SpatialNet: A Declarative Resource for Spatial Relations

Morgan Ulinski and **Bob Coyne** and **Julia Hirschberg**
Department of Computer Science
Columbia University
New York, NY, USA
{mulinski,coyne,julia}@cs.columbia.edu

Abstract

This paper introduces SpatialNet, a novel resource which links linguistic expressions to actual spatial configurations. SpatialNet is based on FrameNet (Ruppenhofer et al., 2016) and VigNet (Coyne et al., 2011), two resources which use frame semantics to encode lexical meaning. SpatialNet uses a deep semantic representation of spatial relations to provide a formal description of how a language expresses spatial information. This formal representation of the lexical semantics of spatial language also provides a consistent way to represent spatial meaning across multiple languages. In this paper, we describe the structure of SpatialNet, with examples from English and German. We also show how SpatialNet can be combined with other existing NLP tools to create a text-to-scene system for a language.

1 Introduction

Spatial language understanding is a research area in NLP with applications from robotics and navigation to paraphrase and image caption generation. However, most work in this area has been focused specifically on English. While there is a rich literature on the realization of spatial relations in different languages, there is no comprehensive resource which can represent spatial meaning in a formal manner for multiple languages. The development of formal models for the expression of spatial relations in different languages is a largely uninvestigated but relevant problem.

By way of motivation, consider the following translation examples. We use an NP in which we have a PP modifier, and we complete the sentence with a copula and an adjective to obtain a full sentence. The prepositions are marked in **boldface**. The English sentence is a word-for-word gloss of the German sentence except for the preposition.

In our first example, English *on* is correctly translated to German *an*:[1]

(1) a. The painting **on** the wall is abstract.
b. Correct translation: Das Gemälde **an** der Mauer/Wand ist abstrakt.
c. Google Translate/Bing Translator (correct): Das Gemälde **an** der Wand ist abstrakt.

However, the correct translation changes if we are relating a cat to a wall:

(2) a. The cat **on** the wall is grey.
b. Correct translation: Die Katze **auf** der Mauer ist grau.
c. Google Translate/Bing Translator (incorrect): Die Katze **an** der Wand ist grau.

The problem here is that the English preposition *on* describes two different spatial configurations: 'affixed to', in the case of the painting, and 'on top of', in the case of the cat.[2]

Similar problems appear when we translate from German to English. The painting again translates correctly:

(3) a. Das Gemälde **an** der Mauer ist abstrakt.
b. Correct translation: The painting **on** the wall is abstract.
c. Google Translate/Bing Translator (correct): The painting **on** the wall is abstract.

[1] Note that English *wall* should be translated to *Wand* if it is a wall which has a ceiling attached to it, and *Mauer* if it is freestanding and does not help create an enclosed three-dimensional space. We ignore this particular issue in this discussion.

[2] We set aside the interpretation in which the cat is affixed to the wall similarly to a clock, which is an extraordinary interpretation and would require additional description in either language.

Proceedings of SpLU-RoboNLP, pages 61–70
Minneapolis, MN, June 6, 2019.

But when we replace the painting with the house, we no longer obtain the correct translation:

(4) a. Das Haus **an** der Mauer ist groß.
b. Correct translation: The house **at** the wall is large/big.
c. Google Translate (incorrect): The house **on** the wall is large.
Bing Translator (incorrect): The house **on** the wall is big.

The problem is again that the German preposition *an* corresponds to two different spatial configurations, 'affixed to' (painting) and 'at/near' (house).

We address the issue of modeling cross-linguistic differences in the expression of spatial language by developing a deep semantic representation of spatial relations called *SpatialNet*. SpatialNet is based on two existing resources: FrameNet (Baker et al., 1998; Ruppenhofer et al., 2016), a lexical database linking semantic frames to manually annotated text, and VigNet (Coyne et al., 2011), a resource extending FrameNet by grounding abstract lexical semantics with concrete graphical relations. VigNet was developed as part of the WordsEye text-to-scene system (Coyne and Sproat, 2001). SpatialNet builds on both these resources to provide a formal description of the lexical semantics of spatial relations by linking linguistic expressions both to semantic frames and to actual spatial configurations. Because of the link to VigNet and WordsEye, SpatialNet can also be used to create a text-to-scene system for a language. This text-to-scene system can be used to verify the accuracy of a SpatialNet resource with native speakers of a language.

SpatialNet is divided into two modules: *Spatiographic primitives* (SGPs) represent possible graphical (spatial) relations. The *ontology* represents physical objects and their classification into semantic categories. Both are based on physical properties of the world and do not depend on a particular language. *Spatial frames* are language-specific (though, like the frames of FrameNet, may be shared among many languages) and represent the lexical meanings a language expresses. *Spatial vignettes* group together lexical items, spatial frames, and SGPs with spatial and graphical constraints from the ontology, grounding the meaning in a language-independent manner.

In Section 2, we discuss related work. In Section 3, we provide background information on FrameNet and VigNet. In Section 4, we describe the SpatialNet structure, with English and German examples. In Section 5, we show how the SpatialNet for a language can be used in conjunction with the WordsEye text-to-scene system to generate 3D scenes from input text in that language. We conclude in Section 6 and discuss future work.

