Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020), pages 5855–5864
Marseille, 11–16 May 2020

From Spatial Relations to Spatial Configurations

Soham Dan [1], Parisa Kordjamshidi[3,4], Julia Bonn[2], Archna Bhatia[3], Jon Cai[2], Martha Palmer[2], Dan Roth[1]
[1]University of Pennsylvania, [2]University of Colorado, Bolder, [3]IHMC, [4]Michigan State University
{sohamdan,danroth}@seas.upenn.edu , {julia.bonn,jon.z.cai,martha.palmer}@colorado.edu ,
kordjams@msu.edu, abhatia@ihmc.us

Abstract

Spatial Reasoning from language is essential for natural language understanding. Supporting it requires a representation scheme that can capture spatial phenomena encountered in language as well as in images and videos. Existing spatial representations are not sufficient for describing spatial configurations used in complex tasks. This paper extends the capabilities of existing spatial representation languages and increases coverage of the semantic aspects that are needed to ground spatial meaning of natural language text in the world. Our spatial relation language is able to represent a large, comprehensive set of spatial concepts crucial for reasoning and is designed to support composition of static and dynamic spatial configurations. We integrate this language with the Abstract Meaning Representation (AMR) annotation schema and present a corpus annotated by this extended AMR. To exhibit the applicability of our representation scheme, we annotate text taken from diverse datasets and show how we extend the capabilities of existing spatial representation languages with fine-grained decomposition of semantics and blend it seamlessly with AMRs of sentences and discourse representations as a whole.

Keywords: Semantics, Knowledge Discovery/Representation

1. Introduction

Spatial reasoning is necessary for many tasks. For example, consider the collaborative building task (Narayan-Chen et al., 2019) where a human architect (knows the target structure but cannot manipulate blocks) issues commands to a builder (who does not know the target structure but can place and manipulate blocks) with the aim of creating a complex target structure in a grounded environment (Minecraft). This task is challenging from a spatial reasoning perspective because the architect does not issue commands in terms of the unit blocks that the builder uses, but complex shapes and their spatial aspects. A single instruction can convey information about complex spatial configurations of several objects and describe nested relationships while shifting focus from describing one object to another. For example, in *Place the red block on the yellow block which is to the left of the blue column.*, the focus shifts from describing the intended location of the red block to describing the location of the yellow block. Or, in a navigation description, there can be a sequence of spatial expressions describing a single object, as in *Over the hill and through the woods....* There is a severe limitation in the ability of existing spatial representation schemes to represent sentences involving complex spatial concepts. An example sentence from the Minecraft corpus is shown in Figure 1.

Table 1 highlights some of the challenging instances from a few recent datasets that cannot be adequately represented by existing spatial representations(for a more detailed comparison with previous schemes refer to Section 6). For instance, the spatial representation in (Kordjamshidi et al., 2010) does not capture the fine-grained semantics of the object properties and does not distinguish between the frame of reference and the perspective. The scheme of (Pustejovsky et al., 2011) does not handle complex landmarks in the first example with the sequence of spatial properties. We observe the ubiquity of these hard spatial descriptions across different datasets. Here is another such instance from NLVR (Suhr et al., 2017a) depicting *spatial focus shift*: *There is a box with a black item between 2 items of the same color and no item on top of that.* We propose an integrated view called SPATIAL CONFIGURATION that captures extensive spatial semantics that is necessary for reasoning. To develop an automatic builder, in the case of the collaborative construction task, we need an explicit representation of all involved spatial components and their spatial relations in the instruction.

One of our key contributions is to represent this information compactly as a part of the configuration. Further, we can represent *spatial focus shift* and *nested concepts* as part of the scheme. We also situate the configuration scheme within a new extended spatial AMR framework (Bonn et al., 2019). Our spatial configuration scheme takes the spatial roles and relations identified in spatial AMR graphs and converts them into an easy to read general framework, resulting in a maximally expressive spatial meaning representation language.

Recent research shows the limitations of training monolithic deep learning models in two important aspects of interpretability and generalizability (Hu et al., 2017). Symbolic representations that can support reasoning capabilities are shown to have a critical role in improving both of the above-mentioned aspects (Goldman et al., 2018; Krishnaswamy et al., 2019; Suhr et al., 2017b). A robust intermediate representation can help an off-the-shelf planner to ground a symbolic representation of the spatial concepts to a final executable output. This representation should be as general and domain independent as possible to eliminate the need to retrain an automatic annotation system for a new domain. Using AMR (Banarescu et al., 2013) as a stepping stone to our spatial configurations helps to ensure generalizability and domain independence. Our goal is to identify

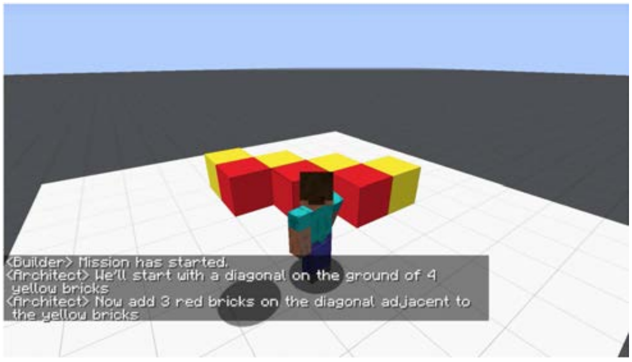

Figure 1: An instance of the collaborative building task. The last instruction was : Now add 3 red bricks on the diagonal adjacent to the yellow bricks.

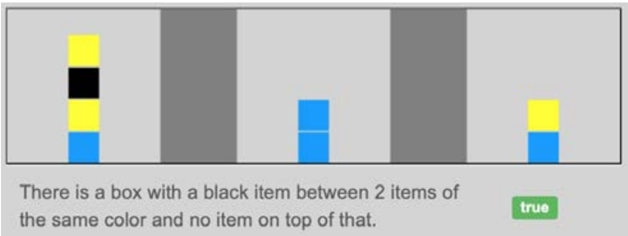

Figure 2: An example from the NLVR corpus that demonstrates *spatial focus shift* from the *black item* to the *yellow item*.

various spatial semantic aspects that are required by a concrete meaning representation for downstream tasks.

