around* South Lawn>, <MBS *to the right of* Graduate House> and so on. In addition, the desired parser output triplets include spatial relations that are binary in nature, Therefore, with ternary relations such as 'A *between* B and C', and 'A *across the street from* B', the parser is required to break them down into simple binary relations that fit with the rest. *Between* is decomposed into <A *between* B> and <A *between* C>, while *across* into <street *between* A> and <street *between* B>. This way, we maintain uniformity in the required output of binary relations' triplets.

3.2 Spatial Property Graph algorithm

In this section, the second step of the suggested framework is described; a graph-producing algorithm that creates the spatial-Property Graph (sPG). Input is the ordered list of triplets produced by the NL parser at the first step. A property graph has a set of nodes and a set of edges. Each node has a unique identifier, a set of incoming and a set of outgoing edges, and a collection of node properties defined as key/value pairs. An edge has a unique identifier, an outgoing and an incoming node, a label denoting the type of spatial relationship between the two nodes and a collection of properties defined as key/value pairs. These key/value pairs make it possible to describe in detail the relationship between two entities. In our framework, the nodes represent ROs and Ls and the edges the spatial relations between them. In the case of a spatial PG, the types of relations may belong to different formal frameworks of relations such as topological, directional, or cardinal relations. The direction of the edge points from the L to the RO, with the exception of *between* relations that are represented with bi-directional edges, and additionally decomposed into pairs of directed edges, showing the existence of one L and two ROs (see Rule 6). There can be multiple directed edges between the same pair of RO and L (i.e., combined with a logical AND in a formal interpretation).

Below, the sPG constructing algorithm is outlined, and an example graph from the case study is presented in Section 4 (Fig. 2).

> ***Algorithm1*: Constructing the spatial Property Graph**
>
> Input = *n* triplets in the ordered list, each assigned with a sequence index *i*:
>
> For *i*=1...n,
>
> a. If not already created, create nodes for the RO and the L in <L *r* RO>$_i$
> b. Create a directed edge E_j pointing to RO and originating from L, labeled with *r*
>
> Next *i*

3.3 Sketch Drawing algorithm

The last step in the text-to-sketch conversion framework is the generation of a sketch guided by the graph structure. In the current state, this process is performed by a person—it is not (yet) expect it to work unsupervised. The expected outcome is a plausible representation of the relationships among spatial objects encoded in the graph, in a topologically conflict-free sketch, called a *topological sketch*.

Related literature suggests that NL place descriptions are sequential and clustered [46]. In all the case study examples analyzed for this paper, people followed the communication maxims [47]. The descriptions are not just a collection of facts, but they reveal each person's preference for what is important, relevant, and in which sequence it should be presented. In their text-based descriptions, volunteers started from a certain part of the environment, described it in sequence of spatially related elements, and when done, moved on to the next, possibly neighboring part. The steps in our algorithm are designed with this characteristic in mind: The parser's output triplets are ordered according to the order with which ROs, Ls, and *r*s are introduced in each description. The algorithm places all ROs and Ls in the sketch, one at a time, starting with the first RO in the sPG.

McNamara *et al*. [41] identified region membership as an important property of spatial environments. Therefore, (small) regions are the main spatial feature to be placed on the sketch canvas. However, in this framework, this idea is taken further than in McNamara *et al.* by giving any entity in the graph the shape of a rectangle that can be dynamically enlarged to include other entities, shrunk, or elongated as needed. Furthermore, a general reference frame is introduced in the form of a double cross of absolute cardinal directions [16], with the true North.

Basic process. The first entity to be placed on the blank canvas is the container feature for all named entities in the description, in our case study the 'campus'. This entity is treated as the reference frame for all subsequent relations. In this bounding feature, a double cross is introduced as reference frame for absolute cardinal directions (Fig. 1a). Next, find in the sPG the reference object RO_*i* with the smallest sequence number and place its rectangle in the double cross—if no cardinal information with respect to the reference frame is given, the rectangle shall be placed in the center region of the double cross. If a cardinal relation is given, then the rectangle is sketched in the appropriate place on the canvas (Fig. 1b).

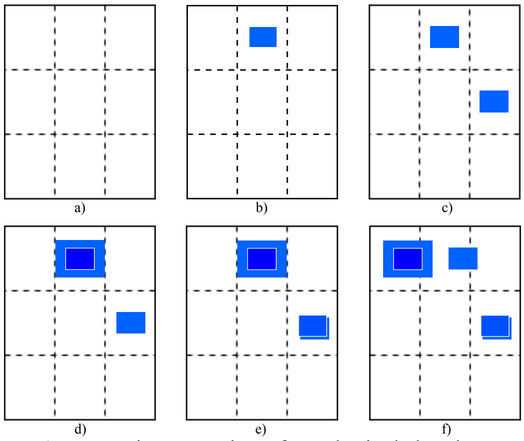

Figure 1. Dynamic generation of topological sketch example.

