Spatial AMR: Expanded Spatial Annotation in the Context of a Grounded Minecraft Corpus

Julia Bonn, Martha Palmer, Jon Cai, Kristin Wright-Bettner

University of Colorado at Boulder

Boulder, Colorado, USA

{julia.bonn, martha.palmer, jon.z.cai, kristin.wrightbettner}@colorado.edu

Abstract

This paper presents an expansion to the Abstract Meaning Representation (AMR) annotation schema that captures fine-grained semantically and pragmatically derived spatial information in grounded corpora. We describe a new lexical category conceptualization and set of spatial annotation tools built in the context of a multimodal corpus consisting of 185 3D structure-building dialogues between a human architect and human builder in Minecraft. Minecraft provides a particularly beneficial spatial relation-elicitation environment because it automatically tracks locations and orientations of objects and avatars in the space according to an absolute Cartesian coordinate system. Through a two-step process of sentence-level and document-level annotation designed to capture implicit information, we leverage these coordinates and bearings in the AMRs in combination with spatial framework annotation to ground the spatial language in the dialogues to absolute space.

Keywords: AMR, Spatial Relations, PropBank, SRL, Semantic Roles, Minecraft, Frame of Reference, Annotation

1. Introduction

This paper presents a spatial addendum to the Abstract Meaning Representation (AMR) semantic annotation schema (Banarescu et al., 2013) with accompanying spatial conceptualization, PropBank rolesets (Palmer et al., 2005), and grounded annotated corpus. AMR in its current form represents sentences as directed, rooted, acyclic graphs that capture core surface-level semantics while abstracting away from syntactic specificity. While AMR has not previously approached spatial semantics with much detail (Bonial et al., 2019), the implicit argument marking capability introduced by O'Gorman et al. (2018b) makes AMR an especially good fit for annotating spatial language. Spatial relations are not only semantically complex, they leverage pragmatically essential knowledge about the spatial characteristics of entities and their frameworks to map from linguistic expression onto space and from one spatial framework to another.

This addendum takes the fine-grained spatial semantics and object grounding strategies of previous schemata (Pustejovsky, 2017; Gotou et al., 2016; Pustejovsky and Krishnaswamy, 2016; Tellex et al., 2011) and folds them into multi-sentence AMR (MS-AMR). The result is a comprehensive annotation tool that can handle fine-grained explicit and implicit nested spatial relationships that are grounded in quantified space and merged fluidly with event dynamics. Because the spatial annotations are incorporated into the domain-general AMR graphs, this approach also captures information about how spatial relations are expressed in the context of whole sentences and overall discourse.

The new set of tools and practices we present span single sentence and multi-sentence annotation. At the single sentence level, we propose a new set of general semantic frames and roles as well as relation-specific rolesets. The rolesets represent the practical application of the new spatial conceptualization we also present here. The conceptualization and rolesets target lexical units from diverse parts of speech (now including prepositions and adverbs) and account for extrinsic relations and intrinsic properties having to do with location, orientation, configuration, extent, direction, topology, and especially frame of reference (FoR) (Levinson, 2003; Zlatev, 2010; Levelt, 1996). At the multi-sentence level, we propose new layers of coreference annotation and bridging that capture implicit spatial knowledge and aid in grounding. An important addition to MS-AMR is the inclusion of an existential dummy AMR graph that sets the configurational stage for the spatial entities represented in the dialogue. This dummy AMR defines specific spatial frameworks for each entity and the environment and describes how these frameworks map together. Accurately representing how frameworks map onto each other and onto language is an essential step in converting spatial language to language-independent spatial representations that are used by NLP systems downstream (see Dan et al., (2019) for discussion of downstream strategies). While Spatial AMR is intended to be adaptable to other environments, we present it here in the specific context of our corpus of Minecraft structure-building dialogues (Narayan-Chen et al., 2019) as an example of its specificity and range. This annotated corpus is being used to train a state-of-the-art semantic parser (Zhang et al., 2019), for which we present preliminary baseline results.

2. AMR

2.1 Single Sentence and Multi-sentence AMR

AMR annotation now occurs in two passes, one for single sentence annotation and another for multi-sentence (O'Gorman et al., 2018b). During single sentence annotation, AMR depicts each sentence as a series of nested predicate argument structures (Banarescu et al., 2013). Predicate-specific argument structures come from PropBank (Palmer et al, 2005), which disambiguates eventualities into senses (rolesets) with shared semantics and associated semantic roles given as numbered arguments. In PropBank, these numbered arguments are each associated with a three-letter semantic function tag,

furthering the semantic specificity of each role (O'Gorman et al., 2018a; Bonial et al., 2014; Bonial et al., 2015). A single roleset groups together etymologically related aliases represented as verbs, nouns, adjectives, light verb constructions, and other multi-word expressions (MWEs). AMRs are thus ambiguous for part of speech in their predicates, and a single AMR may accurately represent a variety of semantically similar sentences. AMRs are also ambiguous for tense, aspect, and discourse structure and are limited in their representation of quantification and scope, although expansions in these areas are currently underway (Pustejovsky et al., 2019; Donatelli et al., 2019; Myers and Palmer, 2019; Van Gysel et al., 2019; and Vigus et al., 2019).

Numbered arguments that aren't represented explicitly in a sentence are left unannotated during the single sentence pass. In the multi-sentence pass, they are brought back into the graphs and marked as potential implicit arguments, available for inclusion in co-reference identity chains and set/member or part/whole bridges. In this way, implicit and explicit information are brought together for a more complete meaning representation without conflating the two.

2.2 Spatial Limitations of AMR

Spatially speaking, AMR's coverage has been limited to a subset of spatial properties and elements that tend to come up as arguments of eventualities. Motion path—both fictive and literal (Talmy, 1996)—and location are well represented because they appear frequently as predicate arguments and general modifiers (**:direction**, **:source**, **:destination**, **:location**, **:path**, and **:extent**). General semantic roles **:consist-of**, **:part** are also often spatial in nature. Properties that don't fall within the scope of any of these categories end up lumped under the general modifier **:mod**.

