,

Qualitative Spatial Ontologies for Robot Dynamics

Jona Thai[1], Michael Grüninger[1]

[1]Department of Mechanical and Industrial Engineering, University of Toronto, Ontario, Canada M5S 3G8

Abstract

Following Smith and Gasser's work on embodied cognition, one can consider a robot as an intelligent agent that interacts with its external environment through sensorimotor activities, such as touching, lifting, standing, sitting, and walking. In this paper we explore the ontologies that are required to represent and reason about robot dynamics. We propose new ontologies for robotic components and poses, including a new nonclassical mereotopology for touch contact. The design of the ontologies is driven by semantic parsing of natural language instructions (e.g. "Lift the box that is beside the chair and place it on the table"), through which we identify the spatial and mereotopological relations among a robot's components and the external world, as well as the activities that the robot can perform to change these relationships.

Keywords

spatial ontologies, mereotopology, robotics, semantic parsing

1. Introduction

Many within the field of artificial intelligence(AI) research are familiar with the "Monkey & Banana Problem" [1] - a famous toy problem aimed at solving and optimizing the best sequence of actions for a monkey to obtain bananas suspended from a ceiling, given a chair and a stick. Although the optimal solution is straightforward to find, what is the background knowledge and context necessary to execute it?

It is typically assumed that the monkey is aware of how to navigate and use tools such as sticks. To instill the same level of spatial awareness, it is common to turn to the quantitative precision of Euclidean geometry. However, a perfect understanding of Euclidean geometry does not necessarily translate to a perfect understanding of verbal instruction - an extension of the symbol-grounding problem [2] in embodied cognition [3]. This is somewhat consistent with the embodiment hypothesis, which descrives intelligence as an emergent phenomena of sensorimotor activity between an agent and its environment [4].Even simple natural-language instructions, such as "stand up" or "raise your arms" do not have a straightforward translation in logic. Perhaps one could argue that it is a problem easily solved through manual programming, but what comes of interpreting prepositions or state-of verbs? "Rinsing off a mug" or "Placing a block behind another" do not directly map to an equation or first-order preposition, and current semantic parsing infrastructure is ill-equipped to perform such operations with rigor, precision and reproducibility. [5][6].

RobOntics 2021: 2nd International Workshop on Ontologies for Autonomous Robotics, held at JOWO 2021: Episode VII The Bolzano Summer of Knowledge, September 11–18, 2021

This paper aims to highlight an ontological approach to supplement current work in robotic-natural-language-processing and spatial understanding. After presenting a set of motivating scenarios in robot dynamics, we introduce a set of new ontologies for qualitative spatial relations that are needed to formalize the logical sentences that correspond to natural language instructions within the scenarios.

2. Motivating Scenarios

We return to the "Monkey & Banana Problem" (or perhaps the "Robot & Banana Problem") in guiding our ontology design. Below are common poses we deem necessary to not only perform the activity of grabbing bananas, but are also indicative of basic self-awareness and qualitative spatial understanding.
Although simple to understand from a human perspective, these poses are not trivial. Mereotopologically speaking, all the poses in the motivating scenarios above are equivalent – in each case, the parthood and connection relations between objects are the same. However, their embedding within the physical space and intended function differ based on assumptions. For instance, Motivating Scenario 1 & 3 are unique in that they are independent of external environment. Regardless of the presence of a floor, gravity or ceiling, it should be achievable with spatial self-awareness alone.

1. Stand Up
    a) Feet are completely touching the ground.
    b) Other limbs and body parts are not touching any external object.

Figure 1: Stand Up

2. Arms Raised Over Head
    a) Feet are completely touching the ground
    b) Hands are above head.
    c) Shoulders and upper arms are next to head.
3. Right Hand Touching Top of Head
    a) Feet are completely touching the ground

Figure 2: Arms Raised Over Head

b) Only the right hand is touching the head.
c) No other body part is touching an external object.