Once the RO_*i* is placed, a spatial relation of the type cardinal direction is selected from the property graph, that relates a L to that RO_*i*. The rectangle for the corresponding L is placed on the canvas in the given cardinal direction relation at a fixed distance from the RO_*i* (Fig. 1c). This is repeated for all relations containing cardinal directions relating to RO_*i*. When no other cardinal direction relations with RO_*i* remain, one of the other relations with RO_*i* shall be selected for inclusion in the sketch, and each corresponding L is placed on the canvas in the given spatial relation. If an L already exists from a previous relation, its position is checked for consistency and readjusted if necessary, maintaining the sketch conflict free. This is repeated until no spatial relations with RO_*i* remain in the graph. Next, RO_*i +1* is selected and its rectangle sketched on the canvas; the procedure is repeated until all ROs and their Ls are placed on the canvas. Below is the outline of the sketch algorithm.

***Algorithm 2*: Generating a topological sketch**

1. If known, draw the RO that is considered the main container and introduce a double cross reference system in it. Else, introduce the double cross and consider it the main reference frame.

2. For *i*=1…*n,* where *n* is the number of unique RO_IDs,
   a. Place RO_*i* as a rectangular feature on the double cross. If there is information available for its absolute cardinal position place it in the appropriate cell; otherwise place it in the central cell.
   b. For all the <L *r* RO_*i*> triplets,
      i. Identify all cardinal directions relations *r* and, if not already there, place the corresponding L on the frame.
      ii. For all other relations, if not already there, place the corresponding L relative to RO_*i* according to *r* on the frame.

   Next *i*

Defining the heading. One characteristic of natural language descriptions of places is the switch of perspective or reference frame. Following Meilinger [48], who proposes to allow different reference frames for each location, we assign each rectangle a heading, which forms one axis of the reference frame. The heading of a rectangle is defined as the direction in which the entrance of a building lays, or in the case of other entities such as the campus or the park, the direction to the North. This rule is justified by recent findings by Meneghetti *et al.* [42], who showed that the mental representation derived from spatial descriptions is North-Up oriented. In the absence of any other orientation information we will thus default the heading value to "North". More complex rules could be made here, e.g. making the heading dependent on the features already placed on the canvas or on the type of the entity. We will however leave this refinement for future work.

Dynamically adjusting the feature placement. All rectangles can be enlarged, shrunk, or moved to accommodate a new rectangle on the canvas. For example, if there exists a spatial relation "IN" between two rectangles then the reference rectangle is enlarged to make space for the new feature (Fig. 1d). On the canvas, the "IN" relation is represented as the topological relation "RO contains L". In the special case of 3D spatial relations, e.g., spatial relation "L underneath RO", two rectangles are allowed to overlap partially, i.e. showing the topological relation "RO overlaps L" (Fig. 1e). Rectangles that represent streets or other linear features (e.g., rivers) are allowed to stretch alongside several other objects when information is available or may be inferred (see for an example Section 4 – case study).

During the placement process it can happen that a rectangle cannot be put on the canvas, because doing so would violate one or several of the relations that already exist. In this case the existing rectangles can also be moved as long as their original relations are not violated. For example in Fig. 1f, the containing rectangle is moved a little to the west, to leave room for the new rectangle that is placed to its east, to satisfy the sequence of information, as this is revealed by the ordered triplets. When a rectangle cannot be placed without producing a conflict with existing features, past decisions on how and where to place features are revisited, aiming to find another solution that still satisfies all existing spatial relations. For this process to work we introduce a data structure called a *placement decision tree*.

Placement decision tree. The placement decision tree records incrementally the history of decisions of where features were placed and their potential alternative placements. For example, when placing a rectangle on the canvas using the relation "entity1 is in the northern part of the campus", three placements of that rectangle could be viable, since a system with eight cardinal directions is used. In the placement decision tree, three new links (north, north-east, north-west) with new leaf nodes are added, one of which (the one attached to the link "north") contains entity1. If at any point in the dynamic sketching process a conflict arises out of a potential placement, then the last decision is revisited and changed, i.e., up in the decision tree, and then the new feature is placed again.

