Augmenting Abstract Meaning Representation for Human-Robot Dialogue

Claire Bonial[1], Lucia Donatelli[2], Stephanie M. Lukin[1], Stephen Tratz[1],
Ron Artstein[3], David Traum[3], and Clare R. Voss[1]

[1]U.S. Army Research Laboratory, Adelphi, MD 20783
[2]Georgetown University, Washington DC 20057
[3]USC Institute for Creative Technologies, Playa Vista, CA 90094
claire.n.bonial.civ@mail.mil

Abstract

We detail refinements made to Abstract Meaning Representation (AMR) that make the representation more suitable for supporting a situated dialogue system, where a human remotely controls a robot for purposes of search and rescue and reconnaissance. We propose 36 augmented AMRs that capture speech acts, tense and aspect, and spatial information. This linguistic information is vital for representing important distinctions, for example whether the robot has moved, is moving, or will move. We evaluate two existing AMR parsers for their performance on dialogue data. We also outline a model for graph-to-graph conversion, in which output from AMR parsers is converted into our refined AMRs. The design scheme presented here, though task-specific, is extendable for broad coverage of speech acts using AMR in future task-independent work.

1 Introduction

We describe an augmented version of Abstract Meaning Representation (AMR) for use as a conduit for natural language understanding (NLU) in a robot dialogue system. We find that while AMR is promising for NLU, refinements are needed in order to capture information critical for live, situated communication. Specifically, we propose the addition of a set of speech acts, tense and aspect information, and parameters that help specify spatial location.

After providing background on our broader research goals and the AMR project, we motivate our choice to explore the use of AMR for NLU (sections 2, 3). We then detail our findings on gaps in the representational coverage of existing AMR for human-robot dialogue (4), and we describe our refinements (5). We next describe ongoing and future work to implement an augmented AMR-based NLU that uses existing parsers and graph-to-graph AMR conversion to replace a more limited statistical classifier (6). We then compare to related work (7) and conclude.

2 Background: Human-Robot Dialogue

The broad goal of this research is to develop a system for conducting dialogue between a person and a remotely located robot in collaborative navigation tasks common to disaster relief and search-and-rescue scenarios. Efficient communication is essential: the robot must be able to interpret both the language used by the human and the intention behind it, as well as to carry out the instructions in these dynamic environments and coordinate with the human by providing appropriate feedback of the status of instructions at different times.

In the language of this domain, we find that people communicating with robots often employ multiple ways of saying the same thing: *Turn/rotate left*, *Drive/move/go forward*. However, they also employ very similar syntactic structures to say different things: *Can you take a picture?*, intended as a polite request for a picture, and *Can you speak Arabic?*, intended as a question of the robot's abilities. To get at the underlying meaning of these utterances despite surface variations and similarities, our goal is to develop semantic representations for this project. We plan to use these representations in an implemented, live system to facilitate both NLU of the robot-directed instructions as well as Natural Language Generation (NLG) of robot responses and feedback.

2.1 Human-Robot Dialogue Corpus

We collected a corpus of observed data from the target domain collected via a phased Wizard-of-Oz approach (Marge et al., 2016, 2017), in which a participant directed what they believed to be an autonomous robot to complete search and navigation tasks. In reality, the participant was speaking with two "wizard" experimenters responsible for the robot's dialogue and navigation capabili-

Proceedings of the First International Workshop on Designing Meaning Representations, pages 199–210
Florence, Italy, August 1st, 2019

	Left floor		Right Floor		Annotations		
#	**Participant**	**DM → Participant**	**DM → RN**	**RN**	**TU**	**Ant**	**Rel**
1	move forward 3 feet				1		
2		ok			1	1	ack-wilco
3			move forward 3 feet		1	1	trans-r
4				done	1	3	ack-done
5		I moved forward 3 feet			1	4	trans-l

Table 1: Example of a Transaction Unit (TU) which contains an instruction initiated by the participant, its translation to a simplified form (DM to RN), and the execution of the instruction and acknowledgement of such by the RN. TU, Ant(ecedent), and Rel(ation type) are indicated in the right columns.

ties. This setup allowed for the creation of a corpus of human-robot interactions that shows how people communicate with a robot in collaborative tasks when they are unconstrained in their communication.