2 Related Work

Spatial relations have been studied in linguistics for many years. One study for English by Herskovits (1986) catalogs fine-grained distinctions in the interpretation of prepositions. For example, she distinguishes among the uses of *on* to mean 'on the top of a horizontal surface' (*the cup is on the table*) or 'affixed to a vertical surface' (*the picture is on the wall*). Likewise, Feist and Gentner (1998) describe user perception experiments that show that the shape, function, and animacy of the figure and ground objects are factors in the perception of spatial relations as *in* or *on*.

Other work looks at how the expression of spatial relations varies across languages. Bowerman and Choi (2003) describe how Korean linguistically differentiates between putting something in a loose-fitting container (*nehta*, e.g. fruit in a bag) vs. in a tight fitting wrapper (*kkita*, e.g. hand in glove). Other languages (English included) do not make this distinction. Levinson (2003) and colleagues have also catalogued profound differences in the ways different languages encode relations between objects in the world. Our work differs from linguistic efforts such as these in that we are building a formal representation of how a language expresses spatial information, which can be applied to a variety of NLP problems and applications. Since the representation is human- as well as machine-readable, it can also be used in more traditional linguistics.

Another area of research focuses on computational processing of spatial language. Pustejovsky (2017) has developed an annotation scheme for labeling text with spatial roles. This type of annotation can be used to train classifiers to automatically perform the task, as demonstrated by the SpaceEval task (Pustejovsky et al., 2015). Although this work provides examples of how a language expresses spatial relations, annotation of spatial roles does not provide a formal description of the link between surface realization and underlying semantics. Our work provides a formal de-

scription and also a semantic grounding that tells us the actual spatial configuration denoted by a set of spatial roles. Also, our work extends to languages other than English.

Petruck and Ellsworth (2018) advocate using FrameNet (Ruppenhofer et al., 2016) to represent spatial language. FrameNet uses frame semantics to encode lexical meaning. VigNet (Coyne et al., 2011) is an extension of FrameNet used in the WordsEye text-to-scene system (Coyne and Sproat, 2001). SpatialNet builds on both FrameNet and VigNet; we will describe FrameNet and VigNet in more detail in the next section.

3 Background on FrameNet and VigNet

FrameNet encodes lexical meaning using a frame-semantic conceptual framework. In FrameNet, lexical items are grouped together in *frames* according to shared semantic structure. Every frame contains a number of *frame elements* (semantic roles) which are participants in this structure. Words that evoke a frame are called *lexical units*. A lexical unit is also linked to sentences that have been manually annotated to identify frame element fillers and their grammatical functions. This results in a set of *valence patterns* that represent possible mappings between syntactic functions and frame elements for the lexical unit. FrameNet already contains a number of frames for spatial language. Spatial language frames in FrameNet inherit from LOCATIVE-RELATION, which defines core frame elements FIGURE and GROUND, as well as non-core frame elements including DISTANCE and DIRECTION. Examples of spatial language frames are SPATIAL-CONTACT, CONTAINMENT and ADJACENCY.

VigNet, a lexical resource inspired by and based on FrameNet, was developed as part of the WordsEye text-to-scene system. VigNet extends FrameNet in several ways. It adds much more fine-grained frames, primarily based on differences in graphical realization. For example, the verb "wash" can be realized in many different ways, depending on whether one is washing dishes or one's hair or a car; VigNet therefore has several different wash frames. VigNet also adds graphical semantics to frames. It does this by adding primitive graphical (typically, spatial) relations between frame element fillers. These graphical relations can represent the position, orientation, size, color, texture, and poses of objects in the scene. The graphical semantics can be thought of as a semantic grounding; it is used by WordsEye to construct and render a 3D scene. Frames augmented with graphical semantics are called *vignettes*.

The descriptions of the graphical semantics in vignettes make use of object-centric properties called *affordances* (Gibson, 1977; Norman, 1988). Affordances include any functional or physical property that allows an object to participate in actions and relations with other objects. For example, a SEAT of a chair is used to support a sitter and the INTERIOR of a box is used to hold the contents. VigNet has a rich set of spatial affordances. Some examples are CUPPED REGIONS for objects to be *in*, CANOPIES for objects to be *under*, and TOP SURFACES for objects to be *on*.

Information about the 3D objects in WordsEye is organized in VigNet into an *ontology*. The ontology is a hierarchy of semantic types with multiple inheritance. Types include both 3D objects and more general semantic concepts. For example, a particular 3D rocking chair is a sub-type of ROCKING-CHAIR.N. Every 3D object has a semantic type and is inserted into the ontology. WordsEye also includes lexicalized concepts (e.g. *chair* tied to CHAIR.N) in the ontology. The ontology includes a knowledge base of assertions that provide more information about semantic concepts. Assertions include sizes of objects and concepts, their parts, their colors, what they typically contain, what affordances they have, and information about their function. Spatial affordances and other properties can be applied to both 3D graphical objects and to more general semantic types. For example, the general semantic type CUP.N has a CUPPED REGION affordance, since this affordance is shared by all cups. A particular 3D graphical object of a cup might have a HANDLE affordance, while another might have a LID affordance, but these spatial affordances are not tied to the super-type CUP.N.