2. Expanded AMRs

In this section, we briefly describe the spatial AMR extension that we designed to produce graphs suitable for mapping directly onto our spatial configurations. AMR is a good starting point for this annotation because it is domain general and relatively syntax-independent. The graphs themselves are comprised of nesting relations and concepts that may be inverted or rearranged depending on focus, which is useful in representing the kinds of complex and nesting spatial relations described above. While the new AMRs cover a richer set of spatial semantics and complex spatial relationships, they also still represent any non-spatial portions of the sentence. One contribution of our Spatial Configuration schema is the ability to extract the spatial relationships and spatial object properties from the AMR and organize them into an easily interpretable format that maintains the complex relationships and nesting.

The corpus that we describe here [1] includes annotation on over 5000 dialogue sentences (170 full dialogues) that discuss collaborative construction events in the Minecraft dataset (Narayan-Chen et al., 2019). We have an additional 7600 annotated automatically-generated sentences representing builder actions, giving a total of 12,600 spatial AMRs. The AMR expansion involves the addition of 150+ new or updated PropBank (Palmer et al., 2005; Gildea and Palmer, 2002) rolesets as well as a dozen general semantic frames, roles, and concepts that signal spatial relationships not previously covered by AMR. Using rolesets to annotate spatial relations allows us to disambiguate senses and to group together etymologically-related relations from different parts of speech within those senses. For example, the roleset **align-02** includes aliases align-v, aligned-j, line-up-v, and in-line-p, and **down-03** includes aliases down-p, down-r, downward-r, etc.. Prepositional and adverbial aliases are new for PropBank and AMR. The roles that make up the rolesets cover both semantically and pragmatically derived information and are labeled with a new set of spatial function tags that map to the elements of our Spatial Configurations. A typical directional spatial roleset includes two primary spatial entities held in comparison to one another (SE1, SE2), an axis or directional line between them (AXS), and an anchor which is used to name the entity whose spatial framework is being activated by the relation (ANC):

above-01: higher on a vertical axis
:ARG1-SE1 entity above
:ARG2-SE2 entity below
:ARG3-ANC anchor
:ARG4-AXS axis

SE1 and SE2 are used for primary spatial entities that have an external (Tenbrink and Kuhn, 2011) relationship to each other; when the entities have an internal relationship, we create a separate sense and call them PRT and WHL. For example, **top-06** entails an internal relationship (the trajector is part of the landmark) and **on-top-03** entails an external relationship (the trajector and landmark are discrete from one another). Other relevant function tags are for angles (ANG), orientations (ORT), and scale (SCL) for use with scalar adjectives.

New general semantic frames and roles allow us to target Spatial Configurations that are triggered outside of standard roleset applications. **Have-configuration-91/:configuration** replaces :mod/:domain roles in cases where a set of entities is arranged into some sort of constellation, for example, *the chairs are in a ring around the fire pit*. **Spatial-sequence-91** provides a means of indicating that a temporal sequence of entities is meant to be interpreted as a spatial sequence, as in (*put down red red green space red*) and *now a green row of 3. Next a green row of 4.* **Cartesian-framework-91** includes arguments for up to 3 axes, an origin, and a spatial entity housing the framework. **Cartesian-coordinate-entity** is an entity frame whose arguments are :x, :y, :z and :framework. Other general semantic roles (with expected reified frames) include **:orientation**, **:size**, **:color**, **:anchor**, **:axis**, and **:pl**. **:Pl** takes '+' as a value to indicate plurality, which AMR omitted and which is vital in grounded spatial annotation.

The recently added prepositional and adjectival spatial relations are prevalent in language. For example, in the NLVR (Suhr et al., 2017a) training dataset (1000 sentences) consisting of 10,710 tokens, we found 1563 occurrences of these newly added frame file concepts, which is almost 15% of all tokens, and in the 3D BLOCKS dataset (Bisk et al., 2018), we found 39,917 occurrences of the new additions among a total of 250,000 tokens, almost 16% of all tokens. The recent GQA dataset (Hudson and Manning, 2019) has 22.4% spatial questions, 28% object attribute questions and

[1]https://github.com/jbonn/CwC-Minecraft-AMR

52% questions on compositionality. Our spatial representation scheme with modular configurations that facilitate compositionality for reasoning and explicit representation of object properties will be helpful for such tasks too.

3. Representation Language

In this section we describe the SPATIAL CONFIGURATION schema and as running examples we present two of our annotated instruction sentences from the Minecraft domain. We also include general examples beyond Minecraft to show our representation language is applicable universally. In Figures 7 and 8 we show the screenshots from the Minecraft environment of the actual execution of the instruction. In Tables 3 and 4 we show the detailed annotation of the spatial configuration elements. In Figures 3, 4, 5 and 6 we show the spatial configuration annotation graph and the aligned extended AMR annotations.

We represent each sentence S as a set of *spatial entities* $\{E\}$ and a set of *configurations* $\{C\}$, that is, $S =< \{E\}, \{C\} >$. Each of our examples have two configurations because there is a shift of spatial focus; Tables 3 and 4 show them in detail. Each **spatial configuration** C is described as a tuple,

C= $<$*tr,*$\{$*lm* $|$*path*$\}$*,m,sp,FoR,v,QT* $>$. The elements of the tuple are described in Table 2.

3.1. Spatial Entities and Properties

Each **spatial entity** E corresponds to an object in the text and can be described by a number of properties. We represent each spatial entity as follows: $E =< id, head, \{prop\} >$ where id is an identifier and $head$ corresponds to the entity (the actual shape in the Minecraft domain) that participates in spatial configurations with varied roles of tr, lm or a part of a $path$. Spatial entities are described by **spatial properties**, $\{prop\}$. Each element in $prop$ is a property. The properties are represented as $prop =< name, span >$ where name takes a set of values, depending on the domain. In Tables 3, 4 $\{shape, size, color, part-of, area, location, quantity\}$ are the properties of our domain and span refers to the actual lexical occurrence of the property. These properties are present as modifiers in the AMR graphs, which represent them via rolesets or general semantic roles like **:color** and **:size**. Notice that *top* can be a *part-of* property , as in, *the top of the tower is touching the roof* or it could be a *area* property , as in, *the book is on top of the table*, where *top* refers to the area on top of the table. Spatial entity heads and their properties are all *consuming tags*, that is, $head$ and $prop$ refer to the actual sub-strings in the sentence that are referring to the entity, or property, respectively. In Figures 3, 4, 5 and 6 we show both of the presented examples with the corresponding AMRs. We can see the tight connection between elements of the spatial configuration and the relevant portions of the AMR and how we fit the spatial concepts as a part of the larger AMR scheme to obtain a richer representation language.

