

Image Annotation with ISO-Space: Distinguishing Content from Structure

James Pustejovsky and Zachary Yochum

Brandeis University
Waltham, Massachusetts USA
{jamesp,zyocum}@cs.brandeis.edu

Abstract

Natural language descriptions of visual media present interesting problems for linguistic annotation of spatial information. This paper explores the use of ISO-Space, an annotation specification to capturing spatial information, for encoding spatial relations mentioned in descriptions of images. Especially, we focus on the distinction between references to representational content and structural components of images, and the utility of such a distinction within a compositional semantics. We also discuss how such a structure-content distinction within the linguistic annotation can be leveraged to compute further inferences about spatial configurations depicted by images with verbal captions.

Keywords: linguistic annotation, ISO-Space, image annotation, image description, spatial relations, spatial reasoning

1. Introduction

Image annotation involves the identification of objects in a scene and the spatial relations that exist between them. Images, landscapes, and scenes pose an interesting challenge for spatial description languages, as they involve a fixed frame of reference, and the description of the objects and their positions is relative to this frame. This is quite different from news articles and narratives, where the description is of events from no fixed frame of references. In this paper, we illustrate the challenges faced by interpreting the verbal descriptions associated with images, and how a general specification language for spatial information, such as ISO-Space, can help in distinguishing the *structural* properties of the image from the *content* properties, denoted by the linguistic expressions. This, in turn, allows us to make a restricted number of compositionally derived inferences, relating content and structure in the image.

The annotation of images with natural language descriptions has been the focus of attention within several research areas, in particular, content-based image retrieval (CBIR). One approach is to examine the different ways that geo-referenced images can be described using geography (Purves et al., 2008; Edwardes et al., 2007), while another elicits human inputs without interactive games (von Ahn and Dabbish, 2004; Von Ahn and Dabbish, 2008). Much of the work on text-based image retrieval has relied on extracting information about the image from image captions, as well as the surrounding text and related metadata, such as filenames and anchor text extracted from the referring web pages, as for example, in Yahoo!'s *Image Search* (Datta et al., 2008; Klavans et al., 2008; Salway et al., 2008).

Another kind of image annotation data has become available with the rise of "citizen geography". User-annotated geo-referenced digital photo collections allowing for image content labeling and annotation are being generated in distributed environments, such as Flickr and GoogleEarth. Images are indexed with user-supplied labels that typically form a particular language subset (Grefenstette, 2008). Other techniques have recently emerged for boosting image annotation performance using textual features (Feng

and Lapata, 2008; Leong and Mihalcea, 2009), as well as the use of crowdsourcing strategies (Rashtchian et al., 2010). In addition to these approaches, some recent work focuses on linking annotations in the image more closely to the textual description in the caption (Elliott and Keller, 2011; Elliott and Keller, 2013). (Kordjamshidi et al., 2010) approach the task of image annotation as one of *spatial role labeling*. The annotation scheme they employ involves identifying and classifying certain elements of a text as spatial arguments and then relating these arguments to each other. Spatial role labeling is analogous to semantic role labeling. The related specification is referred to as Holistic Spatial Semantics (HSS) because the complete utterance rather than an isolated word is the main unit of analysis. In practice, this means that annotation is performed at the sentence level. For annotating images described by captions, (Kordjamshidi et al., 2010) use a mixture of mereotopological and orientational relations.

While we build on prior image labeling work, our focus here comes from an interest in unrestricted linguistic descriptions of the spatial relations between objects in images. More specifically, our goal is to apply the ISO-Space specification language to landscape, scene, and image annotation. In the remainder of this abstract, we illustrate how the sublanguage of image captions, as well as verbal descriptions of landscapes and scenes more generally, make reference to both properties of the image as an artifact (image structure regions), and properties of the objects denoted by the image regions (image content). We then present the basic elements of the ISO-Space specification, which allows this distinction to be made naturally in the annotation. Finally, we discuss the implications of content vs. structure relation annotation for image interpretation and inference.

2. Spatial Relations in Image Annotation

There are three distinctive aspects to the language associated with image captions and landscape descriptions: (1) unlike news articles, narratives, or stories, they consist of a fixed frame, determined by the viewer's perspective, or frame of reference, of the scene; (2) the spatial relations in

captions can refer to both structural features of the image, as well as content-dependent features of the objects denoted in the scene; properties of the objects in the image do not necessarily correspond to those properties in the denoted scene. We examine each of these characteristics briefly below.