The decision on a particular placement thus corresponds to a default hypothesis if alternative solutions exist. By recording not only the decision itself, but also the alternative solutions, we construct incrementally a search space in which we keep track of our hypotheses. In contrast to [49], our data structure is lazy, i.e., we only compute a new solution if a conflict is identified. The data structure reflects the under-specificity of the place description and can be used as a quality measure.

4 Case Study

The case study for testing our conceptual framework consists of a set of four NL descriptions of a university campus. The descriptions were submitted by a group of graduate students with varying degrees of familiarity with the campus. Students were asked to produce the descriptions from memory, as if explaining to a new student the layout of the university environment. No further directions or expectations of the descriptions were suggested. Below is a campus description (*Narrative A*) to which the suggested framework is applied and Table 1 contains the ordered set of triplets expected from the NL parser. As the parser is still not implemented, the triplets were manually produced and ordered according to the sequence of ROs and Ls in the description. Triplets 3 and 4 present an example of interpreting relation *across* as two *between* relations (Rule 6). After extracting the ordered list of triplets, Algorithm 1 (Section 3.2) was manually implemented on the list, and produced the sPG in Figure 2. Double-directed edges correspond to the *between* relationships discussed above, to show the difference between binary and ternary relations, represented also as binary.

Narrative A

"*We're sitting in Baretto's in the Alan Gilbert Building, across Grattan street is one of the medical buildings. Down the hill along Grattan street the new building being constructed is the Peter Doherty Institute and diagonally across the road (Royal Parade) is Melbourne Hospital. In the other direction the open area is University square (there is a carpark underneath) at the city end of University square is the law building. Across Grattan street from University square there is an entrance to the campus, straight ahead is an overpass building and to the right are the various Engineering buildings. The road goes in a big loop around the campus you can either go left towards the Medical buildings or right past the Engineering buildings then head North (away from University square). One other area you may want to explore is South Lawn which you can get to by going underneath the overpass building directly in front of you when you enter the campus.*"

Finally, Algorithm 2 (Section 3.3) is applied on the sPG and (at least) two plausible sketch maps are derived (Fig. 3-4). As is evident, there may be many topological sketches that can come out of one description. In the example above, it is not specified which side of 'Alan Gilbert Building' to place 'Peter Doherty Institute'; therefore, a decision needs to be made and recorded in the placement decision tree. Figure 2 shows the outcome topological sketch when first choosing one side of 'Alan Gilbert Building', while Figure 3 shows the outcome when choosing the other. Due to the description being underspecified, and since the rest of the spatial relations can be realized without conflict, both topological sketches are valid.

Table 1: Expected parser output – ordered list of triplets <L *r* RO>i

<Barreto's *in* Alan Gilbert Building>1
<Medical Building *across (Grattan St)* Alan Gilbert Building>2
<Grattan Street *between* Alan Gilbert Building>3
<Grattan Street *between* Medical Building>4
<Peter Doherty Institute *down the hill* Alan Gilbert Building>5
<Peter Doherty Institute *along* Grattan Street>6
<Melb. Hospital *diagonally across (Royal Parade)* Peter Doherty Institute>7
<Royal Parade *between* Peter Doherty Institute>8
<Royal Parade *between* Melbourne Hospital>9
<University Square *in the other direction from PDI* Alan Gilbert Building>10
<carpark *underneath* University Square>11
<law building *at the city end of* University Square>12
<entrance to the campus *across (Grattan Street)* University Square>13
<Grattan Street *between* University Square>14
<Grattan Street *between* entrance to the campus>15
<overpass building *straight ahead* entrance to the campus>16
<engineering buildings *to the right* entrance to the campus>17
<medial buildings *left* entrance to the campus>18
<overpass building *directly in front* entrance to the campus>19
<overpass building *between* entrance to the campus>20
<overpass building *between* South Lawn>21

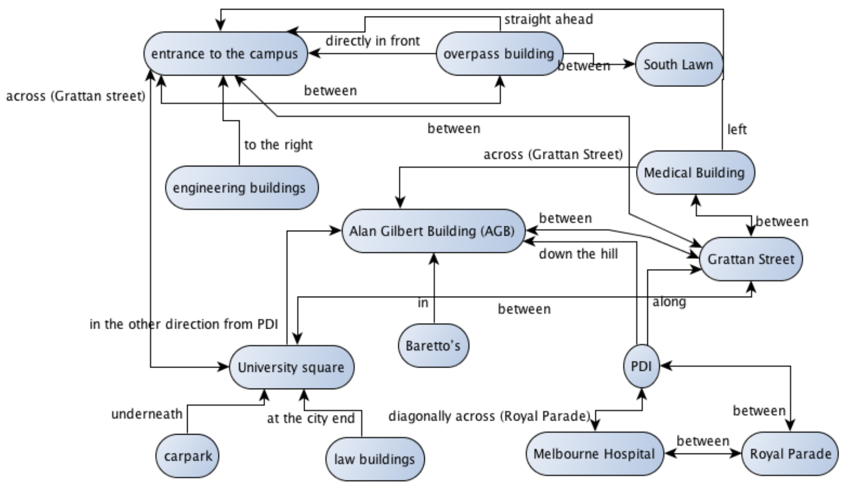

Figure 2. The sPG produced by applying Algorithm 1 on the example campus description