3.2. Spatial Configuration Elements

The roles of Trajector (tr) and Landmarks (lm) are associated with spatial entities. **Trajector** roles are presented as *tr=*$<$*id,e*$>$ where id is the identifier assigned to a tr and e is a spatial entity that plays this role. In our first example, *the large red block* is a spatial entity with the semantic head being *block*, exactly as in the AMR. A trajector can be composed of multiple objects such as *a woman and a child*. Trajectors can also be implicit in a sentence when identifiable from the context. In the sentence, *Just arrived there!*, *I* is implicit. In Minecraft we treat the *space* itself as a trajector (Figure 6), which is again in exact alignment with the spatial AMR.

Landmarks are represented similarly and can be complex. In the sentence *The block is in between the sphere and the cube.*, *the sphere and the cube* is a complex landmark that involves two objects.

Spatial indicators are represented as $<$*id,span,prop*$>$. Each indicator participates in one configuration (unless the configuration describes a path in which we case we expect multiple spatial indicators corresponding to different segments of the path) and its role and properties are described therein. The properties are represented by a name and a span but with different values such as $\{metric, degree, distance, direction, region\}$. In Configuration 2 in Table 4, *in_between* indicates the relation between the *spaces (e3)* and the *blocks (e1,e2)*. and *5 spaces* is the metric property for *from*.

Motion indicator is represented as $<$*id,span,prop*$>$. An example property is *speed*. When the configuration includes a **path** , particularly for the dynamic spatial descriptions or fictive motion (*The river meanders through the valley.*), the path is described using a set of landmarks and spatial indicators: *path*= $<\{$*(lm,sp,path-part)*$\}$*, prop*$>$. Here *prop* is the property associated with the path. For example, in Table 3, *Orientation* with *diagonally* as its value is a property of the path. For the **path** variable, each spatial indicator which is associated with a lm in the path indicates which part of the path it is referring to, where *path-part* $\in \{begin, end, middle, whole\}$. Each of these path-part values are illustrated in the example *I am going from DC (begin) to NY (end) through PHL (middle) along I-78 (whole).* In Table 3, the path includes two landmarks: one is the beginning of the path and the other is the end. In Table 4, the beginning of the path is implicit. Complex configurations having a path with multiple landmarks can have multiple **frames-of-reference**s (FoR). Thus, FoR is represented with respect to a landmark, as $<$*lm, value*$>$ where *value* $\in \{$*intrinsic,relative,absolute*$\}$. Intrinsic FoR (object-centered) activates an internal axis or part of the object of the spatial indicator, e.g., *top of the building*. *Relative FoR (viewpoint-centered)* uses another spatial entity to anchor the location or orientation of the current object's spatial description, e.g., *my left* and *your right*. *Absolute FoR (geo-centered)* is a fixed FoR, e.g., *North*. As an example of a configuration having multiple FoRs with different values, consider : *I walked from the center of the park to the left of the statue.* In the path for this configuration, the first landmark has an intrinsic FoR while the second landmark has a relative FoR. In contrast to the FoR, the **viewer** is unique in a single configuration and is represented as a value $v \in \{first-person, second-person, third-person\}$ (Tenbrink and Kuhn, 2011; Lee et al., 2018). A configuration

Dataset	Example Sentence
BLOCKS	Texaco should line up on the top right corner of BMW in the middle of Adidas and HP (1)
GQA	Are there any cups to the left of the tray that is on top of the table? (2)
NLVR	There is a black square touching the wall with a blue square right on top of it. (3)
SpaceEval	While driving along the village's main road the GPS showed us the direction right ahead, and after two minutes, just few hundred meters from the end of the village, we reached the spot. (4)
SpRL	a white bungalow with big windows, stairs to the left and the right, a neat lawn and flowers in front of the house (5)
Minecraft	place a green block on the edge three spaces from the left side (6)

Table 1: Examples from popular datasets that highlight the need for a more expressive spatial representation language

Trajector *(tr)*	The entity whose location or trans-location is described in a spatial configuration.
Landmark *(lm)*	The reference object that describes the location of the *tr* or is a part of its path of motion.
Motion Indicator*(m)*	Spatial movement usually described by a motion verb.
Frame-of-Ref.*(FoR)*	A coordinate system to identify location of an object, can be intrinsic, relative or absolute.
Path *(path)*	The tr's location can be described via a path of motion instead of a basic lm.
Viewer *(v)*	When FoR is relative, this indicates the viewer as first, second or third person.
Spatial Indicator *(sp)*	The lexical form of the relation between the trajector and landmark.
Qualitative type *(QT)*	The qualitative/formal type of the relation between the tr and lm.

Table 2: The components of a generic spatial configuration

Spatial Entities		
id	head	properties $\{name = span\}$
e1	block	$\{size = large, col = red\}$
e2	column	$\{area = top, col = blue\}$
e3	column	$\{area = top, col = yellow\}$
e4	cube	$\{col = orange\}$

	Configuration 1	Configuration 2
tr	$< t1, e1 >$	$< t2, e3 >$
lm	$< l1, e2 >, < l2, e3 >$	$< l3, e4 >$
sp	*<s1,from >* *<s2,to >*	*<s1,from,* $\{metric = 5spaces\}$*>*
m	*<m1,move, >*	NULL
path	$< l1, s1, begin >$ $< l2, s2, end >$ $\{orientation = diagonally\}$	NULL
FoR	$< l1, relative >$ $< l2, relative >$	$< l3, relative >$
v	first-person	first-person
QT	*<directional, relative>*	*<distal, quantitative>* *<topological, DC>*

Table 3: Relational representation of spatial entities and configurations in the sentence: *Move the large red block diagonally from the top of the blue column to the top of the yellow column, which is 5 spaces from the orange cube.*

Spatial Entities		
id	head	properties $\{name = span\}$
e1	block	$\{col = orange\}$
e2	block	$\{loc = base, area = right, col = red\}$
e3	spaces	$\{quantity = 2\}$