Dialogues in the corpus follow a set procedure: a dialogue manager wizard (DM) listens to the participant's spoken instructions and replies to the participant with feedback and clarification requests via text messages. Executable instructions are passed along by the DM to a robot navigator wizard (RN) via text messages in a separate chat stream unseen by the participant. The RN then tele-operates the robot to complete the participant's instructions. Finally, the RN provides spoken feedback to the DM of completed actions or problems that arose, which are relayed by the DM to the participant. A sample interaction can be seen in Table 1.

The corpus contains dialogues from a total of 82 participants across three separate phased data collections. The participants' speech and the RN's speech are transcribed and time-aligned with text messages generated by the DM and sent either to the participant or the RN.

2.2 Dialogue Structure Annotations

The corpus also includes annotations of several aspects of dialogue structure (Traum et al., 2018) that allow for the characterization of distinct information states (Traum and Larsson, 2003). The portion of the data that we used, constituting about 20 hours of interaction, has been annotated with this scheme, specific to multi-floor dialogue that identifies high-level aspects of initiator intent and signals relations between individual utterances pertaining to that intent.

An example annotation can be seen in Table 1. The scheme consists first of *transaction units (TU)*, which cluster utterances from multiple participants and floors into units according to the joint realization of an initiator's intent. *Relations* indicate the graph structure of utterances within the same TU, and are indicated with a *Relation type (Rel)* (e.g., "ack-done" in row 4 of Table 1, signals that an utterance acknowledges completion of a previous utterance) and an *Antecedent (Ant)* for the relation. The existing annotation scheme highlights *dialogue structure*, but does not provide a markup of the semantic content of participant instructions, which is the goal of our work.

3 Background: AMR

The AMR project (Banarescu et al., 2013) has created a manually annotated semantics bank of text drawn from a variety of genres. Each sentence is represented by a rooted directed acyclic graph in which variables (or graph nodes) are introduced for entities, events, properties, and states; leaves are labeled with concepts (e.g., `(d / dog)`). For ease of creation and manipulation, annotators work with the PENMAN representation of the same information (Penman Natural Language Group, 1989), as in Figure 1.

```
(w / want-01
   :ARG0 (d / dog)
   :ARG1 (p / pet-01
             :ARG0 (g / girl)
             :ARG1 d))
```

Figure 1: AMR of *The dog wants the girl to pet him.*

A goal of AMR research is to capture core facets of meaning while abstracting away from idiosyncratic syntactic structures; thus, the same underlying concept realized alternatively as a noun (*a left turn*), verb (*turn to the left*) or light verb construction (*make/do a left turn*) will all be represented by identical AMRs.

3.1 Motivation for AMR in Human-Robot Dialogue

A primary motivation for using AMR is that there are a variety of fairly robust AMR parsers we can employ for this work, enabling us to forego manual annotation of data and facilitating efficient automatic parsing in a future end-to-end system.

The structured graph representations of AMRs additionally facilitate the interpretation of novel instructions and grounding instructions with respect to the robot's current physical surroundings. This structure allows us to pinpoint those actions that are executable for the robot. This latter motivation is especially important given that the target human-robot dialogue is physically situated and therefore distinct from other dialogue systems, such as chat bots, which do not require establishing and acting upon a shared understanding of the physical environment and often do not require any intermediate semantic representation (see Section 7 for related work). AMR thus offers both efficient and accurate parsing of natural language to a structured representation, as well as ease of conversion of this broad coverage representation to the domain-specific representation discussed in this paper (see 6.2 for more on graph conversion).

The fact that AMRs abstract away from surface variation is a complementary motivation for exploring their use within an NLU component. The AMRs "tame" some of the variation of natural language, representing core concepts in the human's commands, which must ultimately be mapped into the robot's low-level mechanical operations. Therefore, the robot will only be trained to process and execute the actions corresponding to semantic elements of the representation (see Section 6).