Figure 1 shows an example of two vignettes: SELF-MOTION-FROM-FRONT.R and SELF-MOTION-FROM-PORTAL.R. Both are subtypes of SELF-MOTION-FROM.R. The yellow ovals contain semantic constraints on the objects used to instantiate the frame. For example, while the relation SELF-MOTION-FROM-FRONT.R requires only that the source of the motion be a PHYSICAL-ENTITY.N, SELF-MOTION-FROM-PORTAL.R requires that the source has a

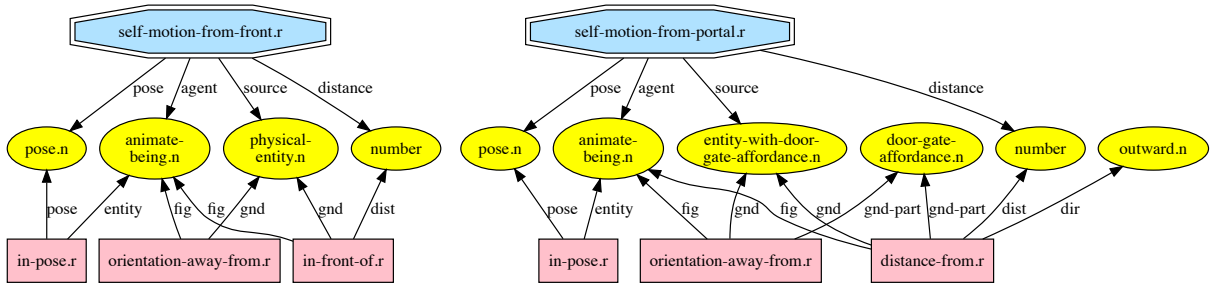


Figure 1: Two frames augmented with primitive graphical relations. The high-level semantics of SELF-MOTION-FROM-FRONT.R and SELF-MOTION-FROM-PORTAL.R are decomposed into semantic types and graphical relations.

DOOR-GATE-AFFORDANCE.N as a part.

4 Structure of SpatialNet

SpatialNet provides a formal description of spatial semantics by linking linguistic expressions to semantic frames and linking semantic frames to actual spatial configurations. To do this, we adopt some conventions from FrameNet and VigNet, making some changes to address some of the shortcomings of these resources.

FrameNet provides semantic frames including frames for spatial language. However, the syntactic information provided in the valence patterns is often insufficient for the purpose of automatically identifying frame elements in new sentences. One example is frames where the target word is a preposition, which includes many of the frames for spatial language. According to the FrameNet annotation guidelines for these (Ruppenhofer et al., 2016, page 50), the GROUND is assigned the grammatical function Obj(ect), and the FIGURE is tagged as an Ext(ernal) argument. Given a previously unseen sentence, automatic methods can identify the object of the preposition and therefore the GROUND, but the sentence may contain several noun phrases outside the prepositional phrase, making the choice of FIGURE ambiguous. FrameNet also does not provide a semantic grounding. To create SpatialNet, we adopt the concept of a FrameNet *frame*, including the definition of *frame elements* and *lexical units*. However, we modify the valence patterns to more precisely define syntactic patterns in a declarative format. In addition, to facilitating the use of SpatialNet across different languages, we specify syntactic constraints in valence patterns using labels from the Universal Dependencies project (Universal Dependencies, 2017).

VigNet does provide a grounding in graphical semantics, but presents other problems. First, VigNet does not currently include a mapping from syntax to semantic frames. Although vignettes provide a framework for linking semantic frames to primitive graphical relations, the VigNet resource does not include frames for spatial prepositions, but only for higher-level semantic constructs. Finally, since VigNet has been developed specifically for English, some parts of the existing resource do not generalize easily to other languages. To create SpatialNet, we adopt from VigNet the concept of a *vignette* and the *semantic ontology*. However, we make the resource more applicable across languages by (a) formalizing the set of primitive graphical relations and constraints used in vignettes into what we call *spatio-graphic primitives* (SGPs), and (b) moving the language-specific mapping of lexical items to semantic categories out of the VigNet ontology and into a separate database. The SGPs and semantic ontology are used to define a language-independent semantic grounding for vignettes.

A SpatialNet for a particular language consists of a set of *spatial frames*, which link surface language to lexical semantics using valence patterns, and a set of *spatial vignettes*, which link spatial frames and lexical units to SGPs based on semantic/functional constraints. We are developing SpatialNet resources for English and German.

4.1 Ontology of Semantic Categories

The ontology in VigNet consists of a hierarchy of semantic types (concepts) and a knowledge base containing assertions. SpatialNet uses the VigNet ontology and semantic concepts directly, under the assumption that the semantic types and assertions are language-independent. Thus far, our work on English and German has not required modification of the ontology; however, since it was de-

veloped for English, it may need to be extended or modified in the future to be relevant for other languages and cultures. VigNet also includes lexicalized concepts (e.g. *chair* tied to CHAIR.N) in the ontology. For SpatialNet, we store this language-dependent lexical information in a separate database.