Figure 3: Right Hand Touching Top of Head

4. Left Leg Lift
   a) Right foot is completely touching the ground.
   b) Left foot is not touching the ground, or any other external object
   c) No other body part, other than the right foot is touching an external object.

3. Ontologies for Spatial Relations in the Motivating Scenarios

We recognize two core components to qualitative spatial reasoning - self awareness and interacting with external objects. An ontology for self-awareness would provide the necessary axioms to define models as described in Motivating Scenarios 1-4. However, we recognize that Motivating Scenarios 1-4 may not be feasible on certain body anatomies. For example, BB8 from the Star Wars franchise would find it impossible to lift its left leg (Motivating Scenario 4), due to its lack of such a limb.

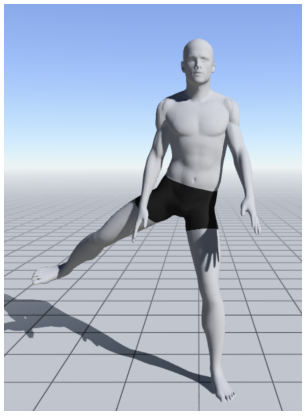

Figure 4: Left Leg Lift

Hence, for the sake of a holistic solution, we also define three large classes of anatomies as a reference. The ontologies for self-awareness and interacting with external objects are relative to these anatomies. Besides effectively constraining the set of feasible models, this approach also uniquely allows the robot to have two frames of reference – one of itself, and possibly one of a human, which could aid in the semantic parsing of human instruction.

3.1. Robot Anatomy

At its core, all anatomies and skeletons can be represented as a connected induced subgraph structure [7]. This is due to the inherent convexity of its mereotopology - for example, the sum of upper limb, joint, lower limb is different from upper limb, lower limb, joint. Convexity is the concept of importance of order in sums. Moreover, it allows us to represent hiearchical relations, which are particularly important in the context of Zoomorphic and Anthropomorphic Anatomies.

To represent differences in direction, such as "right" or "left", which tend to be innate and not relative to an external compass, we utilize the concept of "half-rays" from Hilbert's Foundations of Geometry [8]. We also utilize the *between* relation, where we interpret *between(a,b,c)* as a is between b and c. We translate Hilbert's axioms for half rays below:

$$\begin{aligned} \forall A \forall A' \forall O \forall B \forall a & body\_part(A) \wedge body\_part(A') \wedge body\_part(O) \wedge body(a) \wedge part(A,a) \\ & \wedge part(A'a) \wedge part(O,a) \wedge between(O,A,B) \wedge \neg between(O,A,A') \\ & \supset same\_side(A,A') \wedge \neg same\_side(A,B) \wedge \neg same\_side(A',B) \end{aligned} \tag{1}$$

3.1.1. General Anatomy

The basic building blocks for a skeleton are fixed, immovable parts(bones) and movable parts(joints). Joints dictate possible movements of bones – in other words, if possible movements are models of axioms, joints constrain the number of feasible models. Hence, the weakest anatomy structure contains only three classes - *joint(x)* , *bone(y)* and *limb(z)*(a limb can be any combination of bone and joint).

3.1.2. Zoomorphic Anatomy

The second class of anatomy we define is zoomorphic/quadri-pedal. It builds upon the previously defined General Anatomy, with the added restriction of the existence of a maximal element within the hierarchy of joints, limbs and bones. This maximal element is the *head(h)*.

3.1.3. Anthropomorphic Anatomy

Following the trend above, the Anthropomorphic Anatomy builds upon the Zoomorphic and General Anatomy. The Anthropomorphic Anatomy not only has a single unique maximal element-*head(h)*, but also has two minimal elements - *foot(f)*.

3.2. Ontology for Self-Awareness

Representations of space, and their use in qualitative spatial reasoning, are widely recognized as key aspects in commonsense reasoning, with applications ranging from biology to geography. The predominant approach to spatial representation within the applied ontology community has used mereotopologies, which combine topological (expressing connectedness) with mereological (expressing parthood) relations. A variety of first-order mereotopological ontologies have been proposed, the most widespread being the Region Connection Calculus (RCC) [9], the ontology RT [10], and the ontologies introduced by Casati and Varzi [11].