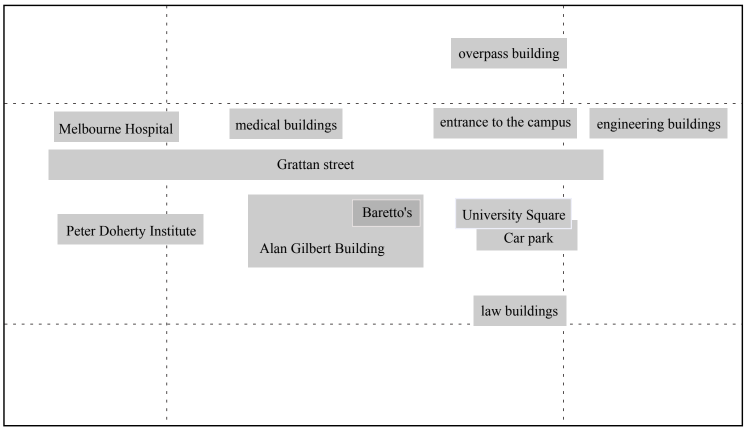

Figure 3. A plausible sketch map produced from the description by applying Algorithm 2

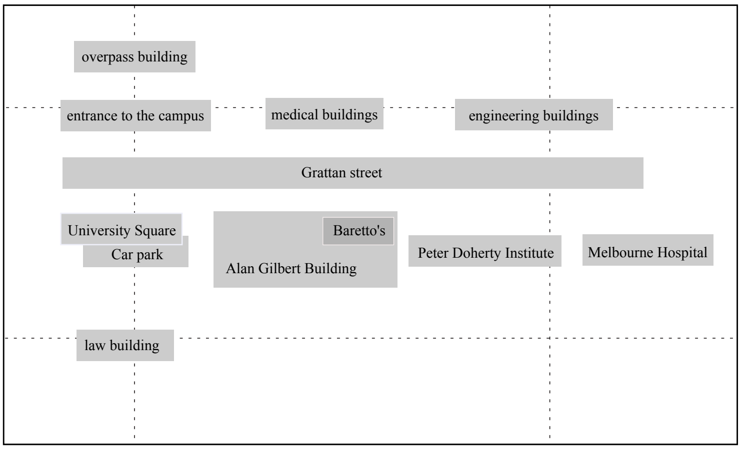

Figure 4. Another plausible sketch map produced by applying Algorithm 2

The suggested framework was also applied in the same manner to the rest of the NL descriptions in the set and produced sPGs and plausible topological sketches for all of them. The fact that plausible sketches could, in fact, be generated supports our hypothesis, considering that all descriptions were very different from each other and varied in style, even within the same description.

5 Discussion

The complexity of putting unrestricted NL onto a sketch map stems from the under-specificity of NL. Even the fact that we can have a multitude of plausible topological sketches is a direct effect of this observation. However, we believe that the suggested framework is a first step toward accommodating this language characteristic.

There are a few observations that can be immediately made from the NL place descriptions we collected for our case study. First of all, people have the tendency to mix different styles of spatial descriptions, alternating between survey and route views, or changing the reference object and switching from an allocentric to an egocentric description and vice versa. Even though this poses additional challenges to the extraction of the appropriate spatial roles (L, *r*, RO), there has been found no contradiction, and the suggested structure of the triplets is flexible enough to accommodate these different styles. For example, in the sentence "[…] the overpass building directly in front of you, when you enter the campus", the implied co-location of the person and the 'entrance of the campus' is enough to suggest that the 'entrance' can be used as the RO in the triplet. Thus, the structure allows for any type of description, and for the co-existence of different types of spatial relations. The interpretation of the relations follows then certain rules (Section 3.1) that make a topological sketch possible.