Some spatial relations that can be lexicalized as adjectives have existing rolesets, for example the original roleset for 'right' in (1):

> (1) **right-04** *be located on the right side*
> **ARG1:** theme, entity on the right
> **ARG2:** to the right of

Of course, some of these rolesets just as easily cover complex prepositional and adverbial variants, e.g. *to-the-right-of*-mwp and *right*-adv. While that was not the intention of the schema, once these rolesets exist, they tend to be used intuitively by annotators anywhere they fit semantically. This creates a discrepancy between spatial prepositions and adverbs that have like adjectival counterparts and those that do not.

Spatial prepositions are omitted if their meaning is adequately conveyed through **:location** or motion path roles (2a); otherwise, they are included in annotation with one or more **:op** roles for their complements (2b).

> (2) a. *I'm at the library*
> (i / i
>   :location (l / library))
> b. *I'm behind the library*
> (i / i
>   :location (b / behind
>   :op1 (l / library)))

Spatial relations involving a directional difference in LOCATION between two entities *se1* and *se2* are annotated with the general frame **relative-position**:

> (3) **(**s / se1 :location (r / **relative-position**
>   :op1 (s2 / se2)
>   :quant (d / distance-quantity)
>   :direction (d2 / direction)))

While this frame provides adequate argument structure for sentences like *the town is 4 miles north of the forest*, it can't accurately treat differences in other intrinsic spatial properties like ORIENTATION. A sentence like *add two 45° to what you just made* isn't annotatable with current tools.

Overall, the **prep** :op1 treatment and **relative-position** frame suffice for spatially-light corpora. For a grounded, multimodal corpus that emphasizes spatial relationships, they are inadequate. Crucially, they cannot accommodate polysemy or FoR. Even the roleset for 'right' in (2) omits an argument for the entity whose right it is (the anchor), a surprise given that 'right' and 'left' are the rare predicates that frequently take an anchor as an explicit role, e.g. *my right is the same as your left*. In section 3, we explain why these considerations are so vital in a grounded corpus.

3. The Minecraft Corpus

The corpus we draw from consists of hundreds of dialogues and accompanying grounding data elicited during a collaborative human-to-human structure-building task in the 3D virtual environment of Minecraft. These dialogues are part of a larger project undertaken under the Communicating with Computers (CwC) DARPA grant, the goal of which is to create an automatic agent capable of communicating back and forth with a human participant while successfully carrying out real-world spatial instructions (Narayan-Chen et al., 2017; and Narayan-Chen et al., 2019). Minecraft using the Malmo platform (Johnson et al., 2016) provides a particularly rich setting for semantic training data of this sort because it automatically tracks locations and orientations of entities in the environment according to an absolute coordinate system. The participants are immersed in the 3D space as they communicate, which means they have access to a wide range of spatial frameworks for their spatial reasoning strategies and FoR selection (Li and Gleitman, 2002). This range exceeds what is available from spatial description tasks and tasks that take place in two-dimensions, which are common in current corpora (Kordjamshidi et al., 2010; Suhr et al., 2017; Johnson et al., 2016; and Mani et al., 2008).

In each dialogue, participants collaborate in building one of a set of predesigned block structures. Some structures are abstract (i.e. intrinsically non-oriented objects), while other structures are representational models of animals, vehicles, letters, etc. (i.e. oriented objects). The variety of target structures used here elicit a range of FoR strategies. In each task, one participant plays the role of Builder and the other plays Architect. Only the Builder has an avatar in the space and can manipulate blocks. Only the Architect can see plans for the target structure. Conversation between them is unrestricted and includes instructions, requests for clarification, corrections, confirmations, etc. The Builder can move readily around

the environment but can only place blocks within the boundaries of an 11 x 11 white grid. The rest of the space extends toward an uninterrupted horizon in all directions with no further landmarks in sight.

The corpus is multi-modal. Each dialogue appears as a set of utterances and automatically generated representations of Builder actions given in the form *[Builder puts down/picks up a red block at X:0 Y:1 Z:0]*, where the coordinates refer to the absolute coordinate system used by Minecraft. Each is accompanied by screenshots from the Builder's and Architect's points of view, four additional fixed views, and orientation and location data for the Builder. Because the Architect either confirms or corrects each block placement, these dialogues provide definitive answers for the intended meaning of each of the Architect's instructions. When a miscommunication occurs, we often have enough data between all modalities to deduce which element of meaning failed.

To date, our annotation of this corpus extends past 5,000 dialogue sentences and 7,600 automatically generated Builder action sentences, for a total of 12,600+ individual AMRs and 185 full dialogues.

4. Spatial Conceptualization and Rolesets

The new inventory of spatial rolesets derives from the language used in these dialogues and gives special consideration to previously unframed or under-framed lexemes. This means that, as the newest additions to PropBank/AMR, spatial prepositions, adverbs, and multi-word expression rolesets are most prevalent, with spatial adjectives appearing most often with added roles. Fewer new verb/noun additions were needed beyond splitting rolesets that previously conflated caused motion with states. The new rolesets are divided and grouped into senses motivated by a new spatial conceptualization. The categories in the conceptualization aim to target elements of meaning that trigger an intuitive understanding of different senses in language users; these different senses can often be identified by their characteristic triggering of slightly different syntactic patterns and argument behavior, also identifiable by systems trained on the data. The categories in the conceptualization correspond to language-independent schematic predicates; each roleset points to a set of these entailed schematic predicates based on its categorization, and as such, propositions annotated using the rolesets can be converted directly into an accurate language-independent simulation of the spatial array the proposition describes.