All of these approaches implicitly propose that there is a single connection relation. The weakest classical mereotopology, T_{mt}, has a signature that consists of two primitive binary relations, parthood (P) and connection (C). The axioms of the theory (Axioms 2 to 7) state that connection is a reflexive and symmetric relation, while parthood is a reflexive, transitive, and anti-symmetric relation. In addition, if one individual is connected to another, then the first one is also connected to any individual which the second is part of.

$$C(x,x). \tag{2}$$
$$C(x,y) \supset C(y,x). \tag{3}$$
$$P(x,x). \tag{4}$$
$$P(x,y) \wedge P(y,x) \supset (x = y). \tag{5}$$
$$P(x,y) \wedge P(y,z) \supset P(x,z). \tag{6}$$
$$P(y,z) \wedge C(x,y) \supset C(x,z). \tag{7}$$

However, the problems encountered in the motivating scenarios presented earlier in this paper lead to the proposal of topological pluralism – there are multiple distinct connection relations with different axiomatizations. Consider the relation $touches(x,y)$ that formalizes the relation that appears in the following natural language sentences:
My feet are touching the floor.
My hands are touching each other.
Touch your head.

Touch is a connection relation distinct from the connection relation between components and it also has an axiomatization different from the connection relation in classical mereotopologies. In

particular, *touches* is not reflexive – a hand cannot touch itself, and a body only touches itself if there are two disjoint proper parts of the body that touch each other. Similarly, a hand (as a proper part of the body) only touches the entire body if there is another disjoint part of the body that touches the hand e.g. two hands can touch each other and in this sense each hand is touching the body.

$$(\forall x,y)\, touches(x,y) \supset physical\_body(x) \wedge physical\_body(y) \tag{8}$$

$$(\forall x,y)\, touches(x,y) \supset touches(y,x) \tag{9}$$

$$(\forall x,y,z)\, touches(x,y) \wedge component\_of(y,z) \supset touches(x,z) \tag{10}$$

touches is a subproperty of physical connection:

$$(\forall x,y,z)\, touches(x,y) wedgeppart(y,z) \supset touches(x,z) \tag{11}$$

$$(\forall x)\, touches(x,x) \supset (\exists y,z)\, ppart(y,x) \wedge ppart(z,x) \wedge \neg overlaps(y,z) \wedge touches(y,z) \tag{12}$$

$$(\forall x,y)\, touches(x,y) \wedge ppart(x,y) \supset (\exists z)\, ppart(z,y) \wedge \neg overlaps(x,z) \wedge touches(x,z) \tag{13}$$

3.2.1. Revisiting the Motivating Scenarios

With an established ontology signature and anatomy, we can now revisit our motivating scenarios and redefine each instruction in first order logic. Each bipedal robot pose can be defined by a set of conditions on the spatial relations between the robot's body parts, as we can see in the next three axiomatization of poses. Note that these are not instruction axioms, rather they are a way of finding out what spatial relations are needed to understand each pose e.g. arms raised over head,

1. Arms Raised Over Head
   a) Feet are completely touching the ground

$$\forall x \forall y foot(x) \wedge foot(y) \wedge x \neq y \supset \exists z\, floor(z) \wedge touches(x,z) \wedge touches(y,z) \tag{14}$$

   b) Hands are above head.

$$\forall x \forall y \forall h\, hand(x) \wedge hand(y) \wedge head(h) \wedge x \neq y \supset above(x,h) \wedge above(y,h) \tag{15}$$

   c) Shoulders and upper arms are next to head.

$$\forall x \forall y \forall h\, shoulders(x) \wedge upper\_arms(y) \wedge head(h) \wedge above(y,x) \supset adjacent\_to(x,h) \wedge adjacent\_to(y,h) \tag{16}$$