The mapping from lexical items to semantic concepts is important for the decomposition of text into semantics. For English SpatialNet, we use the lexical mapping extracted from VigNet. To facilitate creation of lexical mappings for other languages, we mapped VigNet concepts to entries in the Princeton WordNet of English (Princeton University, 2019). An initial mapping was constructed as follows: For each lexicalized concept in VigNet, we looked up each of its linked lexical items in WordNet. If the word (with correct part of speech) was found in WordNet, we added mappings between the VigNet concept and each WordNet synset for that word. This resulted in a many-to-many mapping of VigNet concepts to WordNet synsets. We are currently working on manually correcting this automatically-created map.

To obtain a lexical mapping for German, we use the VigNet–WordNet map in conjunction with GermaNet (Henrich and Hinrichs, 2010; Hamp and Feldweg, 1997). GermaNet includes mappings to Princeton WordNet 3.0. For a given German lexical item, we use the GermaNet links to Princeton WordNet to obtain a set of possible VigNet concepts from the VigNet–WordNet mapping. We are also experimenting with the Open German WordNet (Siegel, 2019), although in general we have found it to be less accurate. Open German WordNet includes links to the EuroWordNet Interlingual Index (ILI) (Vossen, 1998), which are in turn mapped to the Princeton English WordNet. Table 1 shows the VigNet concepts for German words used in the sentences in Figure 2, obtained using GermaNet and Open German WordNet.

Lexical item	**VigNet concepts**	
	GermaNet	**ODE-WordNet**
Mauer	WALL.N RAMPART-WALL.N RAMPART.N	WALL.N
Katze	DOMESTIC-CAT.N HOUSE-CAT.N	DOMESTIC-CAT.N HOUSE-CAT.N TRUE-CAT.N
Gemälde	PAINTING.N PICTURE.N	PICTURE.N ICON.N IMAGE.N
Haus	HOUSE.N	SHACK.N HUTCH.N HOUSE.N FAMILY.N HOME.N

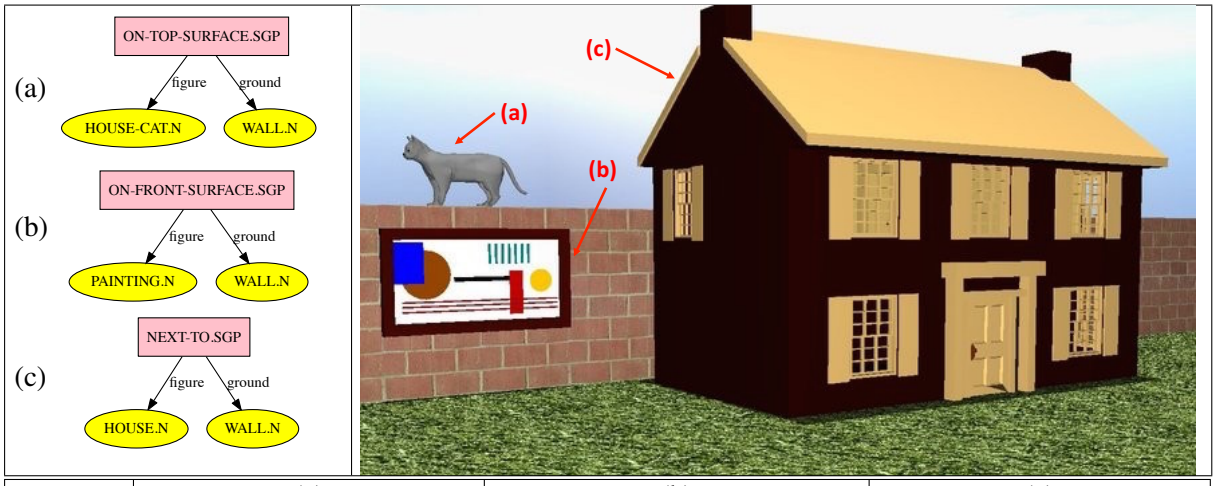
Table 1: Mapping from German lexical items to VigNet semantic categories, obtained using two different German WordNet resources.

4.2 Spatio-graphic Primitives

To create the set of spatio-graphic primitives used in SpatialNet, we began with relations already in VigNet. VigNet contains a range of semantic relations, from high-level abstract relations originating in FrameNet, such as ABANDONMENT.R, to low-level graphical relations, such as RGB-VALUE-OF.R. We extracted from VigNet a list of relations representing basic spatial configurations and graphical properties, separating these from the higher-level relations in VigNet which may be English-specific.

We also wanted to ensure that our list of spatio-graphic primitives was as comprehensive as possible, and not limited to the graphical capabilities of WordsEye. To that end, we annotated each picture in the Topological Relations Picture Series (Bowerman and Pederson, 1992) and the Picture Series for Positional Verbs (Ameka et al., 1999) with the spatial and graphical primitives it represents. When an appropriate spatial primitive did not exist in VigNet, we created a new one. These new SGPs have also been added it to a list of "pending" graphical relations that the WordsEye developers plan to implement in the future. In total, we have about 100 SGPs.