A second observation verifies related literature that suggests that NL place descriptions are sequential and clustered [46]. In all our examples, we found that people like to start from a certain part of the environment, describe it in sequence of spatially related elements, and when done, move on to the next, possibly neighboring part. There is evidence that people move through their mental representation when giving directions or describing places and regions. This is especially true for descriptions using a route perspective, but seems also to be true for a survey perspective [50]. The simplicity of the roles assigned in our framework makes it easy to identify the common spatial elements, which then act as the bridges between the various parts of described environment.

A third observation is the fact that people like to refer to the same things with different names or expressions (e.g., abbreviations), which adds another layer of difficulty when assigning spatial roles to the NL expressions. A different, but related problem is the identification of synonyms that are in different parts of the description. In the current approach, this identification is carried out by a person (who can consult gazetteers and other available knowledge sources), and is expected to be addressed in natural language processing.

The choice of the double cross of absolute cardinal directions to represent the main reference system is motivated by the need to delineate and structure the canvas in a uniform way. Cardinal relations provide the most information when trying to place things on a blank canvas, and are thus used first in our sketch producing algorithm whenever available. In this process, the double cross allows for suggested areas where elements can be placed relative to each other. When absolute relative cardinal relations are missing, we start with the assumption that the first mentioned RO is placed in the center, and things are located around it according to the rest of the available spatial relations.

We need to emphasize here that the production of the topological sketch is an incremental and dynamic sketching process. Therefore, spatial elements are allowed to shrink, expand, and move around in order to avoid topological conflict, as long as they do not violate previously established relations. For example, in the description, "the Arts building is to the north of the Union House", if we later on find out that there is another spatial object between the Arts building and the Union House toward the North, it can still be placed between them, by moving for example the Arts building further North, to make room for the new object. Also, in the case study example (Fig. 3-4), since there is available information about the span of "Grattan st." along certain objects from the extracted triplets (Table 1), the rectangle representing the street is stretched alongside the corresponding objects. Additionally, spatial objects must occupy an area (which can be resized accordingly), and they are not allowed to overlap, unless specifically stated so (e.g., "carpark *underneath* South Lawn").

Finally, by using the term 'plausible' we refer to the fact that the outcome sketches must be free of topological conflict. This means that from a single NL description there may be multiple sketch map interpretations—homomorphic representations are all 'plausible' in this sense. A mental model represents one possibility, capturing what is common to all the different ways in which the possibility may occur [51]. The spatial objects can be placed on the plane in a way that agrees with the incremental revelation of information from the ordered triplets. If, by revisiting the placement decision tree, conflict with a previously established relation cannot be resolved after examining all possible decision points in decision placement tree, we then come to the conclusion that there is an error in the description. And even after all conflicts are resolved, there is no guarantee that the resulting sketch-map accurately represents what the narrator meant.

6 Conclusions and Future Work

The main goal of this paper was to show that a plausible topological sketch can be produced from text-based natural languages descriptions of places, such as a campus, without the use of additional knowledge sources. There are three processing steps in our approach. In the first step, textual descriptions are parsed for spatial relationships and the result is presented in an ordered list of binary spatial relations between objects, i.e., a list of triplets <L *r* RO>*i.* In the second step, objects in this list and the binary relations are stored in a spatial property graph. In the third and final step, the information contained in the property graph is used to dynamically construct a topological sketch, keeping track of the placement decisions made when the spatial relation is underspecified in the placement decision tree. We allow for the fact that some of the placement decisions need to be default solutions. Such default solutions may be inconsistent with reality. This is not critical to our goal, which is to determine if a plausible, rather than the actual topological layout, can be found from the NL descriptions.

We have shown in this paper that our approach to (semi-) automatically produce a topologically plausible sketch from natural language description is feasible, by applying the framework to an example (*Narrative A*) from the case study, and producing plausible sketches (Fig. 3-4). We also applied the algorithm to the full set of NL descriptions and could derive plausible topological sketches for all of them. We have given pseudo-algorithms to describe the necessary processing steps, where an automatic solution and thus implementation is possible. We clarified the rules we impose to arrive at a feasible solution. Such rules mostly stem from research in spatial cognition and psychology. We also identified where more research is needed to fully automate the process.

One of the strengths of the approach lies in the fact that we accept the underspecification of given spatial relations and provide a cognitively motivated default interpretation in order to sketch them. We do not force quantification on the qualitative descriptions, rather allowing for a re-interpretation of previously sketched relations in the dynamic sketching process by shrinking, enlarging or moving features as necessary. To achieve that, we populate the topological sketch by following the order with which people reveal the spatial information in their descriptions, exhausting all information for a part of the description, before moving to the next. This dynamic sketching process reveals how each person wanders in her own mental representation of the environment to produce a place description. We thus exploit and apply results from research in spatial cognition and natural language processing. By producing a list of requirements for what needs to be provided by a spatial parser, we can maybe guide research in NL processing of spatial descriptions.