This processing and execution can be seen with a concrete example. Throughout the corpus data, participants use the commands *Take a picture* and *Send image* (as well as other variants) with the same intention that the robot take a picture of what is in front of it and send that image to the participant's screen. While *take* is a light verb in this usage (and therefore dropped from the representation according to existing AMR guidelines), *send* maintains its semantic weight and argument structure. For the purposes of our task, we can abstract away from this variation and convert both types of utterances into `send-image` commands (see 5.2). Though future work may deem these distinctions of lexical choice and syntax meaningful, the current task generalizes them for ease of task completion.

4 Evaluating Suitability of AMR

We began our assessment of AMR for human-robot dialogue by producing a small, randomly selected sample (137 sentences) of gold standard, manual annotations (provided by one senior and two recently trained AMR annotators), based on existing guidelines.[1] We then examined how effectively these gold, guideline-based AMRs can capture the distinctions of interest for human-robot dialogue and how accurately two available AMR parsers generate those gold annotations.

Common instructions in the corpus include *Move forward 10 feet*, *Take a picture*, and *Turn right 45 degrees*. People also used landmark-based instructions such as *Move to face the yellow cone*, and *Go to the doorway to your right*, although these were less common than the metric-based instructions (Marge et al., 2017). In response to these instructions from the DM to the participant, common feedback would be indications that an instruction will be carried out (*I will move forward 10 feet*), is in progress (*Moving...*), or completed (*I moved forward 10 feet*). Given that current AMR guidelines do not make tense/aspect distinctions, these three types of feedback from the robot are represented identically under the current guidelines (see Figure 2). The distinctions between a promise to carry out an instruction in the future, a declarative statement that the instruction is being carried out, and an acknowledgment that it has been carried out are critical for conveying the robot's current status in a live system.

```
(m / move-01
      :ARG0 (i / i)
      :direction (f / forward)
      :extent (d / distance-quantity
            :quant 10
            :unit (f2 / foot)))
```

Figure 2: Identical AMR for *I will move / I am moving / I moved forward...10 feet.*

Although the imperative *Move forward 10 feet* should receive an AMR marker `:mode imperative`, our evaluation of the existing

[1] https://github.com/amrisi/amr-guidelines/blob/master/amr.md

parsers JAMR (Flanigan et al., 2014) and CAMR (Wang et al., 2015) showed that parser output does not include this marker as it is rare if not entirely missing from the AMR 1.0 or 2.0 training corpora (Section 6).[2] As a result, the command to move forward also received the identical above AMR (Figure 2) in parser output. While this suggests that additional training data is needed that includes imperatives, this speaks to a larger issue of AMR: the existing representation is very limited with respect to speech act information. Current AMR includes `:mode imperative` and represents questions through the presence of `amr-unknown` standing in for the concept or polarity being questioned. All unmarked cases are assumed to be assertions. We found that more fine-grained speech act information is needed for human-robot dialogue.

5 Refinements to AMR

To design a representative set of augmented AMRs that capture the breadth of information necessary for collaborative dialogue in our domain, we started by creating a histogram of existing dialogue annotation categories for the 20 hours of experimental data available (described in Section 2.2). This allowed us to see which types of dialogue utterances are most prevalent in the corpus, as well as to view the range of utterances that comprise each category. Based on this data, we designed a set of AMR "templates"—skeletal AMRs in which the top, anchor node is a fixed relation corresponding to a speech act type (e.g., `assert-02`), one of its arguments is a fixed relation corresponding to an action (e.g., `turn-01`), and arguments of these relations are filled out given the specifics of a particular utterance. These skeletal AMRs can be modified and leveraged for NLU and generation in future human-robot collaboration tasks. We note that our objective is to produce a set of refined AMRs that provide coverage for human-robot dialogue, rather than an attempt to change AMR on a general scale.

We augmented AMR with the following information: i) coarse-grained information related to the *when* (tense) and *how* (aspect) of events (5.1); ii) speech acts (5.2); and iii) basic spatial information pertinent to robot functioning (5.3).

[2] https://catalog.ldc.upenn.edu/LDC2014T12, https://catalog.ldc.upenn.edu/LDC2017T10

5.1 Tense & Aspect

AMR currently lacks information that specifies *when* an action occurs relative to speech time and whether or not this action is completed (if a past event) or able to be completed (if a future event). This information is essential for situated human-robot dialogue, where successful collaboration depends on bridging the gap between differing perceptions of the shared environment and creating common ground (Chai et al., 2014).