	Configuration 1	Configuration 2
tr	$< t1, e1 >$	$< t2, e3 >$
lm	$< l1, e2 >$	$< l2, e2 >, < l3, e1 >$
sp	*<s1,to >*	*<s2,in_between >*
m	*<m1,place >*	NULL
path	$< implicit, begin >$ $< l1, s1, end >$	NULL
FoR	$< l1, relative >$	$< l2, relative >$ $< l3, relative >$
v	first-person	first-person
QT	*<directional, relative>*	*<topological, EC>*

Table 4: Relational representation of spatial entities and configurations in the sentence: *Place an orange block to the right of the base red block with two spaces in between.*

describes at most one trajector based on at most one basic/-complex landmark or a path. It includes various semantic aspects that can make it specific enough for the visualization of a single spatial relation. Therefore, when there is more than one spatial indicator (except for a path which is a complex landmark with multiple spatial indicators), we have one configuration for each spatial indicator. Each configuration will have at most one motion indicator.

3.3. Qualitative Types

Fine-grained spatial relation semantics can be represented using specialized linguistically motivated ontologies such as General Upper Model (Bateman et al., 2010). Another approach has been mapping to formal spatial knowledge representation models that are independent from natural language (Mani et al., 2008; Kordjamshidi et al., 2010; Pustejovsky et al., 2011). The latter provides a more tractable formalism for spatial reasoning. The spatial formal models are divided into three categories: *topological*, *directional* and *distal*, and for each category specific formalisms have been invented (Wallgrün et al., 2007; Liu et al., 2009). We use these qualitative types. Directional systems can be relative or absolute, distal systems can be qualitative or quantitative and topological models can follow

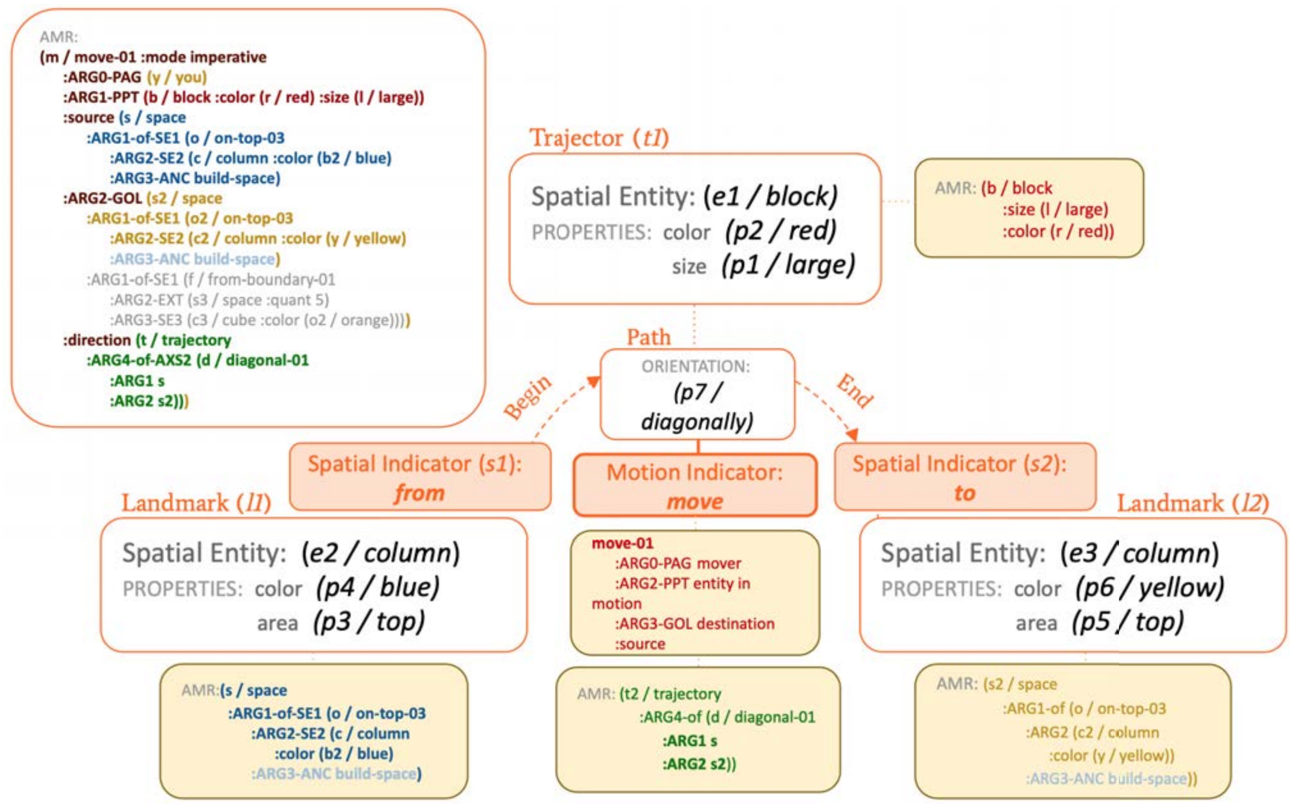

Figure 3: Graphical Representation of Configuration 1 of Table 3 with aligned AMR : ***Move the large red block diagonally from the top of the blue column to the top of the yellow column****, which is 5 spaces from the orange cube.*