4.1 Spatial Conceptualization

The conceptualization includes spatial relations that are eventualities and those that are not (*go* vs. *above*). Eventualities may be static or dynamic (extent *go* vs. motion *go*), and relations may express intrinsic properties of a single entity or event (*flat*) or extrinsic relationships between multiple entities (*parallel)*. Relations target four different basic elements of spatial meaning: LOCATION, ORIENTATION, CONFIGURATION, and EXTENT. EXTENT deals with dimensionality, measurement, and density, with either increasing or decreasing dynamic variants. CONFIGURATION refers to an arrangement of component entities considered holistically, as in *a row of blocks*. LOCATION refers to the overall coordinates of an

entity, whereas ORIENTATION characterizes how a particular face of an entity is aimed in a larger context. Relations highlighting each of these four elements may be described further in terms of DIRECTION, REGION and TOPOLOGY (Zlatev, 2010). DIRECTIONS are vectors characterized by the axes of a framework, and REGIONS are portions of a framework that often correspond to DIRECTIONS but may also be bounded by SCALE (*near* or *far*) or FORM (e.g. *around:* this entails not only **containment** TOPOLOGY but REGIONS defined by spheres and radii).
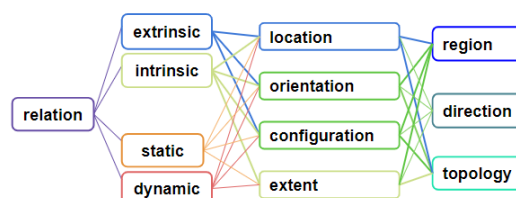

Figure 1: upper conceptualization

TOPOLOGY's primary distinction applies to extrinsic relations, focusing on whether the relationship between two primary spatial entities is **internal** or **external** (Tenbrink and Kuhn, 2011; Randell et al., 1992). For example, **external** CONFIGURATIONS involve components that are separable and may be interchangeable, like *chairs in a ring around a fire pit*, as opposed to **internal** CONFIGURATIONS like *fingers on a hand*. **External** relations further distinguish **containment, contact**, and **adjacency**.
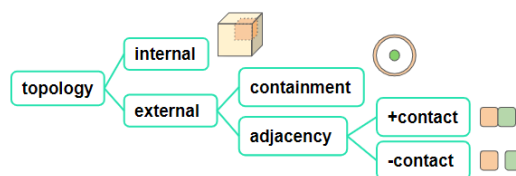

Figure 2: TOPOLOGY

DIRECTION is a key element of most of our prepositional relations, with **inward**, **outward** and **elliptical** trajectories, **vertical** vs **horizontal** contrast, and three axes labelled **up/down**, **left/right**, and **front/back**. We discuss direction in terms of Cartesian axis systems. Our convention is to assume for an animate entity a left-handed framework in which **up** corresponds to the positive y-axis, **front** corresponds to the positive z-axis, and **right** corresponds to the positive x-axis.
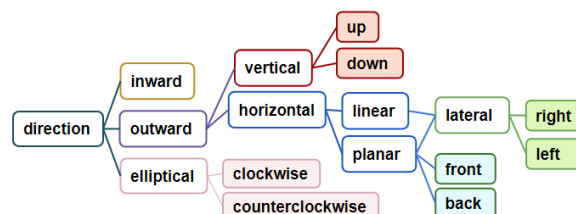

Figure 3: DIRECTION

4.2 New Rolesets

Spatial AMR has added 170 new or expanded relation-specific rolesets, many featuring prepositions and adverbs. As with the rest of the lexicon, these rolesets are ambiguous for part of speech. Rolesets that cover relations

typically expressed as prepositions conventionally include the prepositional complement as ARG2.

4.2.1 Image Schema and Semantic Roles

The rolesets leverage a new set of function tags that label spatial semantic and pragmatic roles. They correspond to the components of a spatial image schema (Figure 4).
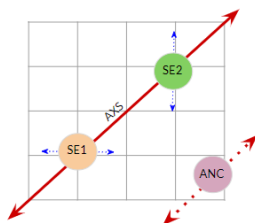

Figure 4: Spatial Image Schema with roles

In its most generalized form, this image schema includes two spatial entities (SE1, SE2), a directional path between them (AXS), and an anchoring entity who projects the spatial framework in which the relation holds meaning (ANC). SE1 and SE2 represent the two primary spatial entities when **external** TOPOGRAPHY is entailed. For **internal** relationships, PRT and WHL replace SE1 and SE2. Some rolesets include function tags for secondary axes (AXS1/AXS2, AXSp when perpendicular, AXSc when an axis of rotation). Others include tags for angle measurements (ANG), 2D planes in 3D space (PLN), scales for scalar relations (SCL), and sources (SRC), which have been split off from the PropBank direction (DIR) tag.

Function tags are syntax independent, as illustrated in (4). This approach provides more detailed semantic role information than that conveyed by the traditional terms figure/ground (Levinson, 2003), locatum/relatum (Levelt, 1996) and trajector/landmark (Lakoff, 1987). Certain senses of prepositions 'up', 'down', 'across', and 'over' take an AXS as opposed to an SE2 as their complement, and we preserve the semantic distinction to facilitate conversion to logical predicates at a later stage.

(4) a. [*John*]$_{SE1}$ *is up [the ladder]*$_{AXS}$ *from [Mary]*$_{SE2}$
    b. (up-03
        :ARG1-SE1 (p / person :name John)
        :ARG2-AXS (l / ladder)
        :ARG4-SE2 (p2 / person :name Mary)
        :ARG5-ANC [implicit: build-space])

    b. [*John*]$_{SE1}$ *is above [Mary]*$_{SE2}$ *on [the ladder]*$_{AXS}$
    c. (above-01
        :ARG1-SE1 (p / person :name John)
        :ARG2-SE2 (p2 / person :name Mary)
        :ARG3-ANC [implicit: build-space]
        :ARG4-AXS (l / ladder))

4.2.2 Entailments and Sense Division

The new rolesets offer several advantages. First, they allow us to treat polysemous senses individually while grouping together related aliases. Second, they provide a home for the fine-grained semantic and pragmatic roles we need to annotate in order to understand the context-dependent meaning of a relation within an absolute framework. Third, the schematic predicate entailments of

each roleset serve as an annotation short-cut by bundling semantics that would otherwise need to be annotated manually. (5) shows the roleset **on-top-03** along with its entailments[1]. This roleset covers aliases *on_top_of*-mwp, *on*-p, and *atop*-p.