Our tense and aspect annotation scheme is based on Donatelli et al. (2018), who propose a four-way division of temporal annotation and three multi-valued categories for aspectual annotation that fits seamlessly into existing AMR annotation practice. We reduced the authors' proposed temporal categories to three, to capture temporal relations before, during, and after the speech time. In addition to the aspectual categories proposed by Donatelli et al. (2018), we added the category `:completable +/-` to signal whether or not a hypothetical event has an end-goal that is executable for the robot (described further in Section 5.3). Our annotation categories for tense and aspect can be seen in Table 2.

TEMPORAL ANNOTATION	ASPECTUAL ANNOTATION
`:time`	
1. `(b / before :op1 (n / now))`	`:stable +/-`
2. `(n / now)`	`:ongoing +/-`
3. `(a / after :op1 (n / now))`	`:complete +/-`
	`:habitual +/-`
	`:completable +/-`

Table 2: Three categories for temporal annotation and five categories for aspectual annotation are used to augment existing AMR for collaborative dialogue.

Notably, this annotation scheme is able to capture the distinctions missing in Figure 2. Updated AMRs for utterances that communicate information about a "move" event relative to the future, present, and past are now re-annotated as in Figure 3. Using the scheme in Table 2, our augmented AMRs allow for locating an event in time and expressing information related to the boundedness of the event, i.e. whether or not the event is a future event with a clear beginning and endpoint, a present event in progress towards an end goal, or a past event that has been completed from start to finish.

```
1. (m / move-01 :completable +
      :ARG0 (i / i)
      :direction (f / forward)
      :extent (d / distance-quantity
            :quant 10
            :unit (f2 / foot))
      :time (a / after
            :op1 (n / now)))

2. (m / move-01 :ongoing + :complete -
      :ARG0 (i / i)
      :direction (f / forward)
      :extent (d / distance-quantity
            :quant 10
            :unit (f2 / foot))
      :time (n / now))

3. (m / move-01 :ongoing - :complete +
      :ARG0 (i / i)
      :direction (f / forward)
      :extent (d / distance-quantity
            :quant 10
            :unit (f2 / foot))
      :time (b / before
            :op1 (n / now)))
```

Figure 3: Updated AMRs for (1) *I will move...*, (2) *I am moving...*, and (3) *I moved...*. New temporal information is in blue; new aspectual information is purple.

5.2 Speech Acts

Annotation of speech acts allows us to capture how dialogue participants use language (its pragmatic effect) in addition to what the language means (its semantic content). The existing annotation on the corpus involves only dialogue structure (section 2.2). Our longer-term goal is to create a set of speech acts that i) cover the range of in-domain language use found in the corpus and ii) are generalizable to speech acts in other dialogue and conversational settings. To inform this work, we drew upon classical speech acts work such as Austin (1975) and Searle (1969).

To capture the range of speech acts present in the corpus, we arrived at an inventory of 36 unique speech acts specific to human-robot dialogue, inspired loosely by the dialogue move annotation of Marge et al. (2017). These 36 speech acts are classified into 5 types. In Figure 4, these are listed with the number of their subtypes in parentheses, along with a list of example subtypes for the type `command`. A full listing of subtypes and can be found in the Appendix.

To integrate speech acts into AMR design, we selected existing AMR/PropBank (Palmer et al., 2005) rolesets corresponding to each speech act (e.g., command-02, assert-02, request-01, etc.)