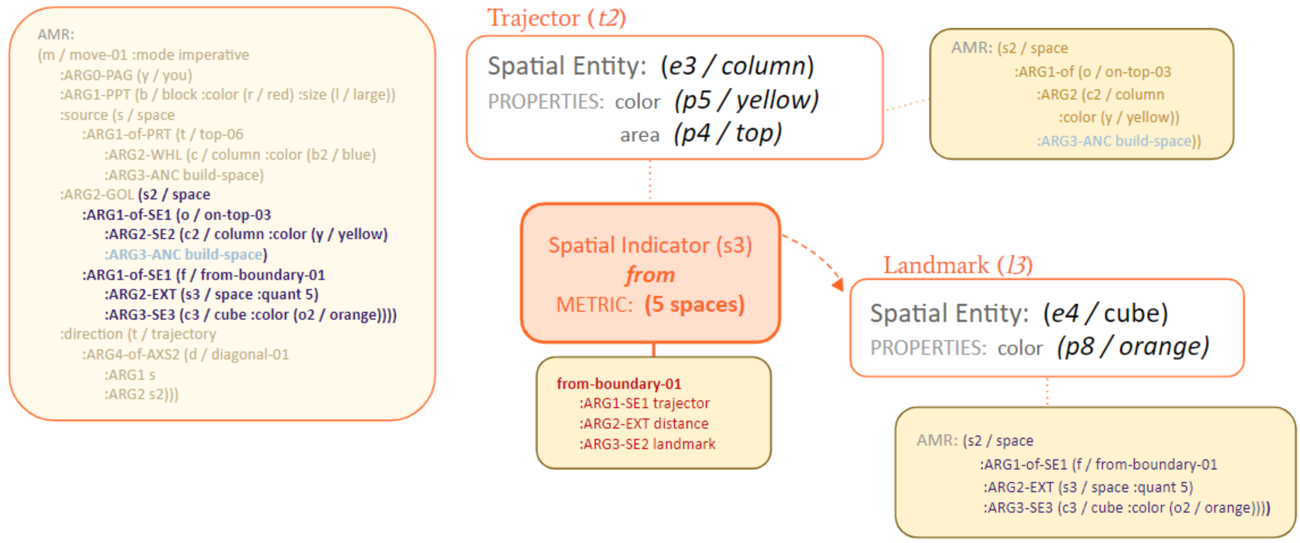

Figure 4: Graphical Representation of Configuration 2 of Table 3 with aligned AMR : *Move the large red block diagonally from the top of the blue column to the top of the yellow column,* ***which is 5 spaces from the orange cube.***

various formalisms such as RCC8, RCC5, etc (Randell et al., 1992). Though these formalisms have the advantage of providing reasoning systems, using them is challenging due to their differences in the level of specificity compared to natural language. Thus, we leave open the option of plugging in a fine-grained formalism in our representation by representing the qualitative type as a pair of general type and formal meaning, QT=<*G-type*,*F-meaning*>, where *G-type* $\in$ {*direction*,*distance*,*topology*} and F-meaning will be a task-specific formalism. We allow the assignment of multiple *G-type* and *F-type* to a single configuration to cover the gap between the level of specificity of language and formal models.

4. Automated Parsing

In this section, we demonstrate that the extended AMR scheme is learn-able by training a parser on the annotated examples. We present results for the state-of-the-art AMR parser (Zhang et al., 2019)(STOG parser) as a baseline for future follow-up works to parse the natural language sur-

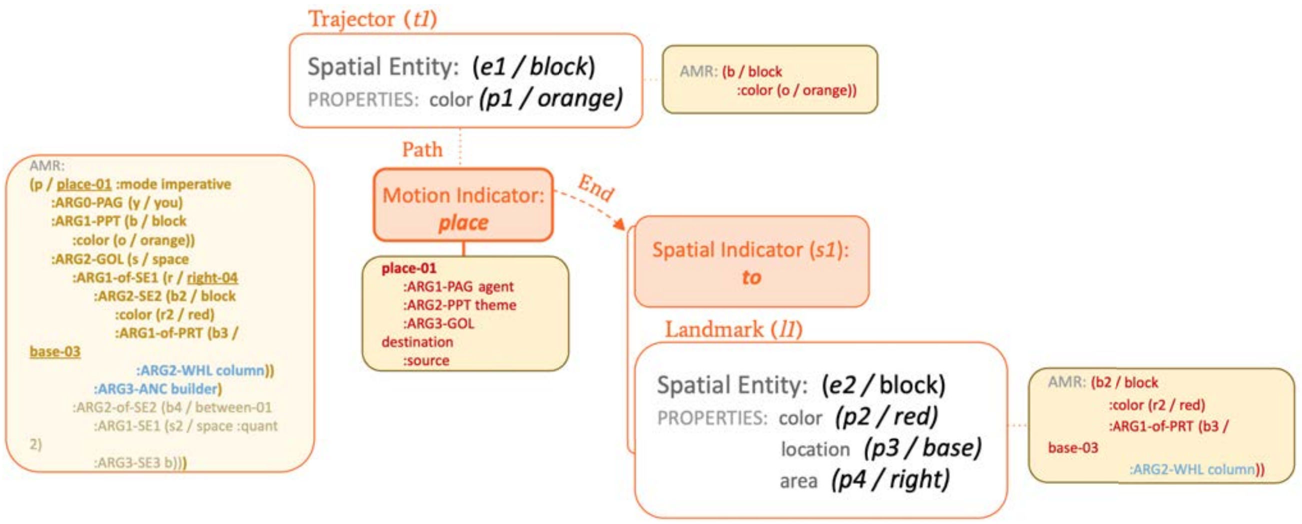

Figure 5: Graphical Representation of Configuration 1 of Table 4 with aligned AMR : ***Place an orange block to the right of the base red block*** *with two spaces in between.*