2. Right Hand Touching Top of Head
   a) Feet are completely touching the ground

$$\forall x \forall y \, foot(x) \wedge foot(y) \wedge floor(z) \wedge x \neq y \supset \exists z \, floor(z) \wedge touches(x,z) \wedge touches(y,z) \tag{17}$$

   b) Only the right hand is touching the head.

$$\forall l \forall o \, body\_part(l) \wedge body\_part(o) \wedge touches(l,o) \supset right\_hand(l) \wedge head(o) \tag{18}$$

   c) No other body part is touching an external object.

$$\forall l \forall o \, body\_part(l) \wedge object(o) \wedge touches(l,o) \supset (foot(l) \wedge floor(o)) \vee (right\_hand(l) \wedge head(o)) \tag{19}$$

3. Left Leg Lift
   a) Right foot is completely touching the floor.

$$\forall l \forall o \, body\_part(l) \wedge object(o) \wedge touches(l,o) \supset right\_foot(l) \wedge floor(o) \tag{20}$$

   b) Left foot is not touching the ground, or any other external object
   c) No other body part, other than the right foot is touching an external object.

$$\forall l \forall o \, body\_part(l) \wedge object(o) \wedge touches(l,o) \supset foot(l) \wedge right(l) \wedge floor(o) \tag{21}$$

   Both 4(b) and 4(c) are represented by axiom 23.

3.3. Ontology for Interacting with External Objects

In a typical role-playing game tutorial (RPG), you first figure out how to move your avatar before interacting with the world within the game. Similarly, the Ontology for Interacting with External Objects builds upon the Ontology for Self-Awareness.
Specifically, the Ontology for Interacting with External Objects differentiates the *touch* relation used for touching/grasping from the connection relation used to describe the mereotopology between a joint and a bone. This is also a reason why Euclidean geometry is simultaneously too strong, yet insufficient to represent physical relations - it does not account for cases of topological pluralism (as demonstrated).

If a robot is standing, then one of its feet touches the floor:

$$(\forall x) \, standing(x) \supset (\exists y) \, component\_of(y,x) \wedge foot(y) \wedge touches(y, Floor) \tag{22}$$

In fact, both of its feet are probably touching the ground, if it is bipedal:

$$(\forall x) \, standing(x) \supset \forall x \forall y \exists z \, foot(x) \, foot(y) \, floor(z) \wedge x \neq y \supset touches(x,z) \wedge touches(y,z) \tag{23}$$

Since it is only described as standing up, it is assumed that the robot is not touching any other external object:

$$(\forall x)\, standing(x) \supset \forall l \forall o\, body\_part(l)\, object(o) \wedge touches(l,o) \supset foot(l) \wedge floor(o) \quad (24)$$

If a robot is sitting, then its torso is touching a chair and its feet are touching the floor:

$$\begin{aligned}(\forall x)\, sitting(x) \supset (\exists y,z,u)\, component\_of(y,x) \wedge torso(y) \\ \wedge chair(u) \wedge touches(y,u) \wedge component\_of(z,x) \wedge foot(z) \wedge touches(z,Floor)\end{aligned} \quad (25)$$

In the walk activity, one foot is always touching the ground:

$$\begin{aligned}(\forall o,x)\, occurrence\_of(o, walk(x)) \supset ((\forall s)\, subactivity\_occurrence(s,o) \\ \supset (\exists y)\, componentOf(y,x) \wedge foot(y) \wedge prior(tc(y,Floor),s)\end{aligned} \quad (26)$$

4. Methodology

Embodied question answering [12], [13] provides an interesting platform for identifying new ontologies for qualitative spatial relations that formalize the semantic properties of the robot and its environment [14]. In order to answer such questions, the agent must first intelligently navigate to explore the environment, gather necessary visual information, and then answer the question. The long-term goal is to build intelligent agents that can perceive their environment (through vision and other sensors), communicate (i.e., hold a natural language dialog grounded in the environment), and act. In this approach, a key capability is for intelligent robots to understand natural language instructions ([15], [16], [17], [18], [19]) This includes the problem of parsing natural language commands to actions and control structures that can be readily implemented in a robot execution system. [20] and the ability for robots to interact with human partners in following spoken instructions [21].