Another strength of our approach is that we can extract a very simple structure of spatial information (ordered triplets) from unrestricted NL place descriptions. The simple structure <L *r* RO>*i* lends itself to a straightforward production of a property graph and a spatial topological sketch, by following the provided rules. An important characteristic of this structure is that it allows for different types of spatial relations and different styles of narrations to be accommodated together and to contribute synergistically to the production of a conflict free sketch.

In our current solution, the input of an intelligent agent (i.e., a person) is still needed at the parsing step. We do realize that we are building our process on a parser that is fairly sophisticated. However, our in-depth study on previous work in NL interpretation shows that an automatic parsing of spatial expressions as we require it, may not actually be that far in the future. Another point of consideration are the rules for sketching, i.e. the default interpretations of spatial relations. The rules we apply were motivated by research in spatial cognition. We will continue to check and refine these rules as more research becomes available.

In the future, apart from actually implementing the suggested algorithms and testing them on a different set of NL place descriptions, we want to enable processing of multiple descriptions of the same place. In our current approach the input is a single description. We have already discussed the resolution of place name synonyms (Section 3.1). The same approach, expanded to accommodate identifying vernacular synonyms from context, can be used to point to identical entities mentioned in different descriptions. This will allow for combining parser outputs into a single list of triplets that contain the extracted information from all descriptions. We are expecting that abundance and redundancy of the combined information will lead to more detailed positioning of elements on the topological sketch, coming closer to enabling an 'actual' representation of the described environment. By actual representation we mean a topological sketch that excludes some homomorphisms as not resembling the layout of the described environment.

So far, sketch evaluation is only about plausibility. Other evaluations can offer different insights to our approach. For example, a comparison of an output sketch with a map would allow to ground-truth our approach. Since ground truth was not an issue for the current research, we did not touch on the potential of using gazetteers and spatial databases for geocoding the known elements in a description and thus enabling alignment of sketches, but this is certainly something to consider for future versions of the algorithm.

Moreover, some additional information can be extracted from our sPGs. For example, the number of incoming edges of a single node in the graph can be considered as a measure of 'popularity' of the element, since it reveals the number of times the element was used as a reference object. Especially in the case where the descriptions of multiple people are combined, such popularity measure would reveal *landmark* candidates. Similarly, counting the number of leaves in the decision tree can be considered as a measure of the quality of the description, where higher number of decisions implies less specific descriptions that can lead to a higher number of plausible topological sketches. Finally, in the current framework, processing is restricted to buildings and street level [45]. Future extensions will deal with other granularity levels such as cities or rooms.

References

1. Sjoo, K., Pronobis, A., Jensfelt, P.: Functional Topological Relations for Qualitative Spatial Representation. In: Proceedings of 15th International Conference on Advanced Robotics (ICAR 2011). pp. 130 –136. IEEE, Tallinn, Estonia (2011).
2. Egenhofer, M.J.: Query Processing in Spatial-Query-by-Sketch. Journal of Visual Languages and Computing. 8, 403–424 (1997).
3. Tappan, D.: Knowledge-Based Spatial Reasoning for Scene Generation from Text Description. Proceedings of Association for the Advancement of Artificial Intelligence. , Chicago, IL (2008).
4. Denis, M.: The description of routes : A cognitive approach to the production of spatial discourse. Cahiers de psychologie cognitive. 16, 409–458 (1997).
5. Wang, J., Schwering, A.: The Accuracy of Sketched Spatial Relations: How Cognitive Errors Influence Sketch Representation. In: Tenbrink, T. and Winter, S. (eds.) International Workshop on Spatial Information: Granularity, Relevance, and Integration. In conjunction with COSIT 2009. pp. 40–47 (2009).
6. Tversky, B.: What do Sketches say about Thinking? AAAI Spring Symposium, Sketch Understanding Workshop, AAAI Technical Report, Stanford University (2002).
7. Tappan, D.A.: Knowledge-Based Spatial Reasoning for Automated Scene Generation from Text Descriptions, PhD Thesis, New Mexico State University, Las Cruces, New Mexico (2004).
8. Kordjamshidi, P., Van Otterlo, M., Moens, M.-F.: Spatial role labeling: Towards extraction of spatial relations from natural language. ACM Trans. Speech Lang. Process. 8, 4:1–4:36 (2011).