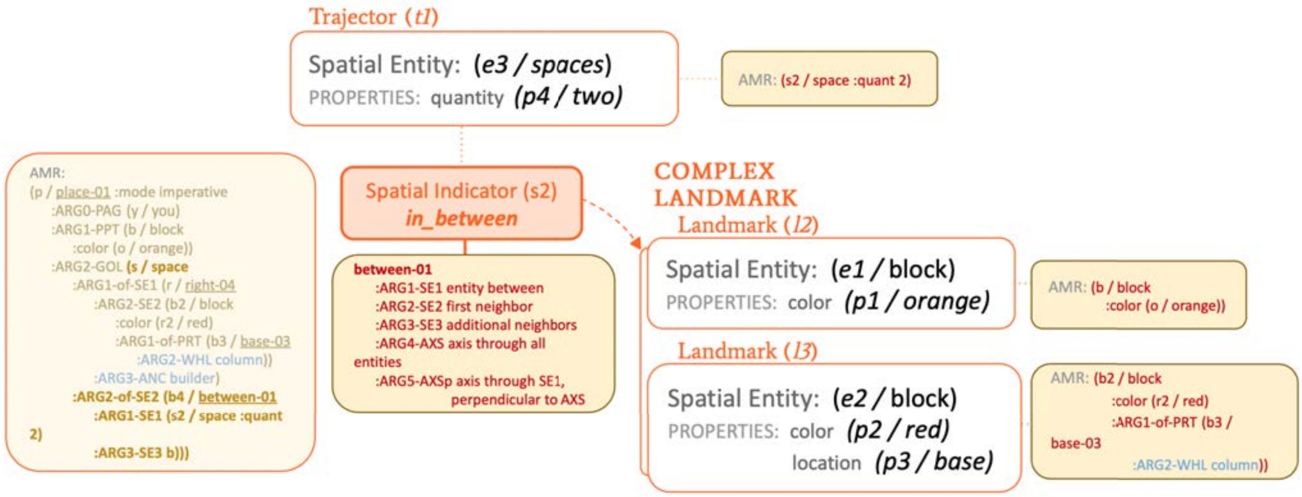

Figure 6: Graphical Representation of Configuration 2 of Table 4 with aligned AMR : *Place an orange block to the right of the base red block* ***with two spaces in between.***

face form into the AMR format. These are preliminary results with a scope for further improvement in the future. The statistics for the data split are shown in Table 5. We achieved an F1 score (calculated through triplet matches) of 66.24% on the test set after training on the filtered training set and validating on the filtered dev set. The parser is trained from scratch instead of relying on a pre-trained version of it, since the domain of LDC2017T10 data which STOG parser was reported on is significantly different than the Minecraft data — we found several preliminary fine-tuning results are not as good as the version trained from scratch. Here we show the predicted AMR output for the "bell" construction from the Minecraft dataset (Figure 9) , which is one of the more challenging cases:

```
(vv1 / place-01
    :ARG0 (vv3 / you)
    :ARG1 (vv4 / block
      :ord (vv5 / ORDINAL_ENTITY_1
        :range-start (vv6 / thing
          :ARG1-of (vv7 / bottom-03
```

	total	used
train	7954	4850
dev	933	604
test	862	583

Table 5: Total number of sentences and the number of sentences used for training, validation and test purposes

```
            :ARG2 (vv8 / bell))
          :ARG1-of (vv9 / middle-01
            :ARG2 (vv10 / bell))
          :ARG1-of (vv11 / middle-01
            :ARG1 vv4))))
    :mode imperative)
```

which predicts correctly most of the important components of the actual gold AMR.

```
(p / place-01 :polite + :mode imperative
    :ARG0 (y / you)
    :ARG1 (b / block :quant 1
```

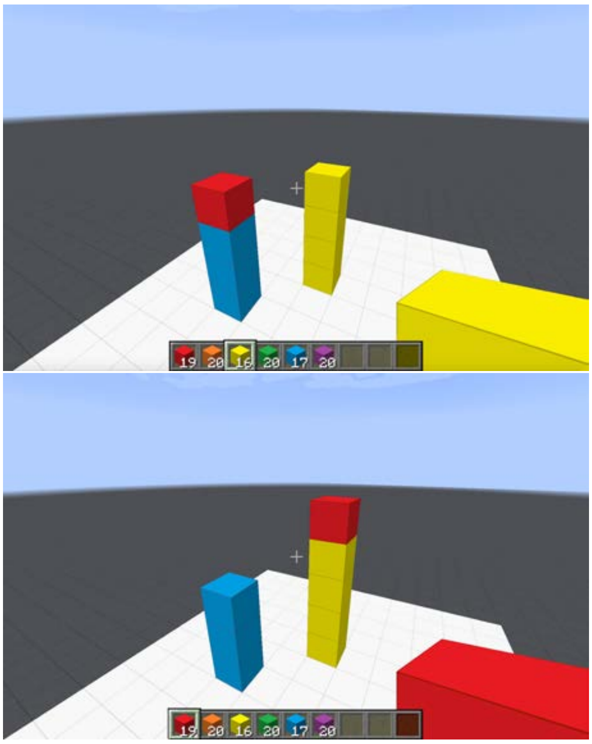

Figure 7: Move the large red block diagonally from the top of the blue column to the top of the yellow column ...

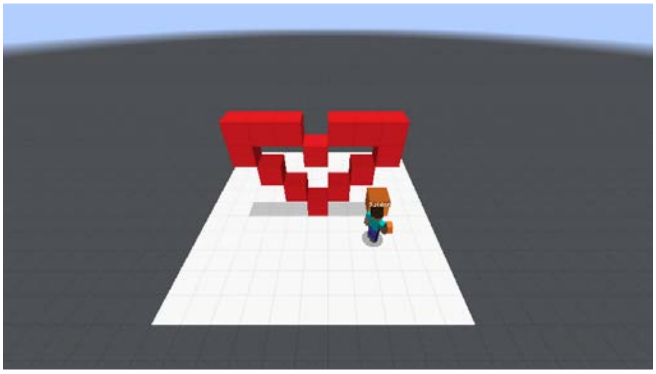

Figure 8: Place an orange block to the right of the base red block with two spaces in between.

```
                                    :color(y2/yellow))
       :ARG2 (s / space
          :ARG1-of (m / middle-01
             :ARG2 (c / composite-entity
                :ARG1-of (b4 / bottom-03
                   :ARG2 (b3 / bell))))))
```

5. Universal Applicability of SPATIAL CONFIGURATION scheme

We emphasize that although we have annotated the Minecraft corpus with spatial AMR to highlight the efficacy in one extremely spatially-challenging domain, we can use the proposed scheme to represent the spatial aspects of any domain. For illustration, we present an example from the NVLR dataset (Figure 10), which is a TRUE statement: There is a blue square closely touching the bottom of a box.

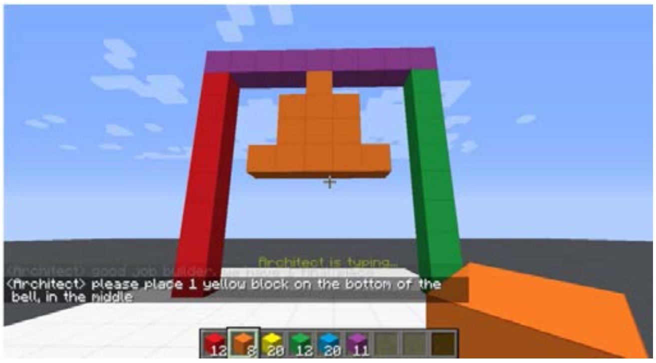

Figure 9: please place 1 yellow block on the bottom of the bell, in the middle.