We use WordsEye as a realization engine for the SGPs. This is done using the WordsEye web API, which can generate a 3D scene from a semantic representation. The semantic representation consists of a list of entities, each with a semantic type from the VigNet ontology, and a list of relations between entities. SpatialNet SGPs can be used as relations in this semantic input; we are working closely with the WordsEye developers to ensure that SGPs in SpatialNet continue to be compatible with the WordsEye system. In some cases, graphical functionality for an



	(a)	(b)	(c)
English	*The cat is **on** the wall.*	*The painting is **on** the wall.*	*The house is **at** the wall.*
German	*Die Katze ist **auf** der Mauer.*	*Das Gemälde ist **an** der Mauer.*	*Das Haus ist **an** der Mauer.*

Figure 2: Examples of spatio-graphic primitives: (a) ON-TOP-SURFACE, (b) ON-FRONT-SURFACE, and (c) NEXT-TO and English/German descriptions.

SGP is not yet supported by WordsEye. For example, WordsEye currently cannot graphically represent a FITTED-ON relation, e.g. a hat on a head or a glove on a hand. When WordsEye encounters a relation that it cannot decompose into supported graphical primitives, the relation is ignored and not included in the 3D graphics. The entities referenced by the relations will be displayed in a default position (side-by-side). Figure 2 shows a scene created in WordsEye that demonstrates the spatio-graphic primitives ON-TOP-SURFACE, ON-FRONT-SURFACE, and NEXT-TO.

4.3 Spatial Frames

Spatial frames represent the lexical meanings a language can express. The structure of spatial frames is closely based on FrameNet frames. We have incorporated many of the FrameNet spatial language frames into SpatialNet, adding to these as needed. For example, for English we have added an ON-SURFACE frame that inherits from SPATIAL-CONTACT. The main difference between SpatialNet frames and FrameNet frames is in the definition of the valence patterns. SpatialNet defines valence patterns by precisely specifying lexical and syntactic constraints, which can be based on the syntactic dependency tree structure, grammatical relations, parts of speech, or lexical items. Figure 4, which provides examples of spatial vignettes, includes a valence pattern for the English lexical unit *on.adp*. This pattern specifies a syntactic structure consisting of a root (which must have part of speech `NOUN`), an `nsubj` dependent, and a `case` dependent (which must be the word "on"). The declarative format used to define this spatial frame is shown in Figure 3 (top).

4.4 Spatial Vignettes

Spatial vignettes use spatial frames, SGPs, and the ontology to interpret prepositions and other lexical information in a language. They relate linguistic realization (e.g. a preposition with its argument structure) to a spatial frame (such as ON-SURFACE), and at the same time to a graphical semantics expressed in terms of SGPs and additional constraints. This lexical information is often ambiguous. Consider the English and German descriptions in Figure 2. In English, the preposition *on* is ambiguous; it can mean either ON-TOP-SURFACE or ON-FRONT-SURFACE. In German, the preposition *an* is ambiguous; it can mean either ON-FRONT-SURFACE or NEXT-TO. To resolve such ambiguities, vignettes place selectional restrictions on frame elements that require fillers to have particular spatial affordances, spatial properties (such as the object size, shape, and orientation), or functional properties (such as whether the object is a vehicle or path). This information is found in the ontology.

Consider the spatial vignettes that would be used to disambiguate the meanings of English *on* from Figure 2. The declarative format used to de-

```
<frame name="On_surface">
  <parent name="Spatial_contact"/>
  <FE name="Figure"/>
  <FE name="Ground"/>
  <lexUnit name="on_top_of.adp">
    <pattern>
      <dep FE="Ground" tag="NOUN">
        <dep FE="Figure" reln="nsubj"/>
        <dep reln="case" word="on">
          <dep word="top" reln="mwe"/>
          <dep word="of" reln="mwe"/>
        </dep>
      </dep>
    </pattern>
  </lexUnit>

  <lexUnit name="on.adp">
    <pattern>
      <dep FE="Ground" tag="NOUN">
        <dep FE="Figure" reln="nsubj"/>
        <dep reln="case" word="on"/>
      </dep>
    </pattern>
  </lexUnit>
</frame>
```

```
<vignette name="on-vertical-surface">
  <input frame="On_surface"
      lexUnit="on.adp"/>
  <type-constraint FE="Ground"
      type="vertical-surface.n"/>
  <type-constraint FE="Figure"
      type="wall-item.n"/>
  <output relation="on-front-surface.r">
    <map FE="Ground" arg="ground"/>
    <map FE="Figure" arg="figure"/>
  </output>
</vignette>

<vignette name="on-top-surface">
  <input frame="On_surface"
      lexUnit="on.adp"/>
  <input frame="On_surface"
      lexUnit="on_top_of.adp"/>
  <type-constraint FE="Ground"
      type="upward-surface.n"/>
  <output relation="on-top-surface.r">
    <map FE="Ground" arg="ground"/>
    <map FE="Figure" arg="figure"/>
  </output>
</vignette>
```

Figure 3: Declarative format for spatial frames (top) and spatial vignettes (bottom)

fine these spatial vignettes is shown in Figure 3 (bottom). A visual representation of the vignettes is shown in Figure 4 (top). The vignettes link the spatial frame ON-SURFACE to different SGPs based on features of the frame element fillers.