(5) a. **on-top-03** *higher on a vertical axis, +contact*
      ARG1-SE1: entity above
      ARG2-SE2: entity below
      ARG3-ANC: anchor
      ARG4-AXS: axis

    b. (forall (SE1 SE2 f ANC)
      (iff (**on-top-03** SE1 SE2 f)
        (exists (a p1 p2 y1 y2)
          (and(**selfAnchoredFramework** f ANC)
            (**yAxis** a f)
            (**parallel** a AXS)
            (**atLoc** SE1 p1)(**atLoc** SE2 p2)
            (**yCoordinate** y1 p1 f)
            (**yCoordinate** y2 p2 f)
            (**lt** y2 y1 a)
            (**contact** SE1 SE2)
            (**externalTo** SE1 SE2)))))

By simply marking a relation with **on-top-03,** an annotator is automatically indicating the following: 1) ARG1 and ARG2 are spatial entities that are topologically discrete but in contact with one another; 2) directionally speaking, SE1 is higher than SE2, and this relationship holds meaning within a spatial framework f projected by ARG3-ANC; 3) within framework f, the line AXS between SE1 and SE2 parallels the y-axis; and 4) because SE1 and SE2 are in contact, the coordinates of the bottom surface of SE1 are equal to the coordinates of the top surface of SE2. Functionally speaking, in Minecraft's coordinate system, the bottom block of SE1 has a y value that is 1 greater than the y value of the top block of SE2. In contrast, the PRT/WHL roleset **top-06** is just like this except that it takes an **internalTo** relationship between its primary arguments that eliminates the need for the **contact** predicate. That roleset includes aliases *top*-v, *top*-j, *at_the_top_of*-mwp, and *in_the_top_of*-mwp.

Polysemous relations receive separate rolesets when two equally viable interpretations project different spatial entailments in the conceptualization. For example, **diagonal-01** and **diagonal-02** differ in that the former targets the spatial element LOCATION, while the latter targets ORIENTATION. **Diagonal-01** says that two entities are LOCATED such that a line drawn between them would be diagonal relative to some external framework. **Diagonal-02** says that two entities are ORIENTED such that a defining axis of one is diagonal relative to a defining axis of the other.

Relations like **on-top** and **above** are categorized as VERTICAL, meaning they always entail a relationship along the y-axis of whatever framework is selected under the ANC role. In absolute terms, these relations often take a geocentric UP interpretation (*into the sky*), but they may also refer to the UP portion of some other non-geocentric

---

framework that points in some other absolute direction. Regardless of what the direction is in absolute terms, because the entailment is always that the y-axis of f is being referenced, we consider these relations to have one sense each. In contrast, **over** is considered to have two separate senses with different entailments. O**ver-05** is categorized as VERTICAL (*the cloud is over the tree)* and o**ver-04** is categorized as HORIZONTAL (*the seat 3 seats over from me)*. More specifically, **over-05** entails a relationship along f's y-axis, while **over-04** entails a relationship along f's x-axis. The difference in entailments motivates the division into separate rolesets.

4.2.3 Anchor and Axis Roles

Rolesets don't entail anything specific about which spatial framework is being referenced, they simply provide an ANC argument slot so that FoR information can be annotated per instance. With ungrounded corpora, FoR annotation may be left unannotated. For grounded corpora like ours, annotating these arguments is essential. We need to know which framework a given utterance references, and we need to know how that framework was oriented, in the moment, in terms of the absolute space. How fine-grained this grounding needs to be is left up to the needs of the corpus. All directed rolesets receive ANC roles, including traditionally 'absolute' relations like **north**, which may be interpreted differently depending upon how the poles that orient the cardinal framework are defined. On Earth, magnetic north has a slightly different pole than true north. Uranus has 3 north poles: one defined in terms of the invariable plane of the solar system and the counter-clockwise rotation of the sun, one defined by the counter-clockwise rotation of Uranus itself, and a magnetic north pole that lies somewhere in the sun-based southern hemisphere[2]. In Minecraft, participants aren't aware of an absolute north, but sometimes use cardinal directions to mean more generally 'forward', 'back', 'right', and 'left'. For systems that wish to identify FoR by types such as intrinsic and relative, cross-referencing the identity of ANC against that of SE2 is one way to automatically generate those values.

The AXS role is frequently used to house a leaf node that serves as a hinge between motion and spatial relation rolesets (see example 7). It is also used for coreference to axis variables in the spatial frameworks of the dummy AMR during MS-AMR annotation.

5. Spatial AMR Annotation

5.1 Single Sentence: New General Tools

Spatial AMR adds new general semantic concepts, roles, entity types, and frames. First, two new general concepts, **space** and **trajectory**, aid in annotating concepts that are beneficial as leaves in the AMRs for structural reasons. **Space** describes a volume of space that may be occupied by a spatial entity. It is used frequently with **:source**, **:destination**, and **:location** roles. Relations that hold for a

volume of space hold for any entity that occupies that space. In (6), **space** is modified by one of the new entity types, **cartesian-coordinate-entity** which provides slots for Cartesian coordinates values. The **:framework** role allow us to co-refer to the **cartesian-framework-91** that projects the coordinates, described in detail in section 6.

(6) a. *[Builder puts down a red block at X:0 Y:1 Z:0]*
   b. (p / put-down-17
       :ARG0 (b / builder)
       :ARG1 (b2 / block :color (r / red))
       :ARG2 (**s / space**
          :location (**c / cartesian-coordinate-entity**
             :x 0 :y 1 :z 0 :framework [build-space])))

Trajectory gives a variable to a rich path (7) (Zlatev, 2010). Both **trajectory** and **space** can be modified by spatial rolesets and included in co-reference chains. Note that in (7), multiple spatial rolesets modify the **trajectory**. The interpretation of this AMR is that there is one direction that is simultaneously UPWARD and LEFTWARD; the direction is the average. Sequential interpretations would be represented as a series of separate motion events.