We also show the detailed annotation in Table 6 and the graphical representation with an aligned AMR in Figure 11.

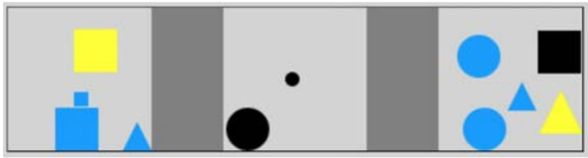

Figure 10: NLVR example : There is a blue square closely touching the bottom of a box.

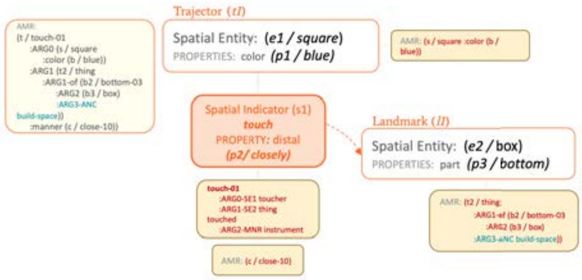

Figure 11: Graphical Representation of the components of the spatial configuration with the aligned AMR for Figure 10 :*There is a blue square closely touching the bottom of a box.* (zoom in for more clarity)

We also present another example from BLOCKS (Figure 12): Move the Mercedes block to the right of the Nvidia block. The annotation is shown in Table 7 and the graphical representation with aligned AMR in Figure 13.

6. Related Representations

The presented scheme is related to previous spatial annotation schemes such as ISO-space (Pustejovsky et al., 2011; Pustejovsky et al., 2015) and SpRL (Kordjamshidi et al., 2010; Kordjamshidi et al., 2012; Kolomiyets et al., 2013; Kordjamshidi et al., 2017). SpRL is based on holistic spatial semantics (Zlatev, 2003) and takes both cognitive-linguistic elements and spatial calculi models into account to bridge natural language and formal spatial representations that facilitate reasoning. The result is a light-weight spatial ontology compared to more extensive, linguistically motivated spatial ontologies such as GUM (Bateman et al., 2010; Hois and Kutz, 2008) or type theory motivated (Dobnik and Cooper, 2017) ones. ISO-Space scheme consid-

Figure 12: Before (top) and After (bottom) screenshots of the execution of the instruction : *Move the Mercedes block to the right of the Nvidia block.*

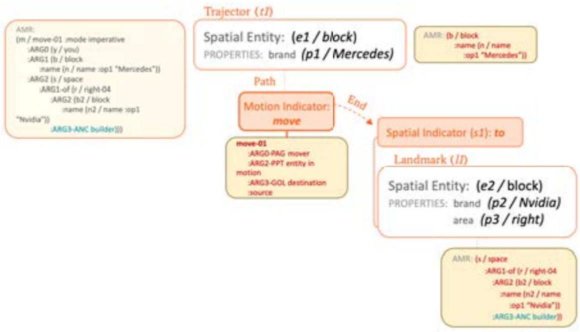

Figure 13: Graphical Representation of the components of the spatial configuration with the aligned AMR for Figure 12 : *Move the Mercedes block to the right of the Nvidia block.* (zoom in for more clarity)

Spatial Entities		
id	head	properties $\{name = span\}$
e1	square	$\{col = blue\}$
e2	box	$\{part - of = bottom\}$

	Configuration 1
tr	$< t1, e1 >$
lm	$< l1, e2 >$
sp	$<s1, touching$ $degree = closely >$
m	NULL
path	NULL
FoR	$< l1, intrinsic >$
v	$first - person$
QT	$<$topological, EC$>$

Table 6: Relational representation of spatial entities and configurations in the sentence: *There is a blue square closely touching the bottom of a box.*

ers very fine-grained properties of spatial objects based on external information and ontologies Dynamic spatial relations are extensively investigated in ISO-space and its recent extensions (Lee et al., 2018). Our proposed representation extends ISO-space and SpRL by decomposing complex spatial descriptions into modular configurations that are easier to formalize, ground and visualize. This facilitates incorporation of additional ISO-space concepts, extending our coverage to more complex and varied motion events. These configurations can be composed during grounding via shared indicators and arguments. Composition rules are domain dependent and formalizing them for a specific domain is future work. We represent spatial entities independently from the configurations. Entities can play

Spatial Entities		
id	head	properties $\{name = span\}$
e1	block	$\{brand = Mercedes\}$
e2	block	$\{brand = Nvidia,$ $area = right\}$

	Configuration 1
tr	$< t1, e1 >$
lm	$< l1, e2 >$
sp	$<s1, to >$
m	$<m1, move >$
path	$< implicit, begin >$ $< l1, s1, end >$
FoR	$< l1, relative >$
v	first-person
QT	$<$directional, relative$>$

Table 7: Relational representation of spatial entities and configurations in the sentence: *Move the Mercedes block to the right of the Nvidia block.*

various roles inside each configuration and the indicators can have properties. Our representation of motion along a path is more expressive than ISO-space by allowing multiple frames-of-references with respect to each constituent landmark connection in a path. The spatial configuration integrates these components and we tie them to AMR, enabling representation of spatial semantics as a subgraph of the complete semantic representation of the sentence; the first integration of spatial semantics with a rich general meaning representation scheme (AMR).

7. Conclusion

We presented a new spatial representation language and showed how it can compactly represent the spatial aspect of complex sentences that were challenging for existing schemes. We also integrated the scheme with AMR to achieve a richer meaning representation language. Further, we annotated sentences from the Minecraft corpus and trained an automatic parser to convert sentences into the extended AMR. The spatial configuration schema gives a succinct way to represent the spatial aspects of the full AMR annotation which we depict through examples. The adoption of the spatial representation scheme and the extended AMR as general-purpose tools by the research community can be beneficial particularly in spatially involved domains.