The first vignette, ON-FRONT-SURFACE, adds semantic type constraints to both the FIGURE and the GROUND. The Figure must be of type WALL-ITEM.N and the Ground must be of type VERTICAL-SURFACE.N. If these constraints are met, the vignette produces the SGP ON-FRONT-SURFACE as output, mapping FIGURE to the SGP argument `figure`, and GROUND to the SGP argument `ground`. The second vignette, ON-TOP-SURFACE, has a semantic type constraint only that GROUND be of type UPWARD-SURFACE.N. If this constraint is met, the vignette produces the SGP ON-TOP-SURFACE. Note that, while in this case the frame elements and SGP arguments have the same names, this is not necessarily true for all vignettes (cf. the vignettes in Figure 1). Note also that in English, *painting **on** wall* is actually ambiguous, since a painting can technically be balanced on the top of a wall rather than hanging on its front surface. The spatial vignettes allow for either interpretation.

Figure 4 also shows the two vignettes which would be used to disambiguate the meanings of German *an* from Figure 2. The German vignettes link the spatial frame ADJACENCY to SGPs. The first vignette, ON-FRONT-SURFACE, is identical to the English vignette of the same name, except for the input frame and lexical unit. The semantic type constraints, SGPs, and frame element to SGP argument mappings are the same. The second vignette, NEXT-TO, does not have any semantic type constraints and thus outputs the SGP NEXT-TO with the familiar FIGURE–`figure` and GROUND–`ground` argument mappings. In the next section, we provide a complete example of using spatial vignettes to interpret these German sentences.

5 Using SpatialNet for Text-to-Scene Generation

SpatialNet can be used in conjunction with the graphics generation component of the WordsEye text-to-scene system to produce a 3D scene from a spatial description which can be used to verify the spatial frames and vignettes defined in SpatialNet. Figure 5 shows an overview of our system for text-to-scene generation. Although SpatialNet focuses on semantics, the system also requires modules for morphological analysis and syntactic parsing. For English and German, we use the Stanford CoreNLP Toolkit (Manning et al., 2014). In this section, we describe how we use Stanford CoreNLP, SpatialNet, and WordsEye to convert text into a 3D scene. We illustrate using German sentences (b) and (c) from Figure 2.