(7) a. *move the block 1 upward and toward the left*
   b. (m / move-01 :mode imperative
       :ARG0 (y / you)
       :ARG1 (b / block)
       :extent (d / distance-quantity :quant 1)
       :direction (**t / trajectory**
          :ARG4-of-AXS (u / up-03)
          :ARG4-of-AXS (l / left-20)))

In AMR, the -91 tag indicates a relation non-specific roleset with numbered arguments. Some new -91 frames correspond to new general semantic roles as reifications, while others capture more specific behavior. **Have-anchor-91** and **have-axis-91** reify **:anchor** and **:axis**, which are equivalent to the function tags of the same names. **:Anchor** is useful as a modifier for oriented spatial entities (*row* or *column)* that don't fall under roleset coverage but whose orientation may need grounding, as in (8).

(8) a. 6th column, 2nd row, from any side of the *square*
   b. (s / space
       :location (c / column :ord 6
          :anchor (s2 / side :mod (a / any)
             :part-of (s3 / square))
       :location (r / row :ord 2 **:anchor** s2))

Of the four basic spatial elements described in section 4.1, ORIENTATION and CONFIGURATION did not have general semantic roles, but do now. ORIENTATION has an additional frame that mirrors **relative-position** called **relative-orientation**, as discussed in example (3). **:Color** (**have-color-91**) streamlines color property annotation and breaks up the static/dynamic conflation of the **color-01** roleset. **:Size** (**have-size-91**) often accompanies **dimension-entity**, which is used for general dimensions like '2x6' where the values aren't specified as lengths, widths, heights, or depths. Note the new **:pl** + marker and CONFIGURATION frame in (9). The ability to mark plurals greatly improves our grounding abilities by improving quantification.

(9) a. the *3x3x5 boxes are in a row*
    b. (**h / have-configuration-91**
       :ARG1 (b / box **:pl** +
         :size (d / dimension-entity :value 3)
         :size (d2 / dimension-entity :value 3)
         :size (d3 / dimension-entity :value 5))
       :ARG2 (r / row))

Certain spatial relations are expressed through spatiotemporal metaphor. Take the sequence of instructions in (10):

(10) a. *a row of orange*
    b. *then 3*
    c. *then five*
    d. *and one more in the same direction*

The adverb 'then' would traditionally have been annotated with **:time**. However, the intended interpretation here is that the rows should be placed in a spatial sequence, not a temporal one. **Spatial-sequence-91** (11) now captures this behavior. This frame frequently links entities across sentence boundaries during multi-sentence annotation but can also be useful within a sentence, as in *now put down red, red, green, space, blue.*

(11) **spatial-sequence-91**
    ARG1-AXS: trajectory of the sequence
    ARG2-SE1: first entity in sequence
    ARG3-SE2: second entity
    ARG4-SE3: third entity
    etc.

In (10d), *in the same direction* is annotated under **:ARG1**, while the blocks indicated in (10a-d) fall under the subsequent ARGs according to the order in which they appear in the dialogue.

5.2 Multi-sentence: Partial Grounding and Speech Act Clusters

During multi-sentence annotation, we track coreference between grounded blocks as well as ungrounded blocks. Grounded blocks are specific blocks that have been placed, and ungrounded blocks are hypothetical block mentions that co-refer from sentence to sentence, but which are not yet tied to a specific block from the inventory. The identity chains are kept separate but bridged. In (12), the blocks in bold track a partially grounded identity chain. Within that, the block in (b) is grounded to (f). Coreference between spaces is handled in the same way.

(12) a. <Architect> **one more red** attached to that
    b. [Builder places a **red block** at X:1 Y:1 Z:0]
    c. <Architect> the red you placed
    d. <Builder> here?
    e. <Architect> no
    f. [Builder picks up a **red block** at X:1 Y:1 Z:0]
    j. [Builder places a **red block** at X:2 Y:1 Z:0]
    l. <Architect> yup!

We also create set/member bridging for clusters of dialogue and action sentences that relate to a single instruction. All sentences in (12) are grouped together as members under the instruction given in (a) because they all relate to accomplishing (a).

6. Annotation of Spatial Frameworks

The most radical convention added by Spatial AMR is the inclusion of the document-level dummy AMR in which spatial frameworks are defined and oriented in terms of the space and each other. These spatial frameworks are represented using the new **cartesian-framework-91**, shown in (13).

(13) **cartesian-framework-91**
    ARG1-ANC: spatial entity projecting framework
    ARG2: x-axis
    ARG3: y-axis
    ARG4: z-axis
    ARG5: origin
    FR: frame of reference type
    hand: handedness of framework

This frame does not in itself dictate how it should be used. Axes are left unspecified for direction and polarity, and two-axis frameworks are accommodated by simply leaving one of the axis roles unmarked. It is also not limited to the dummy AMR; the Minecraft corpus contains several dialogues in which the Builder and Architect attempt to set up their own Cartesian framework as an instructional tool, and this frame provides all arguments needed in such scenarios including a role for the origin. Outside of these rare explicit uses, however, this frame's predominant function is to bring essential implicit FoR knowledge into the graphs via the ANC roles in directed rolesets, either by explicitly mentioning the relevant framework during single sentence annotation, or by linking the implicit ANC role slot to the appropriate framework in the dummy AMR during MS-AMR.

The advantage of treating FoR frameworks in this way is that we can define multiple frameworks for a single entity, we can clearly define the spatial configuration of each framework (how many axes it contains and how the polarity of each axis is configured relative to the other axes), we can describe how the frameworks map onto each other spatially, and by including all of this information in AMR form, we have variables for each concept that are available for coreference annotation during MS-AMR. Defining multiple frameworks for a single entity is especially helpful in clearing up ambiguity that arises when more than one framework can be identified in terms of a FoR type/anchor pair. Take two opposite interpretations of *in front of the block*. In one, we get 'on the other side of the block relative to me', where the "*front*" is codirectional with my FRONT; in the other we get 'between me and the block', where the "*front*" is the block's FRONT as defined by my location relative to the block. The problem is that both interpretations would be classified as relative FoR with me as an anchor (1st person), yet their absolute interpretations are diametrically opposed. Using our approach, we point the former to the Builder's intrinsic framework (which is not the same as saying this relation is categorized under the intrinsic FoR type), and the latter to a Builder-relative framework projected by the block. The flexibility of this approach also supports annotation of atypically anchored frameworks like those tied to direction of motion and function. As AMR expands cross-linguistically, the ability to accommodate the different FoR rotation and projection patterns will become even more useful (Levinson, 2003).