4.1. From Instructions to Process Descriptions

Semantic parsing maps natural language sentences to first-order logic formulae ([22], [23]). The ontology is correct and complete with respect to the natural language corpus iff any conclusion that the answer to a question about the natural language sentences is mapped to a logical formula that is entailed by the ontology.

We are therefore interested in using the ontology to represent the intended semantics of the terms that appear in the corpus.

Instructions are mapped to process descriptions with the PSL Ontology, which are logical formulae representing activities and the constraints on their occurrences. A process description for an atomic activity contains constraints that arise from the following two questions:

- Under what conditions does an atomic activity occur?
- How do occurrences of atomic activities change fluents?

Classes of complex activities are defined with respect to the following two questions:

- What is the relationship between the occurrence of the complex activity and occurrences of its subactivities?
- Under what conditions does a complex activity occur?
  An activity may have subactivities that do not occur; the only constraint is that any subactivity occurrence must correspond to a subtree of the activity tree that characterizes the occurrence of the activity.

Within the PSL Ontology, the notion of state is represented by reified fluents. Intuitively, a change in state is captured by fluents that are either achieved or falsified by an activity occurrence. The prior relation is used to specify the fluents that are intuitively true prior to an activity occurrence and the holds relation specifies the fluents that are intuitively true after an activity occurrence. Furthermore, a fluent can only be changed by the occurrence of activities. Thus, if some fluent holds after an activity occurrence, but after an activity occurrence later along the branch it is false, then an activity must occur at some point between that changes the fluent. This also leads to the requirement that the fluents holding after an activity occurrence will be the same fluents that are prior to any successor occurrence, since there cannot be an activity occurring between them.

The design of the ontologies is driven by semantic parsing in two ways. In the first approach, stative verbs and participles of dynamic verbs are mapped to relations between entities within the signature of the ontology. For example, *sitting*, *standing*, *touching* are all present participles of the dynamic verbs *sit*, *stand*, *touch*, respectively, Each of the first three are represented by a relationship between an entity and its environment, while each of the last three are represented by an activity that can possibly change the relationship. This leads to a series of domain ontologies for new mereotopologies and spatial relations. Each domain ontology is translated into a domain state ontology. Relations in the domain ontology are mapped to fluents in the domain state ontology. Given the domain state ontology, we axiomatize the domain process ontology. Classes of activities in the domain process ontology are associated with dynamic verbs within the phrase map for the semantic parser.

In the converse direction, we start with the dynamic verbs, which are mapped to activities within the signature of the ontology. We identify the fluents that are changed when the process corresponding to the verb occurs. Axiomatize the domain state ontologies that contain the fluents in their signature Identify the domain ontologies for these domain state ontologies. Axiomatize the domain process ontologies. The process corresponding to the verb will either be an atomic activity in one of the domain process ontologies or it is a complex activity that is composed of activities in the domain process ontologies. The fluents which are changed by the activities within the domain process ontology are associated with stative verbs and participles of dynamic verbs within the phrase map for the semantic parser.

One technique for ontology validation is to demonstrate that these two directions indeed converge on the same domain state and process ontologies. For example, the activities that change the $standing(x)$ fluent include $stand(x)$, $sit(x)$, $walk(x)$.

The domain process methodology guarantees that we have a complete classification of activities that change a given set of fluents. In turn, the set of fluents is a complete characterization of the possible states of the world given the set of domain ontologies.

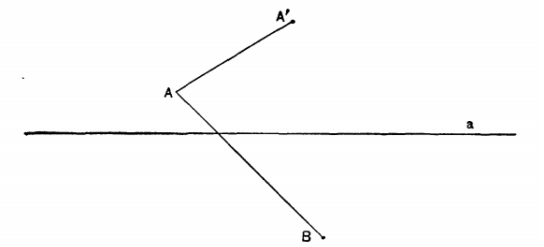

Fig. 5.