SPEECH ACT TYPES

`c / command` (6)	→	`command:move`
`a / assert` (9)		`command:turn`
`r / request` (4)		`command:send-image`
`q / question` (3)		`command:repeat`
`e / express` (5)		`command:cancel`
		`command:stop`

Figure 4: Five proposed speech act types for human-robot dialogue are listed on the left with number of subtypes in parentheses. Examples of the range of subtypes for `:command` are given to the right.

that serve as the anchor node in our augmented AMR. One argument of each of these top-level speech act relations corresponds to the action being commanded or asserted, or in general the content of a question, command, or assertion (e.g., turn-01, move-01, picture-01, etc.). For each speech act constituting the top relation and each action constituting one argument of the speech act relation—i.e. each speech act subtype in Figure 4—there is a corresponding AMR template. All utterances of a particular speech act and action combination are mapped to one template. For example, see (1) in Figure 6 for a blank `assert:turn` template and (2) and (3) for completed AMRs using that template. Note that semantically similar utterances using different vocabulary choices (e.g., *rotate, spin*), which would have slightly distinct AMRs under existing guidelines, would all map to the same AMR template using turn-01 (see Section 6.2 for plans on how to map parser output to templates).

5.3 Spatial Information

A key component of successful human-robot collaboration is whether or not robot-directed commands are executable. In the dialogues represented in the corpus, for a command to be effectively executable by the robot, it must have a clear beginning and endpoint and comprise a basic action. For example, *Move forward* is not executable, since it lacks a clear endpoint; *Move forward two feet*, which identifies an endpoint, is executable. Additionally, a command such as *Explore this room* is currently too high-level for our robot to execute. For implementation within a robot's system, a semantic representation must include well-defined, low-level actions that can then be combined into more complex actions.

Thus, our set of AMRs make explicit any implicit spatial roles in the PropBank/AMR verb role sets (in this sense, we follow the annotation prac-

tices of O'Gorman et al. 2018 for Multi-Sentence AMR). Our AMRs also specify additional spatial parameters necessary for a command to be executable, in the form of new core and non-core roles, when these are not already present in the original relation's set of arguments. If all required roles are present and instantiated by an utterance, then our AMR is marked with `completable +`; if any required roles are missing, the AMR is marked with `completable -`. For example, see Figure 5 for a non-executable command that requires more information to be carried out.

```
(c / command-02
    :ARG0 (c2 / commander)
    :ARG1 (r / robot)
    :ARG2 (m / move-01 :completable -
          :ARG0 r
          :direction (f / forward)
          :extent (a / amr-unknown)
          :time (a2 / after
                :op1 (n / now))))
```

Figure 5: *Move forward* (non-executable) is missing spatial information to complete the action. An existing AMR concept, `a / amr-unknown`, is employed to stand in for the missing parameter.

5.4 Final AMR Templates

Our final set of AMRs needed to provide coverage for the search and navigation domain includes 36 templates (one template corresponding to each speech act and action combination), which capture i) tense and aspect information; ii) speech acts; and iii) spatial parameters required for robot execution. In addition to a command example in Figure 5, we provide an example of a blank `assert:turn` template with filled-in examples of assertions about the future and present moments in Figure 6.

Note that we do not yet know how effective these templates will be in facilitating task-oriented human-robot dialogue. Future evaluation will include examining the coverage of these templates in mapping to a robot-specific action specification as well as generating appropriate responses and feedback. Our plans for implementation for further evaluation are presented in the next section.

6 Implementation

The intent behind our exploration of AMR for human-robot dialogue is to create a representation that is useful for an eventual live implemented system. To accomplish this goal we intend to i) leverage existing parsers to gain automatic AMR parses for the corpus data; ii) use graph-to-graph transformations to move from parser output to one of the 36 augmented in-domain AMRs; and iii) integrate the resulting AMRs with a language understanding component.[3] Our planned pipeline is presented in Figure 7. Ongoing work on each of these components is described in the sections to follow.

```
1. (a / assert-02
    :ARG0-speaker
    :ARG2-listener
    :ARG1 (t / turn-01
        :ARG1-thing turning
        :direction
        :extent
        :destination))

2. (a2 / assert-02
      :ARG0 (r2 / robot)
      :ARG1 (t / turn-01 :completable +
            :ARG1 r2
            :direction (r / right-04
                  :ARG2 r2)
            :extent (a / angle-quantity
                  :quant 90
                  :unit (d / degree))
            :time (a2 / after
            :op1 (n / now)))
      :ARG2 (c / commander))

3. (a2 / assert-02
      :ARG0 (r2 / robot)
      :ARG1 (t / turn-01 :ongoing +
            :complete -
            :ARG1 r2
            :direction (r / right-04
                  :ARG2 r2)
            :extent (a / angle-quantity
                  :quant 90
                  :unit (d / degree))
            :time (n / now))
      :ARG2 (c / commander))
```

Figure 6: Final AMR template of `assert:turn`. Blank template in (1), followed by a future *I will turn right 90 degrees* and a present, follow-up *turning*.