6.1 Dummy AMR

The dummy AMR includes at least one **cartesian-framework-91** for each spatial entity involved in the dialogues. We generalize away from individual blocks, letting each block inherit a framework from the larger structure it belongs to. Example (14) shows frameworks for a block structure model of a cat lying on its side and the build space, as well as predicates showing how these frameworks map together:

```
(14) :snt1 (c / composite-entity
        :ARG1-of (c2 / cartesian-framework-91
            :ARG2 (x / x-axis)
            :ARG3 (y / y-axis)
            :ARG4 (z / z-axis)
            :hand (r / right-handed)
            :FR (r2 / relative-to-builder))
        :ARG1-of (c3 / cartesian-framework-91
            :ARG2 (x2 / x-axis)
            :ARG3 (y2 / y-axis)
            :ARG4 (z2 / z-axis)
            :hand (l / left-handed)
            :FR (i2 / intrinsic))
        :configuration (c7 / cat))
    :snt2 (b / build-space
        :ARG1-of (c4 / cartesian-framework-91
            :ARG1 (x3 / x-axis)
            :ARG2 (y3 / y-axis)
            :ARG3 (z3 / z-axis)
            :hand (r3 / right-handed)
            :FR (a / absolute))
    :snt3 (c5 / codirectional-01
        :ARG1-AXS1 y2
        :ARG2-AXS2 x3)
    :snt4 (c6 / codirectional-01
        :ARG1-AXS1 z2
        :ARG2-AXS2 z3)
```

This says that the structure has two frameworks: one is a right-handed relative-to-builder framework, which means that its FRONT (positive z-axis) is aimed at the Builder at all times, and that if the Builder is facing this framework in return, their respective RIGHTs point in the same direction; the second is intrinsic to the cat, with directions anchored to the cat's body parts and perspective. **(c5)** says that the top of the cat is pointing towards the positive x-axis in the absolute framework, and **(c6)** says that the cat's front is pointing towards the absolute positive z-axis.

6.2 Minecraft Frameworks

After annotating 185 dialogues, we find that we can convert from the directed rolesets to the correct absolute direction/coordinates by referencing a baseline of five frameworks in the ANC roles. :

The **absolute 3D build-space** framework used by Minecraft: this framework is used for VERTICAL relations with *geocentric* interpretation (*into the sky)*.

A **Builder-intrinsic** framework in which 'forward' is the direction the Builder is facing.

A **relative-to-builder 2D framework for the white square** that forms the ground surface of the build-space: this framework is used for VERTICAL relations with a 3D-*horizontal* interpretation; here, **up** is necessarily equivalent to what is **forward** for the Builder.

A **relative-to-builder framework for the structure and its component blocks** in which the FRONT is the side facing the Builder no matter where the Builder is located.

An **intrinsic framework for oriented representational structures**: humanoids, animals, vehicles, graphemes, etc. have *tops* or *fronts* that will define positive z or positive y axes.

Some of these frameworks are left-handed frameworks and some are right-handed depending on the polarity of the x-axis. The build-space has a right-handed framework (+z → 'South', +y → 'up', +x → 'east'), while the Builder has a left-handed framework (+z → 'front', +y → 'top', +x → 'right'). The build-space's framework is independently defined, but frameworks that are Builder-relative are, by definition, enantiomorphic to the Builder's (for English). Some structures' intrinsic frameworks are right-handed and others' are left-handed, possibly hinging on how likely a speaker is to imagine embodying that entity. Because we define each framework individually and orient its axes in relation to the absolute space in the dummy AMR, none of these variations pose a problem for deterministic mapping from one framework to another by the planner; in fact, they streamline the process.

6.3 Mapping Rolesets to Absolute Predicates

Imagine that the sentence in (15) is included in the dialogue that corresponds to the dummy AMR in (14). The annotator decides that the FoR being used is the cat's intrinsic framework, and so in the AMR for this sentence, the :ARG3-ANC role for **on-top-03** is annotated to point to **(c3)** in the dummy AMR (the cat's intrinsic cartesian-frameworks).

```
(15) a. a green block is on top of the blue block
     b. (o / on-top-03
         :ARG1-SE1 (b / block :color green)
         :ARG2-SE2 (b2 / block :color blue
             :location (c / cart-coord-ent :x 1 :y 3 :z 1)
         :ARG3-ANC [implicit: c3]
         :ARG4-AXS [implicit: y2])
```

In order to convert the AMR in (15) to absolute coordinates, we simply combine the information from the roleset's entailments, the framework mappings in the dummy AMR **(c5, c6)**, and the necessary location/bearing data collected by Minecraft. The entailments for **on-top-03** tell us that we are concerned with the y-coordinate values of the two blocks, and that the y-value of SE1 will be greater than that of SE2 by 1. For other applications, the exact value difference might be factored differently, but Minecraft values are all integers. Importantly, this is the y-value within the ANC framework, not the absolute framework. The dummy AMR tells us that the y-axis **(y2)** in the cat's intrinsic framework is codirectional with the absolute framework's x-axis **(x3)**, so we know that we are actually dealing with the absolute x-coordinate values. Minecraft provided us with the absolute cartesian-coordinate-entity values for the blue block when it was placed, listed in **(c)** in (15). Taken all together, we know that if the absolute coordinates of the blue block are (1, 3, 1), then the absolute coordinates of the green block must be (2, 3, 1). The dummy AMR should provide enough information to convert deterministically between axes in any of the frameworks.