Static spatial relations as expressed in language employ a combination of three semantic properties (Herskovits, 1986; Vandeloise, 1991; Randell et al., 1992; Wolter and Zakharyashev, 2000):

- (1) a. Mereotopological: *in, touching, outside*;
- b. Orientational: *behind, left of, in front of*;
- c. Metric: *near, far, close by*;

Mereotopological relations (within 2D) can be captured with the relations shown in Table (1) below.

Relation	Description
DC	Disconnected
EC	External Connection
PO	Partial Overlap
EQ	Equal
TPP	Tangential Proper Part
TPP _i	Inverse of TPP
NTPP	Non-Tangential Proper Part
NTPP _i	Inverse of NTPP

Table 1: RCC8 Relations.

Orientational (or projective) relations are typically interpreted relative to a specific frame of reference. We follow Levinson (2003) in distinguishing between three frames of reference (FRs) for spatial relations:

- (2) a. ABSOLUTE: bird’s eye view of a scene;
- b. RELATIVE: viewer perspective;
- c. INTRINSIC: makes reference to inherent orientation of an object.

The default assumption in image captioning is that orientational expressions, such as *left of* and *behind*, are anchored from the perspective of the viewer, hence a relative FR. There are exceptions, however, and captions can often express an intrinsic FR. Consider the images of a tree and a bench in Figure (1).



Figure 1: “The tree is behind the bench.”

The caption is inherently underspecified as to what perspective is being referenced. In Figure (1b), the viewer’s relative FR aligns with the intrinsic FR associated with the bench, while in Figure (1a), they are unaligned, thereby permitting additional, but consistent orientational descriptions. That

is, two additional captions accompany (Figure 1a) but not (Figure 1b), as shown in (3).¹

- (3) a. “The tree is to the left of the bench.”
- b. “There is a bench to the right of the tree.”

While such ambiguities are present when doing spatial annotation over most natural language texts, one problem that is unique to image annotation is the ability for captions to reference structural properties of the image itself. As mentioned above, spatial relations in captions can refer to “content configurations”, such as where one object is relative to another object in the depicted situation, or “structural configurations”, such as where one object is positioned within the structure of the image. Consider the following additional annotations for Figure (1).

- (4) a. On the left side of the picture is a big tree. (Figure 1a)
- b. A tree is in the center of the scene. (Figure 1b)
- c. The tree’s shadow is in the lower left corner. (Figure 1b)

Positional terms, such as *side, corner, center, middle, top*, are orientational relations inherent to the viewer (relative) frame of reference. When the frame is an image, however, they can refer to relations between structural regions within the image. Formally, such expressions map from image structure to image content. This distinction is brought out very clearly in the image in Figure (2) along with its caption.²



Figure 2: “View of a city with green fields in the center and several mountains with snow covered peaks in the background.”

The content configurations inherent in this caption include relations such as the following (informally stated):³

¹Notice that “A bench is in front of a tree.” is acceptable for (Figure 1b), but not (Figure 1a).

²IAPRTC image 2157 (Müller et al., 2010; Müller et al., 2012; Kiros and Szepesvári, 2012).

³“Backgrounding” and “foregrounding” are, in fact, content-denoting spatial functions, interpreted within an orientational domain, where *front_of_r* is relative to the viewer: e.g., $\forall x[\text{foreground}(x) \leftrightarrow \neg \exists y[\text{front_of}_r(y, x)]]$.

- (5) a. *in(fields,city)*
- b. *in_front_of(city,mountains)*
- c. *in(mountains,background)*

There are also two structural configurations associated with the caption. Here, we capitalize reference to structural aspects of the image.

- (6) a. *in(city,IMAGE)*
- b. *in(fields,CENTER)*

In this photo in particular, we see that it is crucial to distinguish the structural placement of objects from their content configurations, when annotating. For example, while the fields are centered within the image, they are not in the center of the city. We return to this issue in the final section of the paper.

In the next section, we illustrate how ISO-Space represents spatial relations, and in particular, how it can distinguish between structural and content spatial configurations.