Figure 5: Two-dimensional cross-section

Fig. 6.

Figure 6: One-dimensional cross-section

4.2. Hilbert's Foundations of Geometry

In Section 1, we argued the case for qualitative spatial reasoning; the primary reason is the lack of a clear mapping between equation and natural language statement. The secondary reason is that Euclidean geometry (the underlying framework of quantitative reasoning) is both too weak and too strong to account for the nuances of spatial relationships between active agents and external objects. The development of a qualitative spatial reasoning framework is to combat both of the above issues.

A tertiary problem we seek to solve is that of undecideability. Other mereotopologies for qualitative spatial reasoning, such as the region-connection-calculus (RCC8) and classical extensional mereology are not decideable. Hence, concepts of parthood and connection are also undecideable. However, we have found a potential workaround through Hilbert's geometry (which is decideable, due to the fact that it is interpretable by real arithmetic, which is decideable).

Our ontology is faithfully interpretable in Hilbert's geometry. This was a solution devised to combat the problem of not being able to discern poses mereotopologically. By mapping each skeleton(anatomy) to a polygon under Hilbert's geometry, we can utilize Hilbert's theories of connection and order to discern direction and contact without relying on an external surface or compass. Hilbert's geometry is traditionally interpreted as second-order due to the definition of a line of a set of segments – where a "set" is mathematically a second order concept. However, we propose an alternative interpretation of Hilbert's Geometry – as a mereology of lines. Not only is it applicable to describe different parthood relations, it is allows us to represent planes (2D) and polygons(3D) as one dimensional lines. For example, Figure 5 can be represented by Figure 6, as

taken from Hilbert's Foundations in Geometry [8]. In Figure 5, the line A-A' is on a different side as the line A-B relative to the plane a. All of this information is consolidated within Figure 6, and describable with the *between* relation.

5. Conclusion

We began this paper with the goal of building the case for embodied AI and design the ontological foundations necessary for robot understanding of natural language instruction. This is to build a logically consistent, modular framework for performing qualitative physical reasoning. However, there were two immediate challenges to this endeavor. There was the problem of differentiating different directions and poses, independent of external environment, despite the fact that they are mereotopologically equivalent. There was also the issue of differentiating the weakness and strength of different connection relations (e.g. touching or being fused together). Hence, we developed a general methodology based on Hilbert's Geometry and the Process Specification Language(PSL). We then applied this methodology to design an ontology for Robot Anatomy, as well as an Ontology for Self-Awareness and an Ontology for Touch/Contact relations, which are relative to the robot's type and specification of anatomy. We then used these ontologies to define axioms to describe poses related to sub-problems of the "Monkey and Banana Problem".

From a research program perspective, we will follow up this paper by further applying our established methodology to discover new domain and process ontologies related to additional natural language instructions in a robot context. From an ontology perspective, we will continue fine-tuning other aspects of spatial awareness. For example, defining concepts of floor and ceiling, and how they relate to each other e.g. the relationship between the first and second floor, or between the ground floor ceiling and the top of the building.

References

[1] Wikipedia, The monkey and banana problem, Web Article Entry, 2021. https://en.wikipedia.org/wiki/Monkey_and_banana_problem.

[2] S. Harnad, The symbol grounding problem, CoRR cs.AI/9906002 (1999). URL: https://arxiv.org/abs/cs/9906002.

[3] L. Smith, M. Gasser, The development of embodied cognition: six lessons from babies, Artificial Life 11 (2005) 1–2.

[4] A. Das, S. Datta, G. Gkioxari, S. Lee, D. Parikh, D. Batra, Embodied question answering, 2017. arXiv:1711.11543.

[5] M. Shridhar, J. Thomason, D. Gordon, Y. Bisk, W. Han, R. Mottaghi, L. Zettlemoyer, D. Fox, ALFRED: A benchmark for interpreting grounded instructions for everyday tasks, CoRR abs/1912.01734 (2019). URL: http://arxiv.org/abs/1912.01734. arXiv:1912.01734.