7. Semantic Parser and Results

We are training a state-of-the-art AMR parser (Zhang et al., 2019) (STOG parser) on the Minecraft spatial AMRs. The preliminary results we present here form a baseline for follow-up work in parsing the natural language surface form into AMR format. We expect to continue to see improvement as we add to and fine-tune the dataset. Table 1 shows the data splits statistics:

	total	used
train	7954	4850
dev	933	604
test	862	583

Table 1: sentences total and sentences used for training, validation, and test purposes

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 differs significantly from the Minecraft data. We found several preliminary fine-tuning results to be not as good as the version trained from scratch. (16) shows one of the more challenging corpus sentences with its gold (16b) and parser-predicted (16c) graphs. The parser correctly predicts most of the important components of the actual gold AMR.

(16) a. *please place 1 yellow block on the bottom of the bell, in the middle*

b. (p / place-01 :polite + :mode imperative
   :ARG0 (y / you)
   :ARG1 (b / block :quant 1 :color (y2 / yellow))
    :ARG2 (s / space
      :ARG1-of (m / middle-01
        :ARG2 (c / composite-entity
          :ARG1-of (b2 / bottom-03
            :ARG2 (b3 / bell))))))

c. (vv1 / place-01 :mode imperative
    :ARG0 (vv3 / you)
    :ARG1 (vv4 / block)
      :ord (vv5 / ORDINAL_ENTITY_1
        :range-start (vv6 / thing
          :ARG1-of (vv7 / bottom-03
            :ARG2 (vv8 / bell))
          :ARG1-of (vv9 / middle-01
            :ARG2 (vv10 / bell))
          :ARG1-of (vv11 / middle-01
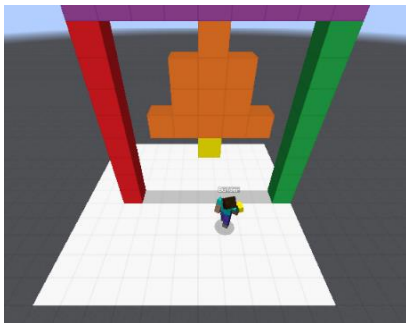   :ARG2 vv4)))))

Figure 5: Screenshot associated with (16)

Inter-annotator agreement was measured on a selection of nine dialogues taken from the 4-10 dataset (dialogues elicited on April 10, 2018, halfway through the elicitation period), a total of 282 sentences, with an overall smatch score of P: 0.88. Automatically-generated builder action AMRs and dummy AMRs were not included. Four different Builders and four different Architects participated in these nine dialogues, with five of them having participated in tasks on previous days.

8. Conclusion

In this paper, we have presented an extension to AMR that leverages new rolesets, general tools, and the implicit argument capturing capabilities of MS-AMR to represent fine-grained spatial semantic and pragmatic information in the context of wider, domain-general discourse. The tools we have presented, especially those that enable identification of specific spatial frameworks tied to entities in the text, allow us to ground the linguistic layers of multimodal corpora to quantified space. The spatial language patterns that have emerged through annotation of the Minecraft corpus are tied to the Minecraft environment and associated spatial placement tasks. We look forward to applying Spatial AMR to more diverse corpora in the future.

9. Acknowledgments

10. Bibliographical References

Banarescu, L., Bonial, C., Cai, S., Georgescu, M., Griffitt, K., Hermjakob, U., Knight, K., Palmer, M., and Schneider, N. (2013). Abstract meaning representation for sembanking. In S. Dipper, M. Liakata, & A. Pareja-Lora (conference chairs), Proceedings of the Seventh Linguistic Annotation Workshop & Interoperability with Discourse (LAW VII & ID), pages 178—186, Sofia, Bulgaria, August. Association for Computational Linguistics (ACL).

Bonial, C., Donatelli, L., Ervin, J., and Voss, C. (2019). Abstract meaning representation for human-robot dialogue. In G. Jarosz et al., editors, Proceedings of the Society for Computational Linguistics (SCiL) 2019, pages 236—246, New York, USA, January.

Bonial, C., Bonn, J., Conger, K., Hwang, J., Palmer, M. (2014). PropBank: semantics of new predicate types. In N. Calzolari et al., editors, Proceedings of the Ninth

International Conference on Language Resources and Evaluation (LREC'14), pages 3013-3019, Reykjavik, Iceland, May. European Language Resources Association (ELRA).

Dan, S., Kordjamshidi, P., Bonn, J., Bhatia, A., Cai, J., Palmer, M., and Roth, D. (2019). From Spatial Relations to Spatial Configurations. Manuscript submitted for publication.

Donatelli, L., Schneider, N., Croft, W., and Regan, M. (2019). Tense and aspect semantics for sentential AMR. *Proceedings of the Society for Computation in Linguistics*, 2(1):346—348.

Gotou, D., Nishikawa, H., and Tokunaga, T. (2016). An extension of ISO-space for annotating object direction. In K. Hasida et al., editors, Proceedings of the 12th Workshop on Asian Language Resources (ALR12), pages 1—9, Osaka, Japan, December. The 26th International Conference on Computational Linguistics (COLING '16).

Johnson, M., Hofmann, K., Hutton, T., Bignell, D. (2016). The Malmo platform for artificial intelligence experimentation. In S. Kambhampati, editor, Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI-16), pages 4246–4247, New York, USA, July. AAAI Press.

Kordjamshidi, P., van Otterlo, M., and Moens, M.-F. (2010). Spatial role labeling: task definition and annotation scheme. In N. Calzolari et al., editors, Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC'10), pages 413—420, Valletta, Malta, May. European Language Resources Association (ELRA).

Lakoff, G. (1987). Women, fire and dangerous things: what categories reveal about the mind. Chicago, the University of Chicago Press.

Levelt, W. (1996). Perspective taking and ellipsis in spatial description. In P. Bloom et al., editors, *Language and Space*. Cambridge, MA: The MIT Press, pp. 77—107.