6.1 AMR Parsers

We initially developed a triple-annotated and adjudicated gold standard sample of 137 sentences from the given corpus to serve as a test set for evaluating the performance of the existing AMR parsers. Inter-annotator agreement (IAA) among the initial independent annotations obtained ade-

[3] Although we do plan to explore the utility of AMR for NLG, we focus first on the NLU direction of communication.

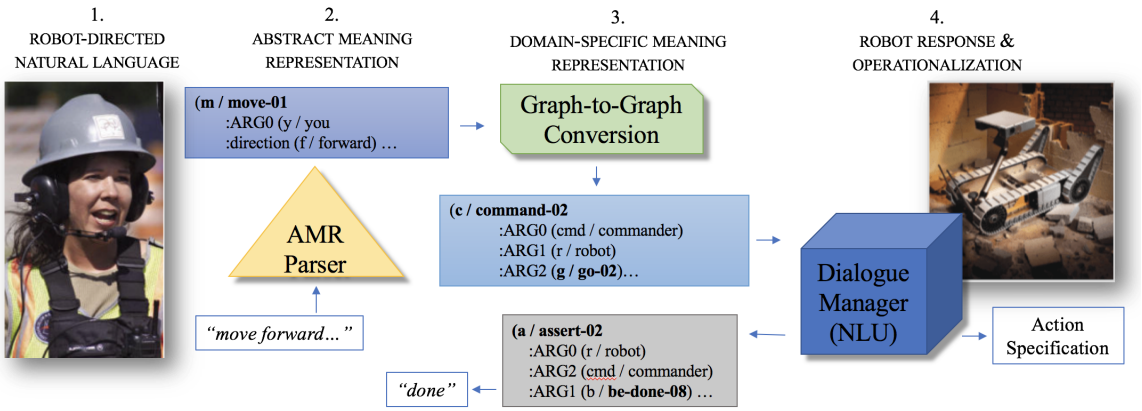

Figure 7: Planned pipeline for implementing AMRs into our human-robot dialogue system: natural language instructions are parsed using AMR parsers into existing AMR, which is then converted via graph-to-graph transformation into one of our augmented AMR templates. If all required parameters in the template are complete and the instruction executable, it will be mapped onto one of the robot's action specifications for execution. Clarifications and feedback from the robot are generated from the AMR templates.

quate scores of .82, .82, and .91 using the Smatch metric (Cai and Knight, 2013). According to AMR development group communication, 2014, IAA Smatch scores on AMRs are generally between .7 and .8, depending on the complexity of the data.

Having created a gold standard sample of our data, we ran both JAMR[4] (Flanigan et al., 2014) and CAMR[5] (Wang et al., 2015) on the same sample and obtained the Smatch scores when compared to the gold standard. We selected these two parsers to explore because JAMR was one of the first AMR parsers and uses a two-part algorithm to first identify concepts and then to build the maximum spanning connected subgraph of those concepts, adding in the relations. CAMR, in contrast, starts by obtaining the dependency tree—in this case, using the Charniak parser[6] and Stanford CoreNLP toolkit (Manning et al., 2014)—and then applies a series of transformations to the dependency tree, ultimately transforming it into an AMR graph. As seen in Table 3, CAMR performs better on both precision and recall when trained on AMR 1.0, thus obtaining the higher F-score. However, compared to their self-reported F-scores (0.58 for JAMR and 0.63 for CAMR) on other corpora, both under-perform on the human-robot dialogue data.