3. Image Structure in ISO-Space

The annotation of spatial information as conveyed through language involves a considerable number of concepts, including at least the following: a PLACE tag (for locations, entities participating in spatial relations, and paths); LINK tags (for topological relations, direction and orientation, time and space measurements, and frames of reference); and a SIGNAL tag (for spatial prepositions)⁴. ISO-Space has been designed to capture both spatial and spatiotemporal information as expressed in natural language texts (Pustejovsky et al., 2012). We have followed a strict methodology of specification development, as adopted by ISO TC37/SC4 and outlined in (Bunt, 2010) and (Ide and Romary, 2004), and as implemented with the development of ISO-TimeML (Pustejovsky et al., 2005) and others in the family of SemAF standards.

ISO-Space allows one to identify the source and type of the text being annotated. This is done with the document creation location (DCL) attribute. This is a distinguished location that serves as the “narrative or reference location”. For example, if the document type is a typical news article or narrative, then the DCL is generally identified in the beginning of the text (e.g., as part of the byline), similarly to how the Document Creation Time is specified in TimeML (Pustejovsky et al., 2005). Image captions are different, however, in that the document describes a *representational artifact*,⁵ such as an image or a Google *Street View* scene; hence, the document type is distinguished as an IMAGE. When this is the case, location tags (PLACE, PATH, SPATIAL_ENTITY, MOTION, and NON-MOTION_EVENT) assume a new attribute, domain, which has two values: STRUCTURE and CONTENT.

- (7) domain ::= STRUCTURE | CONTENT

⁴For more information, cf. (Pustejovsky et al., 2012).

⁵These are represented as *phys_obj • info* complex types (dot objects), and inherit the properties of both type elements (Pustejovsky, 1995).

This allows the spatial relations to differentiate the kinds of regions being identified in the caption. Furthermore, this means that the DCL can take two values: an *Image Structure Location*, for reference to the image as an object; and an *Image Content Location*, which is what the picture refers to (as in the default DCL for most texts).

Let’s see how this plays out with one of the captions we encountered earlier, that in Figure (2).

- (8) a. [**View**_{pl1}] of a [**city**_{pl2}] with green [**fields**_{pl3}] in the [**center**_{pl4}]....
- PLACE(id=pl1, domain=STRUCT, dcl=TRUE)
- PLACE(id=pl2, domain=CONT, dcl=FALSE)
- PLACE(id=pl3, domain=CONT, dcl=FALSE)
- PLACE(id=pl4, domain=STRUCT, dcl=FALSE)

The fragment of the caption above (8) introduces two image structure locations (*view* and *center*), and two image content locations (*city* and *fields*). Furthermore, *view* is marked as the DCL, meaning it is the embedding region for all image structure locations.

Now let us see how these two types of locations participate in spatial relations. There are four relation tags in ISO-Space, defined as follows:

- (9) a. QSLINK – qualitative spatial relations;
- b. OLINK – orientation relations;
- c. MLINK – dimensions of a region or the distance between them.
- d. MOVELINK – for movement relations;

QSLINKs are used in ISO-Space to capture topological relationships between tag elements captured in the annotation. The relType attribute values come from an extension to the RCC8 set of relations that was first used by SpatialML. The possible RCC8+ values include the RCC8 values (Randell et al., 1992), in addition to IN, a disjunction of TPP and NTPP (cf. Table 1).

Orientation links describe non-topological relationships. A SPATIAL_SIGNAL with a DIRECTIONAL semantic_type triggers such a link. In contrast to qualitative spatial relations, OLINK relations are built around a specific frame of reference type and a reference point. The referencePt value depends on the frame_type of the link. The ABSOLUTE frame type stipulates that the referencePt is a cardinal direction. For INTRINSIC OLINKs, the referencePt is the same identifier that is given in the ground attribute. For RELATIVE OLINKs, the identifier for the viewer should be provided as to the referencePt. When the document type is IMAGE, all olinks are interpreted as relative FR relations (unless otherwise stated), with the “VIEWER” as the referencePt.

4. Inferring Structure-Content Relations

In this section, we illustrate briefly some of the consequences of making a distinction between image structure and image content directly in the annotation. Most significantly, if the resulting relations are to be used for performing inferencing (e.g., constraint satisfaction, deductive reasoning, probabilistic techniques), then it is important to distinguish those configurations referencing the region structure of the image from those denoting spatial relations in the depicted situation.