Levinson, S. (2003). Space in Language and Cognition: Explorations in Cognitive Diversity. Cambridge, UK: Cambridge University Press.

Li, P. and Gleitman, L. (2002). Turning the tables: language and spatial reasoning. *Cognition*, 83(3):265—94.

Liu, W., Li, S., and Renz, J. (2018). Combining RCC-8 with qualitative direction calculi: algorithms and complexity. In C. Boutilier, editor, Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI-09), pages 854—859, California, USA, July. AAAI Press.

Mani, I., Hitzeman, J., Richer, J., Harris, D., Quimby, R., Wellner, B. (2008). SpatialML: annotation scheme, corpora, and tools. In N. Calzolari et al., editors, Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08), pages 410-415, Marrakech, Morocco, May. European Language Resources Association (ELRA).

Myers, S. and Palmer, M. (2019). ClearTAC: verb tense, aspect and form classification using neural nets. In N. Xue et al., editors, Proceedings of the First International Workshop on Designing Meaning Representations (DMR-2019), pages 136-140, Florence, Italy, August. Association for Computational Linguistics (ACL).

Narayan-Chen, A., Jayannavar, P., and Hockenmaier, J. (2019). Collaborative Dialogue in Minecraft. In A. Korhonen et al., editors, Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL'19), pages 5405–5415, Florence, Italy, July. Association for Computational Linguistics (ACL).

Narayan-Chen, A., Graber, C., Das, M., Islam, M. R., Dan, S., Natarajan, S., Doppa, J. R., Hockenmaier, J., Palmer, M., and Roth, D. (2017). Towards Problem Solving Agents that Communicate and Learn. In M. Bansal et al., editors, Proceedings of the First Workshop on Language Grounding for Robotics, pages 95—103, Vancouver, Canada, August. Association for Computational Linguistics (ACL).

O'Gorman, T., Pradhan, S., Bonn, J., Conger, K., Gung, J., and Palmer, M. (2018a). The new PropBank: aligning PropBank with AMR through POS unification. In N. Calzolari et al., editors, Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC'18), pages 1457–1463, Miyazaki, Japan, May. European Language Resource Association (ELRA).

O'Gorman, T., Regan, M., Griffitt, K., Hermjakob, U., Knight, K., and Palmer, M. (2018b). AMR beyond the sentence: the multi-sentence AMR corpus. In E. Bender et al., editors, Proceedings of the 27th International Conference on Computational Linguistics, pages 3693–3702, New Mexico, USA, August. Association for Computational Linguistics (ACL).

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. (2017). ISO-Space: Annotating static and dynamic spatial information. In N. Ide & J. Pustejovsky (Eds.), *Handbook of Linguistic Annotation*. Dordrecht: Springer, pp. 989–1024.

Pustejovsky, J. and Krishnaswamy, N. (2016). VoxML: a visualization modeling language. In N. Calzolari et al., editors, Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), Potorož, Slovenia, May. European Language Resources Association (ELRA).

Pustejovsky, J., Xue, N., and Lai, K. (2019). Modeling Quantification and Scope in Abstract Meaning Representations. In N. Xue et al., editors, Proceedings of the First International Workshop on Designing Meaning Representations (DMR-2019), pages 28–33, Florence, Italy, August. Association for Computational Linguistics (ACL).

Randell, D., Cui, Z., and Cohn, A. (1992). A spatial logic based on regions and connection. In B. Nebel et al., editors, Proceedings of the Third International Conference on the Principles of Knowledge

Representation and Reasoning (KR'92), pages 165—176, Massachusetts, USA, October.

Suhr, A., Lewis, M., Yeh, J., and Artzi, Y. (2017). A Corpus of Natural Language for Visual Reasoning. In R. Barzilay and M. Kan, editors, Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, pages 217–223, Vancouver, Canada, July. Association for Computational Linguistics (ACL).

Talmy, L. (1996). Fictive motion in language and "ception". In P. Bloom, M. A. Peterson, L. Nadel, & M. Garrett (Eds.), *Language and space*. Cambridge, MA: The MIT Press, pp. 211—276.

Tenbrink, T. and Kuhn, W. (2011). A model of spatial reference frames in language. In M. Egenhofer, N. Giudice, R. Moratz, & M. Worboys (Eds.), *Spatial Information Theory: 10th International Conference, COSIT 2011, Belfast, ME, USA, September 12-16, 2011, Proceedings*. Berlin, Heidelberg: Springer-Verlag, pp. 371–390.

Tellex, S., Kollar, T., Dickerson, S., Walter, M., Banerjee, A., Teller, S., and Nicholas, R. (2011). Understanding Natural Language Commands for Robotic Navigation and Mobile Manipulation. In W. Burgard and D. Roth (Program Cochairs) et al., editors, Proceedings of the 25th AAA Conference on Artificial Intelligence, pages 1507-1514, San Francisco, CA, USA, August.

Van Gysel, J. E. L., Vigus, M., Kalm, P., Lee, S., Regan, M., and Croft, W. (2019). Cross-lingual semantic annotation: reconciling the language-specific and the universal. In N. Xue et al., editors, Proceedings of the First International Workshop on Designing Meaning Representations (DMR-2019), pages 1–14, Florence, Italy, August. Association for Computational Linguistics (ACL).

Vigus, M., Van Gysel, J. E. L., and Croft, W. (2019). A dependency structure annotation for modality. In N. Xue et al., editors, Proceedings of the First International Workshop on Designing Meaning Representations (DMR-2019), pages 182–198, Florence, Italy, August. Association for Computational Linguistics (ACL).

Zhang, S., Ma, X., Duh, K., and Van Durme, B. (2019). AMR parsing as sequence-to-graph transduction. In A. Korhonen et al., editors, Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 80–94, Florence, Italy, July. Association for Computational Linguistics (ACL).

Zlatev, J. (2010). Spatial semantics. In D. Geeraerts & H. Cuyckens (Eds.), *The Oxford Handbook of Cognitive Linguistics*. New York: Oxford University Press., pp. 318-350.