9. Kordjamshidi, P., Frasconi, P., Otterlo, M.V., Moens, M.-F., Raedt, L.D.: Relational Learning for Spatial Relation Extraction from Natural Language. In: Muggleton, S.H., Tamaddoni-Nezhad, A., and Lisi, F.A. (eds.) Inductive Logic Programming. pp. 204–220. Springer Berlin Heidelberg (2012).
10. Coyne, B., Sproat, R., Hirschberg, J.: Spatial Relations in Text-to-Scene Conversion. In: Ross, R.J., Hois, J., and Kelleher, J.D. (eds.) Proceedings of 1st Workshop on Computational Models of Spatial Language Interpretation (COSLI'10). pp. 9–16. , Mt Hood/Portland, OR, USA (2010).
11. Coyne, B., Rambow, O., Hirschberg, J., Sproat, R.: Frame semantics in text-to-scene generation. Proceedings of the 14th international conference on Knowledge-based and intelligent information and engineering systems: Part IV. pp. 375–384. Springer-Verlag, Berlin, Heidelberg (2010).
12. Wiebrock, S., Wittenburg, L., Schmid, U., Wysotzki, F.: Inference and Visualization of Spatial Relations. In: C. Freksa, C. Habel, W. Brauer, and K. Wender (eds.) Spatial Cognition II, LNAI 1849. pp. 212–224. Springer (2000).
13. Leidner, J.: Toponym Resolution in Text: Annotation, Evaluation and Applications of Spatial Grounding of Place Names, PhD Thesis, Institute for Communicating and Collaborative Systems, School of Informatics, University of Edinburgh (2007).
14. Egenhofer, M., Herring, J.: Categorizing Binary Topological Relations between Regions, Lines and Point in Geographic Databases. Department of Surveying Engineering, University of Maine, Orono (1990).
15. Randell, D.A., Cui, Z., Cohn, A.G.: A Spatial Logic based on Regions and Connection. Proceedings of 3rd International Conference on Principles of Knowledge, Representation and Reasoning (KR'92). pp. 165–176. Morgan Kaufmann (1992).
16. Frank, A.U.: Qualitative spatial reasoning: cardinal directions as an example. International journal of geographical information systems. 10, 269–290 (1996).
17. Hernández, D.: Relative Representation of Spatial Knowledge. In: Mark, D. and Frank, A. (eds.) Cognitive and Linguistic Aspects of Geographic Space. pp. 373–385. Kluwer Academic, Dordrecht (1991).
18. Kelleher, J.D.: A perceptually based computational framework for the interpretation of spatial language, (2003).
19. Li, H., Zhao, T., Li, S., Zhao, J.: The Extraction of Trajectories from Real Texts Based on Linear Classification. In: Joakim Nivre, Heiki-Jaan Kaalep, Kadri Muischnek, and Mare Koit (eds.) Proceedings of the 16th Nordic Conference of Computational Linguistics (NODALIDA-2007). pp. 121–127. , University of Tartu, Tartu (2007).
20. Bateman, J.A., Hois, J., Ross, R., Tenbrink, T.: A linguistic ontology of space for natural language processing. Artificial Intelligence. 174, 1027–1071 (2010).
21. Bateman, J.A.: Language and Space: a two-level semantic approach based on principles of ontological engineering. Int J Speech Technol. 13, 29–48 (2010).
22. Bateman, J., Tenbrink, T., Farrar, S.: The Role of Conceptual and Linguistic Ontologies in Interpreting Spatial Discourse. Discourse Processes: A Multidisciplinary Journal. 44, 175–212 (2007).