When reasoning over temporal or spatial relations, it is convenient to think in terms of a *composition table* of all possible relations in the domain (Allen, 1983; Cohn and Hazarika, 2001). A composition table is a mapping, $\mathcal{R} \times \mathcal{R} \rightarrow 2^{\mathcal{R}}$, where \mathcal{R} is the set of possible relations in the domain. The result for a finite \mathcal{R} is an $n \times n$ table. Within RCC8, for example, if we know both $TPP(x, y)$ and $NTPP(y, z)$, then we can infer that $NTPP(x, z)$. This model, however, results in what is called a weak composition table (Bennett et al., 1997; Renz and Ligozat, 2005). Nevertheless, it is extremely useful for inference, and can be applied to the problem of reasoning over spatial relations in image annotation.

As mentioned above, locations in image captions refer to either STRUCTURE or CONTENT. Since structural relations (R_s) only denote regions within the embedding 2D space (the image itself), these can be modeled with RCC8, enriched with orientation (Zimmermann and Freksa, 1996). Let us assume a set of image structural relations, such as *corner_of(x,y)*, *middle_of(x,y)*, *left_side(x,y)*. Hence, for any pair of relations, R_s^1 and R_s^2 , we can use the appropriate composition table for computing further inferences, i.e., the set of $R_s^1 \circ R_s^2$ (Cohn and Hazarika, 2001; Wallgrün et al., 2007). For image content relations (R_c), we assume the set of RCC8 relations, along with orientational and metric relations, as represented within ISO-Space.

Now let us examine how content relates structure in an image, i.e., the relation $R(c, s)$ ⁶. For a content relation, R_c , and a structure relation, R_s , let us define the composition as: $R_c \circ R_s = \{(a, c) | \exists b[(a, b) \in R_c \wedge (b, c) \in R_s]\}$. Some of the values of the weak composition table resulting from these relation pairs are shown below in Table 2.

$R_c \circ R_s$	<i>Left</i> (b, c)	<i>Mid</i> (b, c)	<i>Cor</i> (b, c)	...
<i>front</i> (a, b)	<i>Left</i> (a, c)	<i>Mid</i> (a, c)	<i>Cor</i> (a, c)	
<i>beh</i> (a, b)	<i>Left</i> (a, c)	<i>Mid</i> (a, c)	<i>Cor</i> (a, c)	
<i>IN</i> (a, b)	<i>Left</i> (a, c)	<i>Mid</i> (a, c)	<i>Cor</i> (a, c)	
<i>DC</i> (a, b)	*	*	*	

Table 2: Content-Structure Relation Composition Table
front: in.front_of, *beh*: behind, *in*: TPP or NTPP, *Left*:
on.left_side, *Mid*: middle_of, *Cor*: corner_of

Given the composition table in Table 2, let us see how it can be applied to reason over the image and caption shown in Figure 3.

The relations directly annotated from the image caption in Figure 3 are shown in (10); the subscript s denotes a structure location, and the subscript c denotes a content location.⁷

- (10) Dirt [**road**_{t1}]_c in [**middle**_{t2}]_s of a [**landscape**_{t3}]_s ... a [**tree**_{t4}]_c in [**foreground**_{t5}]_c on the [**left**_{t6}]_s.
QSLINK(IN,11,12)
OLINK(middle,12,13)

⁶This is the relation introduced by the dot in the complex type of representational artifacts, as modeled in Generative Lexicon (Pustejovsky, 2006; Asher and Pustejovsky, 2006).

⁷The extent tag types in (10) have been generalized to a location tag l for this example.



Figure 3: “Dirt road in the middle of a landscape with wooded hills; a tree in the foreground on the left.”

QSLINK(IN,14,15)
OLINK(foreground,15,13)
QSLINK(TPP,14,16)
OLINK(left_side,16,13)

Reasoning from the composition tables, we can now infer the additional relations shown in (11).

- (11) a. Behind the tree is a large shrub.
b. The tree is to the left of the road.
c. The large shrub is on the left.
d. The large shrub is to the left of the road.