First, Stanford CoreNLP is used to perform

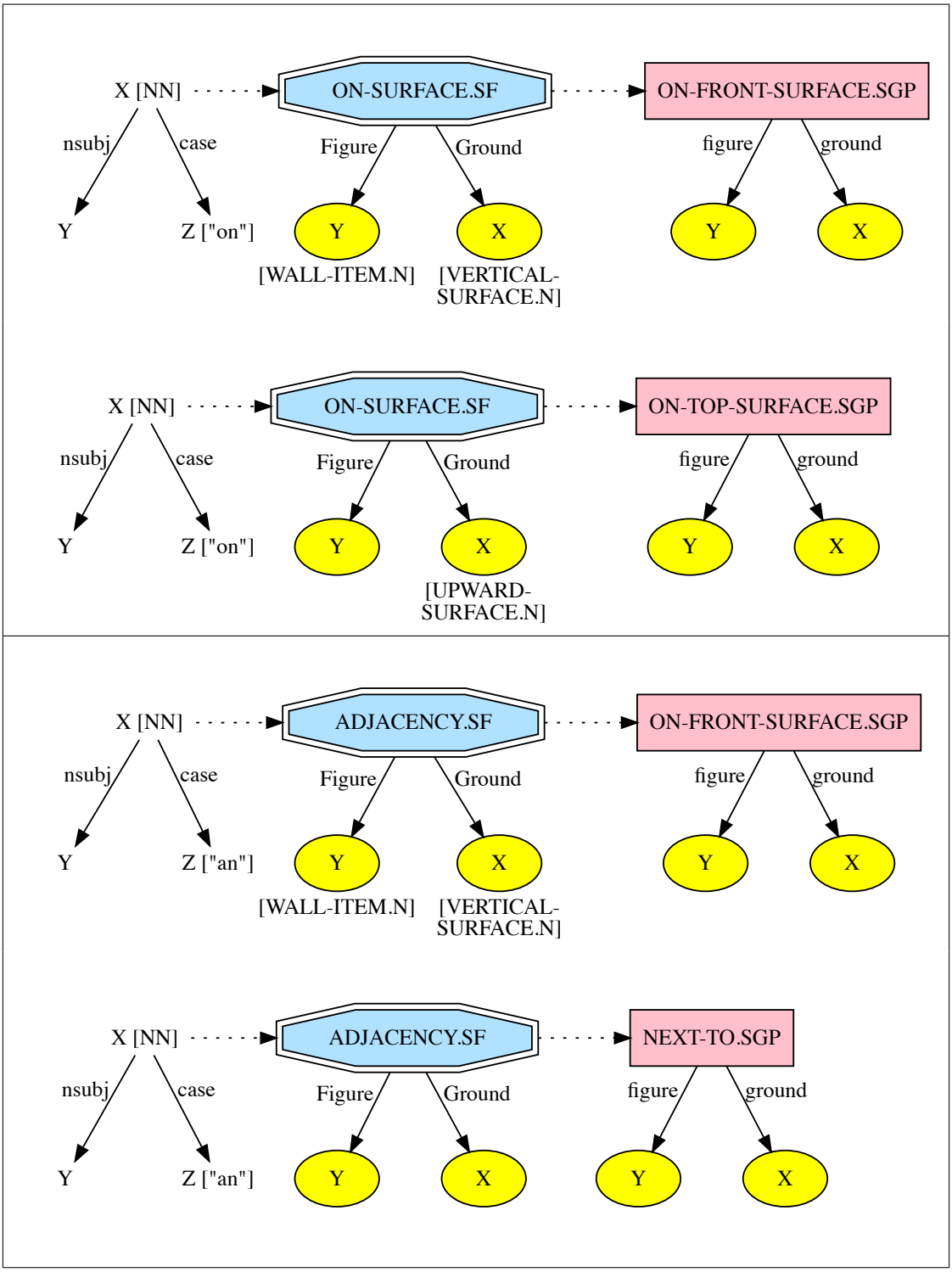


Figure 4: Spatial vignettes for different meanings of English *on* (top) and German *an* (bottom). Vignettes resolve the spatial relation given the spatial and functional object features. Spatial frames are represented by blue octagons, and SGPs by pink rectangles.

lemmatization, part-of-speech tagging, and dependency parsing. Figure 6 shows the resulting dependency structures. The dependency structures are matched against the valence patterns in spatial frames. Sentences (b) and (c) both match the valence pattern for the lexical unit *an.prep* in the ADJACENCY frame. The valence pattern identifies which lexical items in the sentence will act as frame element fillers. These lexical items are converted into semantic concepts using the lexical

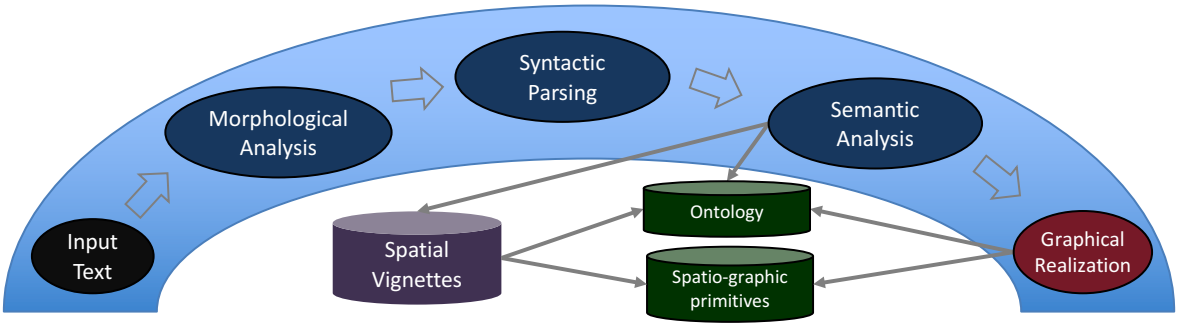
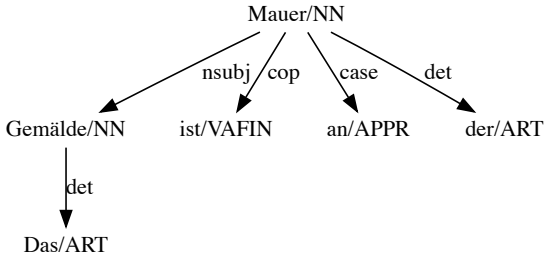
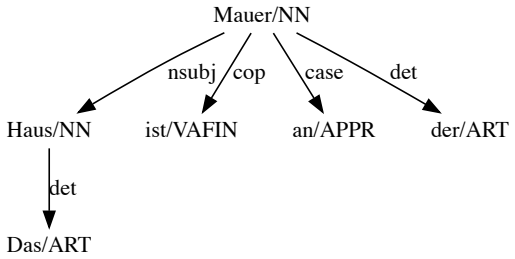


Figure 5: Pipeline for text-to-scene generation with SpatialNet



(b)



(c)

Figure 6: Results of morphological and syntactic analysis for German sentences (b) and (c)

mapping from Section 4.1. We refer to Table 1 to obtain the semantic concepts for the German lexical items. For the purposes of this example, we select the first semantic concept from the GermaNet mapping, which maps *Gemälde* to PAINTING.N, *Mauer* to WALL.N, and *Haus* to HOUSE.N.

The system then identifies the spatial vignettes which accept the frame and lexical unit as input. The features of the semantic concepts obtained for each frame element are checked against the semantic constraints in these spatial vignettes. For German sentence (b), since a WALL.N has a VERTICAL-SURFACE and a PAINTING.N is a WALL-ITEM, the ON-FRONT-SURFACE vignette is a possible match. Since a WALL.N also has an UPWARD-SURFACE, the ON-TOP-SURFACE vignette is also a possible match. For now, we select the first matching vignette, which produces the SGP ON-FRONT-SURFACE with `figure` PAINTING.N and `ground` WALL.N. For German sentence (c), since HOUSE.N is not a WALL-ITEM, only the NEXT-TO vignette is matched. This produces the SGP NEXT-TO, with `figure` HOUSE.N and `ground` WALL.N. The entities and SGPs for each sentence are then converted into a semantic representation compatible with the WordsEye web API, which is used to generate a 3D scene.

6 Summary and Future Work

We have described our development of a novel resource, SpatialNet, which provides a formal representation of how a language expresses spatial relations. We have discussed the structure of the resource, including examples from the English and German SpatialNets we are developing. We have also introduced a text-to-scene generation pipeline for using SpatialNet to convert text into 3D scenes.