8. Acknowledgements

This work was supported by Contract W911NF-15-1-0461 with the US Defense Advanced Research Projects Agency (DARPA) and the Army Research Office (ARO) and by NSF CAREER award #1845771. Approved for Public Release, Distribution Unlimited. The views expressed are those of the authors and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

9. References

Banarescu, L., Bonial, C., Cai, S., Georgescu, M., Griffitt, K., Hermjakob, U., Knight, K., Koehn, P., Palmer, M., and Schneider, N. (2013). Abstract meaning representation for sembanking. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pages 178–186.

Bateman, J. A., Hois, J., Ross, R., and Tenbrink, T. (2010). A linguistic ontology of space for natural language processing. *Artificial Intelligence*, 174(14):1027–1071.

Bisk, Y., Shih, K. J., Choi, Y., and Marcu, D. (2018). Learning interpretable spatial operations in a rich 3d blocks world. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

Bonn, J., Palmer, M., Cai, J., and Wright-Bettner, K. (2019). Spatial amr: Expanded spatial annotation in the context of a grounded minecraft corpus. manuscript submitted for publication.

Dobnik, S. and Cooper, R. (2017). Interfacing language, spatial perception and cognition in type theory with records. *Journal of Language Modelling*, 5(2):273–301.

Gildea, D. and Palmer, M. (2002). The necessity of parsing for predicate argument recognition. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, pages 239–246. Association for Computational Linguistics.

Goldman, O., Latcinnik, V., Nave, E., Globerson, A., and Berant, J. (2018). Weakly supervised semantic parsing with abstract examples. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1809–1819. Association for Computational Linguistics.

Hois, J. and Kutz, O. (2008). Counterparts in Language and Space - Similarity and $\mathcal{S}$-Connection. In Carola Eschenbach et al., editors, *Proceedings of the International Conference on Formal Ontology in Information Systems (FOIS-2008)*, pages 266–279, Amsterdam. IOS Press.

Hu, R., Andreas, J., Rohrbach, M., Darrell, T., and Saenko, K. (2017). Learning to reason: End-to-end module networks for visual question answering. *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 804–813.

Hudson, D. A. and Manning, C. D. (2019). Gqa: a new dataset for compositional question answering over real-world images. *arXiv preprint arXiv:1902.09506*.

Kolomiyets, O., Kordjamshidi, P., Moens, M.-F., and Bethard, S. (2013). Semeval-2013 task 3: Spatial role labeling. In *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, pages 255–262, Atlanta, Georgia, USA, June.

Kordjamshidi, P., van Otterlo, M., and Moens, M.-F. (2010). Spatial role labeling: task definition and annotation scheme. In Nicoletta Calzolari, et al., editors, *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC'10)*, pages 413–420.

Kordjamshidi, P., Bethard, S., and Moens, M.-F. (2012). SemEval-2012 task 3: Spatial role labeling. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval)*, volume 2, pages 365–373.

Kordjamshidi, P., van Otterlo, M., and Moens, M.-F. (2017). Spatial role labeling annotation scheme. In N. Ide James Pustejovsky, editor, *Handbook of Linguistic Annotation*. Springer Verlag.

Krishnaswamy, N., Friedman, S., and Pustejovsky, J. (2019). Combining deep learning and qualitative spatial reasoning to learn complex structures from sparse examples with noise. In *Proceedings of the thirty-third AAAI Conference on Artificial Intelligence*.

Lee, K., Pustejovsky, J., and Bunt, H. (2018). The revision of iso-space,focused on the movement link. In *Proceedings 14th Joint ACL - ISO Workshop on Interoperable Semantic Annotation*, pages 35–44. Association for Computational Linguistics.

Liu, W., Li, S., and Renz, J. (2009). Combining RCC-8 with qualitative direction calculi: algorithms and complexity. In *IJCAI*.

Mani, I., Hitzeman, J., Richer, J., Harris, D., Quimby, R., and Wellner, B. (2008). SpatialML: annotation scheme, corpora, and tools. In N. Calzolari, et al., editors, *Proceedings of the Sixth International Language Resources and Evaluation (LREC)*. European Language Resources Association (ELRA).

Narayan-Chen, A., Jayannavar, P., and Hockenmaier, J. (2019). Collaborative dialogue in Minecraft. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5405–5415, Florence, Italy, July. Association for Computational Linguistics.

Palmer, M., Gildea, D., and Kingsbury, P. (2005). The proposition bank: An annotated corpus of semantic roles. *Computational linguistics*, 31(1):71–106.

Pustejovsky, J., Moszkowicz, J., and Verhagen, M. (2011). ISO-Space: The annotation of spatial information in language. In *ACL-ISO International Workshop on Semantic Annotation (ISA'6)*.

Pustejovsky, J., Kordjamshidi, P., Moens, M.-F., Levine, A., Dworman, S., and Yocum, Z. (2015). SemEval-2015 task 8: SpaceEval. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), 9th international workshop on semantic evaluation (SemEval 2015), Denver, Colorado, 4-5 June 2015*, pages 884–894. ACL.

Randell, D. A., Cui, Z., and Cohn, A. G. (1992). A spatial logic based on regions and connection. In *Proceedings of the 3rd International Conference on the Principles of Knowledge Representation and Reasoning, KR'92*, pages 165–176.

Suhr, A., Lewis, M., Yeh, J., and Artzi, Y. (2017a). A corpus of natural language for visual reasoning. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, volume 2, pages 217–223.

Suhr, A., Lewis, M., Yeh, J., and Artzi, Y. (2017b). A corpus of natural language for visual reasoning. In *ACL*.

Tenbrink, T. and Kuhn, W. (2011). A model of spatial reference frames in language. In Max Egenhofer, et al., editors, *Conference on Spatial Information Theory (COSIT'11)*, pages 371–390. Springer.

Wallgrün, J. O., Frommberger, L., Wolter, D., Dylla, F., and Freksa, C. (2007). Qualitative spatial representation and reasoning in the SparQ-Toolbox. In Thomas Barkowsky, et al., editors, *Spatial Cognition V Reasoning, Action, Interaction*, volume 4387, chapter 3, pages 39–58. Springer Berlin/Heidelberg.

Zhang, S., Ma, X., Duh, K., and Van Durme, B. (2019). AMR Parsing as Sequence-to-Graph Transduction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Florence, Italy, July. Association for Computational Linguistics.

Zlatev, J. (2003). Holistic spatial semantics of Thai. *Cognitive Linguistics and Non-Indo-European Languages*, pages 305–336.