The interpretation of a spatial relation between locations with respect to an image depends on the values of the domain attributes of that relation’s arguments. That is, relations with location arguments from different domains (such as some of the relations in (10)) require coercion when deriving its compositional interpretation. As such, despite the exclusive disjunction between STRUCTURE and CONTENT values for the domain attribute in the annotation specification, we assume all locations marked as CONTENT to be complex objects⁸. This enables spatial relations to access the corresponding structural aspects of their arguments’ content within the structure domain. On the other hand, except for the DCL, locations whose domain is marked as STRUCTURE need not have a complex type.

Under the assumption of complex types for locations referred to within the content domain, we can provide post-annotation rules for coercing relational arguments when they have incompatible domain values. For example, in (12), the locations in subject position are typed (annotated) as CONTENT, and they appear in a location relation with a STRUCTURE argument.

- (12) a. [The road]_c is in [the middle]_s.
b. [The tree]_c is in [the corner]_s.

These are interpreted as STRUCTURE-denoting elements through a coercion operation (Pustejovsky et al., 2009), as shown in (13).

- (13) a. $R(x_c, y_s) \mapsto R(x_s, y_s)$
b. $R(x_s, y_c) \mapsto R(x_s, y_s)$

⁸To use dot-object notation, the type for such a location may be written as CONTENT • STRUCTURE.

We must make special exceptions for the DCL, since it is distinguished explicitly as an embedding space for both structure and content domains, thus, for each relation it participates in, its type can be matched to that of whatever argument it is related to:

- (14) a. $x : \text{DCL} = \top \rightarrow R(x_s, y_c) \mapsto R(x_c, y_c)$
 b. $y : \text{DCL} = \top \rightarrow R(x_c, y_s) \mapsto R(x_c, y_c)$

According to the axioms outlined in (13) and (14), then we can appropriately interpret a caption such as the one for the landscape painting in Figure (4) (Smithsonian Cooper-Hewitt, 1880), in which the representational content of the image referred to as *sky* in the caption (i.e., the representation of a volumetric expanse of light-refracting gases) is not related to any structural region of the image, since it is of the wrong domain. However, the structural, 2D region corresponding to the asserted complex *sky* object does relate to the structural region of the canvas referred to by *quarter*.⁹ The annotation for Figure (4) is included in (15).



Figure 4: “An orange, yellow, and blue sky fills the upper left quarter of the sheet.”

- (15) ...[sky]_{l1}]_c fills the upper left [quarter]_{l2}]_s of the [sheet]_{l3}]_s ...
 QSLINK(PO,11,12)
 OLINK(corner_of,12,13)
 OLINK(left_side,12,13)
 OLINK(top,12,13)

5. Conclusion

In this paper we have presented a formalism and strategy for spatially annotating the captions associated with visual images. We hope to have demonstrated that, by making a distinction between structure- and content-based relations, properties of the objects in the image can be revealed through compositionally generated inferences. The specification language ISO-Space has been enriched to reflect this distinction. While still preliminary, our initial results suggest that a weak composition table is both sound and informative for deriving new spatial relations.

⁹The content-to-structure coercion of 11 in the `qslink` corresponds to the axiom in (13a).

Acknowledgements

This research was supported by grants from NSF’s IIS-1017765 and NGA’s NURI HM1582-08-1-0018. All errors and mistakes are, of course, the responsibilities of the authors.

6. References

- Allen, J. F. (1983). Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11):832–843.
- Asher, N. and Pustejovsky, J. (2006). A type composition logic for generative lexicon. *Journal of Cognitive Science*, 6:1–38.
- Bennett, B., Isli, A., and Cohn, A. G. (1997). When does a composition table provide a complete and tractable proof procedure for a relational constraint language. In *Proceedings of the IJCAI-97 workshop on Spatial and Temporal Reasoning, Nagoya, Japan*. Citeseer.
- Bunt, H. (2010). A methodology for designing semantic annotation languages exploiting syntactic-semantic isomorphisms. In *Proceedings of ICGI 2010, Second International Conference on Global Interoperability for Language Resources*.
- Cohn, A. G. and Hazarika, S. M. (2001). Qualitative spatial representation and reasoning: An overview. *Fundamenta Informaticae*, 46(1):1–29.
- Edwardes, A., Purves, R., Birche, S., and Matyas, C. (2007). Deliverable 1.4: Concept ontology experimental report. Technical report, TRIPOD Project.
- Elliott, D. and Keller, F. (2011). A treebank of visual and linguistic data.
- Elliott, D. and Keller, F. (2013). Image description using visual dependency representations. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, page 12921302. Association for Computational Linguistics.
- Feng, Y. and Lapata, M. (2008). Automatic image annotation using auxiliary text information. In *ACL*, pages 272–280.
- Herskovits, A. (1986). *Language and Spatial Cognition: an Interdisciplinary Study of the Prepositions in English*. Cambridge University Press.
- Ide, N. and Romary, L. (2004). International standard for a linguistic annotation framework. *Natural Language Engineering*, 10(3-4):211–225.
- Kiros, R. and Szepesvári, C. (2012). Deep representations and codes for image auto-annotation. In *Advances in Neural Information Processing Systems 25*, pages 917–925.
- Kordjamshidi, P., Moens, M., and van Otterlo, M. (2010). Spatial role labeling: task definition and annotation scheme. In *Proceedings of LREC 2010 - The seventh international conference on language resources and evaluation*.
- Leong, C. W. and Mihalcea, R. (2009). Explorations in automatic image annotation using textual features. In *Proceedings of the Third Linguistic Annotation Workshop*, pages 56–59. Association for Computational Linguistics.