23. Kordjamshidi, P., Van Otterlo, M., Moens, M.-F.: From language towards formal spatial calculi. In: Ross, R.J., Hois, Joana, and Kelleher, J.D. (eds.) Proceedings of 1st Workshop on Computational Models of Spatial Language Interpretation (COSLI'10). pp. 17–24. , Mt.Hood/Portland, OR, USA (2010).
24. Zlatev, J.: Spatial Semantics. In: Hubert Cuyckens and Dirk Geeraerts (eds.) The Oxford Handbook of Cognitive Linguistics. pp. 318–350. Oxford University Press (2007).
25. Litkowski, K., Hargraves, O.: SemEval-2007 Task 06: Word-Sense Disambiguation of Prepositions. Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval). pp. 24–29. Association for Computational Linguistics (2007).
26. Fellbaum, C. ed: WordNet: An Electronic Lexical Database. MIT Press, Cambridge, MA (1998).
27. Baker, C., Fillmore, C., Lowe, J.: The Berkeley FrameNet Project. COLING-ACL (1998).
28. Hidders, J.: Typing Graph-Manipulation Operations. Proceedings of the 9th International Conference on Database Theory. pp. 394–409. Springer-Verlag, London, UK, UK (2002).
29. Güting, R.H.: GraphDB: Modeling and Querying Graphs in Databases. Proceedings of the 20th International Conference on Very Large Databases (VLDB). pp. 297–308. , Santiago (1994).
30. Kuper, G.M., Vardi, M.Y.: A new approach to database logic. Proceedings of the 3rd ACM SIGACT-SIGMOD symposium on Principles of Database Systems. pp. 86–96. ACM, New York, NY, USA (1984).
31. Rodriguez, M.A., Neubauer, P.: Constructions from dots and lines. Bulletin of the American Society for Information Science and Technology. 36, 35–41 (2010).
32. Mainguenaud, M.: Modelling the network component of geographical information systems. International journal of geographical information systems. 9, 575–593 (1995).
33. Freeman, L.C.: The Development Of Social Network Analysis: A Study In The Sociology Of Science. Empirical Press (2004).
34. Tversky, B.: Cognitive maps, cognitive collages, and spatial mental models. Presented at the Spatial Information Theory: A Theoretical Basis for GIS (1993).
35. Downs, R.M., Stea, D.: Cognitive Maps and Spatial Behaviour: Process and Products. In: Dodge, rtin, Kitchin, R., and Perkins, C. (eds.) The Map Reader. pp. 312–317. John Wiley & Sons, Ltd (2011).
36. Gould, P., White, R.: Mental Maps. Routledge, New York.
37. Lynch, K.: The Image of the City. MIT Press (1960).
38. Tversky, B.: Visualizing Thought. Topics in Cognitive Science. 3, 499–535 (2011).
39. Freksa, C.: Using Orientation Information for Qualitative Spatial Reasoning. In: Frank, A., Campari, I., and Formentini, U. (eds.) Theories and Methods of Spatio-Temporal Reasoning in Geographic Space, LNCS. pp. 162–178. Springer-Verlag (1992).

40. Hirtle, S., Jonides, J.: Evidence of hierarchies in cognitive maps. Memory & Cognition. 13, 208–217 (1985).
41. McNamara, T.P., Hardy, J.K., Hirtle, S.C.: Subjective hierarchies in spatial memory. Journal of Experimental Psychology: Learning, Memory and Cognition. 15, 211–227 (1989).
42. Meneghetti, C., Pazzaglia, F., De Beni, R.: The Mental Representation Derived From Spatial Descriptions Is North-Up Oriented: the Role of Visuo-Spatial Abilities. LNAI 7463. 1–17 (2012).
43. Agrawala, M., Stolte, C.: Rendering effective route maps: improving usability through generalization. SIGGRAPH '01: Proceedings of the 28th annual conference on Computer graphics and interactive techniques. ACM Request Permissions (2001).
44. Fraczak, L.: Generating "mental maps" from route descriptions. In: Olivier, P. and Gapp, K.-P. (eds.) Representation and Processing of Spatial Expressions. pp. 185–200. L. Erlbaum Associates Inc, Mahwah, New Jersey (1998).
45. Richter, D., Vasardani, M., Stirling, L., Richter, K.-F., Winter, S.: Zooming In - Zooming Out: Hierarchies in Place Descriptions. Proceedings of the 9th Symposium on Location Based Services (LBS 2012). , Munich, Germany (2012).
46. Brunyé, T.T., Mahoney, C.R., Taylor, H.A.: Moving through imagined space: Mentally simulating locomotion during spatial description reading. Acta Psychologica. 134, 110–124 (2010).
47. Grice, P.: Logic and Conversation. Syntax and Semantics. 3, 41–58 (1975).
48. Meilinger, T.: The Network of Reference Frames Theory: A Synthesis of Graphs and Cognitive Maps. In: Freksa, C., Newcombe, N.S., Gärdenfors, P., and Wölfl, S. (eds.) Proceedings of the international conference on Spatial Cognition VI: Learning, Reasoning, and Talking about Space. pp. 344–360. Springer-Verlag, Berlin, Heidelberg (2008).
49. Wallgrün, J.O.: Qualitative spatial reasoning for topological map learning. Spatial Cognition & Computation. (2010).
50. Barsalou, L.W.: Grounded Cognition. Annual Review of Psychology. 59, 617–645 (2008).
51. Johnson-Laird, P.N., Byrne, R.M.J.: Conditionals: A theory of meaning, pragmatics, and inference. Psychological review. 109, 646–678 (2002).