In future, we will extend our semantic representation to handle motion as well as static spatial relations. A *motion vignette* could be represented by a labeled sequence of SGPs associated with key stages of the action, e.g. INITIAL-STATE, START-OF-ACTION, MIDDLE-STATE, END-OF-ACTION, FINAL-STATE. For example, *The dog jumped off the log* could be represented by the dog standing on the log, the dog leaping off with legs still on the log, the dog in mid air, the front paws touching the ground, and the dog on the ground.

In addition, we hope to extend SpatialNet to other languages, particularly low-resource and endangered languages, by incorporating it into the WordsEye Linguistics Tools (Ulinski et al., 2014a,b).

References

Felix Ameka, Carlien de Witte, and David P. Wilkins. 1999. Picture series for positional verbs: Eliciting the verbal component in locative descriptions. In David P. Wilkins, editor, *Manual for the 1999 Field Season*, pages 48–54. Max Planck Institute for Psycholinguistics, Nijmegen.

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet Project. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, Volume 1*, pages 86–90, Montreal, Quebec, Canada.

Melissa Bowerman and Soonja Choi. 2003. Space under Construction: Language-Specific Spatial Categorization in First Language Acquisition. In *Language in Mind: Advances in the Study of Language and Thought*, pages 387–428. MIT Press, Cambridge, MA, US.

Melissa Bowerman and Eric Pederson. 1992. Topological relations picture series. In Stephen C. Levinson, editor, *Space Stimuli Kit 1.2*, volume 51. Max Planck Institute for Psycholinguistics, Nijmegen.

Bob Coyne, Daniel Bauer, and Owen Rambow. 2011. VigNet: Grounding Language in Graphics using Frame Semantics. In *Proceedings of the ACL 2011 Workshop on Relational Models of Semantics*, pages 28–36, Portland, Oregon, USA.

Bob Coyne and Richard Sproat. 2001. WordsEye: An Automatic Text-to-scene Conversion System. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '01, pages 487–496, New York, NY, USA.

Michele I. Feist and Dedre Gentner. 1998. On Plates, Bowls, and Dishes: Factors in the Use of English IN and ON. In *Proceedings of the Twentieth Annual Meeting of the Cognitive Science Society*, pages 345–349, Hillsdale, NJ. Erlbaum.

James J. Gibson. 1977. The Theory of Affordances. In *The Ecological Approach to Visual Perception*. Erlbaum.

Birgit Hamp and Helmut Feldweg. 1997. GermaNet - a Lexical-Semantic Net for German. In *Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications*.

Verena Henrich and Erhard Hinrichs. 2010. GernEdiT - The GermaNet Editing Tool. In *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC'10)*, Valletta, Malta.

Annette Herskovits. 1986. *Language and Spatial Cognition: An Interdisciplinary Study of the Prepositions in English*. Cambridge University Press.

Stephen C. Levinson. 2003. *Space in Language and Cognition: Explorations in Cognitive Diversity*. Cambridge University Press, Cambridge, UK.

Christopher Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven Bethard, and David McClosky. 2014. The Stanford CoreNLP Natural Language Processing Toolkit. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 55–60, Baltimore, Maryland.

Donald A Norman. 1988. *The Psychology of Everyday Things*. Basic Books, New York. OCLC: 874159470.

Miriam R L Petruck and Michael J Ellsworth. 2018. Representing Spatial Relations in FrameNet. In *Proceedings of the First International Workshop on Spatial Language Understanding*, pages 41–45, New Orleans. Association for Computational Linguistics.

Princeton University. 2019. WordNet: A Lexical Database for English. https://wordnet.princeton.edu/.

James Pustejovsky. 2017. ISO-Space: Annotating Static and Dynamic Spatial Information. In Nancy Ide and James Pustejovsky, editors, *Handbook of Linguistic Annotation*, pages 989–1024. Springer Netherlands, Dordrecht.

James Pustejovsky, Parisa Kordjamshidi, Marie-Francine Moens, Aaron Levine, Seth Dworman, and Zachary Yocum. 2015. SemEval-2015 Task 8: SpaceEval. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 884–894, Denver, Colorado.

Josef Ruppenhofer, Michael Ellsworth, Miriam R. L Petruck, Christopher R. Johnson, Collin F. Baker, and Jan Scheffczyk. 2016. *FrameNet II: Extended Theory and Practice*.

Melanie Siegel. 2019. Open German WordNet.

Morgan Ulinski, Anusha Balakrishnan, Daniel Bauer, Bob Coyne, Julia Hirschberg, and Owen Rambow. 2014a. Documenting Endangered Languages with the WordsEye Linguistics Tool. In *Proceedings of the 2014 Workshop on the Use of Computational Methods in the Study of Endangered Languages*, pages 6–14, Baltimore, Maryland, USA.

Morgan Ulinski, Anusha Balakrishnan, Bob Coyne, Julia Hirschberg, and Owen Rambow. 2014b. WELT: Using Graphics Generation in Linguistic Fieldwork. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 49–54, Baltimore, Maryland.

Universal Dependencies. 2017. Universal Dependencies. https://universaldependencies.org/.

Piek Vossen, editor. 1998. *EuroWordNet: A Multilingual Database with Lexical Semantic Networks*. Springer Netherlands.