- Müller, H., Clough, P., Deselaers, T., and Caputo, B. (2010). *ImageCLEF: Experimental Evaluation in Visual Information Retrieval*, volume 32. Springer.
- Müller, H., Kalpathy-Cramer, J., Demner-Fushman, D., and Antanid, S. (2012). Creating a classification of image types in the medical literature for visual categorization. In *Proc. of SPIE Vol.*, volume 8319, pages 83190P–1.
- Purves, R., Edwardes, A., and Sanderson, M. (2008). Describing the where—improving image annotation and search through geography.
- Pustejovsky, J., Knippen, R., Littman, J., and Saurí, R. (2005). Temporal and event information in natural language text. *Language Resources and Evaluation*, 39:123–164, May.
- Pustejovsky, J., Rumshisky, A., Moszkowicz, J., and Batiukova, O. (2009). Gml: Annotating argument selection and coercion. submitted to IWCS-8.
- Pustejovsky, J., Moszkowicz, J., and Verhagen, M. (2012). A linguistically grounded annotation language for spatial information. *TAL*, 53(2).
- Pustejovsky, J. (1995). *The Generative Lexicon*. MIT Press, Cambridge, MA.
- Pustejovsky, J. (2006). Type theory and lexical decomposition. *Journal of Cognitive Science*, 6:39–76.
- Randell, D., Cui, Z., and Cohn, A. (1992). A spatial logic based on regions and connections. In Kaufmann, M., editor, *Proceedings of the 3rd International Conference on Knowledge Representation and REasoning*, pages 165–176, San Mateo.
- Rashtchian, C., Young, P., Hodosh, M., and Hockenmaier, J. (2010). Collecting image annotations using amazon’s mechanical turk. In *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon’s Mechanical Turk*, pages 139–147. Association for Computational Linguistics.
- Renz, J. and Ligozat, G. (2005). Weak composition for qualitative spatial and temporal reasoning. In *Principles and Practice of Constraint Programming-CP 2005*, pages 534–548. Springer.
- Smithsonian Cooper-Hewitt, N. D. M. (1880). Drawing, “carter dome from carter lake, new hampshire”, ca. 1880.
- Vandeloise, C. (1991). *Spatial prepositions: A case study from French*. University of Chicago Press.
- von Ahn, L. and Dabbish, L. (2004). Labeling images with a computer game. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 319–326. ACM Press New York, NY, USA.
- Von Ahn, L. and Dabbish, L. (2008). Designing games with a purpose. *Communications of the ACM*, 51(8):58–67.
- Wallgrün, J. O., Frommberger, L., Wolter, D., Dylla, F., and Freksa, C. (2007). Qualitative spatial representation and reasoning in the sparq-toolbox. In *Spatial Cognition V Reasoning, Action, Interaction*, pages 39–58. Springer.
- Wolter, F. and Zakharyashev, M. (2000). Spatio-Temporal Representation and Reasoning based on RCC-8. *KR 2000: Principles of Knowledge Representation and Reasoning*, pages 3–14.
- Zimmermann, K. and Freksa, C. (1996). Qualitative spatial reasoning using orientation, distance, and path knowledge. *Applied Intelligence*, 6(1):49–58.