[6] J. Thai, M. Gruninger, Robot meets world, in: RobOntics2020, 2020.

[7] M. Gruninger, C. Chui, Y. Ru, J. Thai, A mereology of connected structures, in: FOIS, 2020.

[8] D. Hilbert, The Foundations of Geometry - Translation by E.J. Townsend, volume 1 of

1, 2 ed., The Open Court Publishing Company, University of Illinois, 1950. Authorized Translation by E.J. Townsend.

[9] A. G. Cohn, B. Bennett, J. Gooday, N. Gotts, RCC: a calculus for region-based qualitative spatial reasoning, GeoInformatica 1 (1997) 275–316.

[10] N. Asher, L. Vieu, Toward a geometry of common sense: A semantics and a complete axiomatization of mereotopology, in: IJCAI, 1995, pp. 846–852.

[11] R. Casati, A. Varzi, Parts and places: The structures of spatial representation, MIT Press, 1999.

[12] A. Das, S. Datta, S. Gkioxari, Lee, D. Parikh, D. Bhatra, Embodied question answering, in: CVPR 2018, 2018, pp. 1–9.

[13] M. Savva, A. Kadian, O. Maksymets, Y. Zhao, E. Wijmans, B. Jain, J. Straub, J. Liu, V. Koltlun, J. Malik, D. Prikh, D. Batra, Habitat: A platform for embodied ai research, in: ICCV 2019, 2019, pp. 9339–9347.

[14] M. R. Walter, S. Hemachandra, B. Homberg, S. Tellex, S. Teller., Learning semantic maps from natural language descriptions, in: Proceedings of the 2013 Robotics: Science and Systems IX Conference, June 24-28, 2013, Berlin, Germany., 2013, pp. 2–9.

[15] M. Shridar, J. Thmason, D. Gordon, Y. Bisk, H. W., R. Mottaghi, L. Zettlemoyer, D. Fix, Alfred: A benchmark for interpreting grounded instructions for everyday tasks, in: CVPR 2020, 2020, pp. 10740–10749.

[16] P. Anderson, Q. Wu, D. Teney, J. Bruce, M. Johnson, N. Suenderhauf, I. Reid, G. S., A. van den Hengel, Vision-and-language navigation: Interpreting visually-grounded navigation instructions in real environments, in: CVPR 2018, 2018, pp. 3674–3683.

[17] Y. Bisk, A. Holtzman, J. Thomason, J. Andreas, Y. Bengio, J. Chai, M. Lapata, A. Lazaridou, J. May, A. Nisnevich, N. Pinto, J. Turian, Experience grounds language, 2020. arXiv:2004.10151.

[18] Y. Wu, Y. Wu, G. Gkioxari, Y. Tian, Building generalizable agents with a realistic and rich 3d environment, 2018. arXiv:1801.02209.

[19] J. Thomason, S. Zhang, R. Mooney, P. Stone, Learning to interpret natural language commands through human-robot dialog, in: IJCAI 2015, 2015, pp. 1923–1929.

[20] C. Matuszek, E. Herbst, L. Zettlemoyer, D. Fox, Learning to Parse Natural Language Commands to a Robot Control System, Springer International Publishing, Heidelberg, 2013, pp. 403–415.

[21] S. Tellex, T. Kollar, S. Dikerson, M. Walter, A. Bnnerjee, S. Teller, N. Roy, Understanding natural language commands for robotic navigation and mobile manipulation, in: AAAI 2011, 2011, pp. 1507–1514.

[22] J. Berant, A. Chou, R. Frostig, P. Liang, Semantic parsing on Freebase from question-answer pairs, in: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Seattle, Washington, USA, 2013, pp. 1533–1544. URL: https://aclanthology.org/D13-1160.

[23] Y. Wang, J. Berant, P. Liang, Building a semantic parser overnight, in: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Association for Computational Linguistics, Beijing, China, 2015, pp. 1332–1342. URL: https://aclanthology.org/P15-1129. doi:10.3115/v1/P15-1129.