Given the relatively poor performance of both parsers on the human-robot dialogue data and error analysis of the output, we concluded that additional in-domain training data was needed. To this end, we manually selected 504 sentences (distinct from the original 137 test set) made up of short, sequential excerpts of the corpus data representative of the variety of common exchange types that we see. These sentences were independently double-annotated (IAA 87.8%) and adjudicated to create our new small training set. We retrained JAMR in several iterations. First, we retrained JAMR on the larger AMR 2.0 corpus (which includes and expands upon the AMR 1.0 corpus), then we retrained JAMR on the AMR 2.0 corpus and our in-domain data combined. Comparative results are summarized in Table 3. We are currently exploring retraining CAMR and plan to investigate other more recent parsers, such as Lyu & Titov (2018).

Although F-score improvements are modest, they are trending upward, and qualitative analysis of the output of the system making use of in-domain training data shows notable improvements

Parser	Data	Precision	Recall	F-score
CAMR	1.0	0.33	**0.51**	**0.40**
JAMR	1.0	0.27	0.44	0.33
JAMR	2.0	0.46	0.28	0.35
JAMR	2.0+D	**0.56**	0.27	0.36

Table 3: Parser performances on human-robot dialogue test set after being trained on AMR 1.0, AMR 2.0 corpus and on AMR 2.0 corpus combined with small in-domain training set of human-robot dialogue data.

[4] https://github.com/jflanigan/jamr
[5] https://github.com/c-amr/camr
[6] https://github.com/BLLIP/bllip-parser

in some of the common navigation-related language. For example, compare the output of the system trained on AMR 2.0 to the system trained on AMR 2.0 plus in-domain data for a common instruction, shown in Figure 8.

```
1. (m / move-01
      :ARG1 (f / foot
            :quant 15)
      :direction (f2 / forward))

2. (m / move-01
      :direction (f2 / forward)
      :extent (d / distance-quantity
            :quant 15
            :unit (f / foot)))
```

Figure 8: (1) Output from JAMR trained on AMR 2.0 for *move forward 15 feet.* Note that *foot* is incorrectly represented as the ARG1 of move, or the *thing-moved*. (2) Output from JAMR trained on AMR 2.0 plus in-domain data. Note that *15 feet* is correctly treated as an extent of the movement

Despite improvements, the system trained on the small sample of in-domain data still fails to represent :mode imperative and also fails to include implicit subjects. Thus, we conclude that additional data more similar to the corpus is still needed, and we are currently working with other research groups to develop a larger training sample of human-agent dialogue that includes movement direction-giving. However, note that we do not yet know what downstream impact improvements in F-score will have on the final system. Since we do not plan for the robot to act upon parser output AMRs, but rather in-domain AMRs, it may be that the a graph-to-graph transformation algorithm could be robust to some noise in the parser output but still map to the correct in-domain AMR.

6.2 Graph-to-Graph Transformations

We are in the early stages of exploring graph-to-graph transformations that will allow us to move from the parser-output AMRs to our set of in-domain AMRs. Rather than train parsers to parse directly into the augmented AMRs described here, a graph-to-graph transformation allows us to maintain the parser output as a representation of the sentence meanings themselves as input, while the output captures our contextual domain-specific layer and includes speaker intent on top of the sentence meaning. To create training data for graph-to-graph transformation algorithms and to evaluate the coverage and quality of the set of in-domain AMRs, we have begun this exploration by manually mapping a set of our gold-standard AMRs to the 36 in-domain AMR templates.[7]

Necessary transformations so far include the following: i) changing participant roles, for example *I/you* to *robot/commander*); ii) creating a merge step for all actions of similar type, for example merging movement commands of *move, go, walk, back up* into the go-02 frame (following our command:move template); and iii) expanding AMR frames to include implicit roles. Next steps will include the general tasks of pairing utterances with one of the 36 speech act types, making use of linguistic cues (for example, when an utterance lacks a personal pronoun or named entity like "robot", it is likely a command), and identifying when a command is not executable and further information is necessary.

6.3 Revising, Adapting NLU Component

In previous work using the same human-robot dialogue corpus, Lukin et al. (2018) implemented a preliminary dialogue system which uses a statistical classifier for NLU (NPCEditor, Leuski and Traum, 2011). The classifier relies on language model similarity measures to associate an instruction with either a "translation" to be sent forward to the RN-Wizard or a clarification question to be returned to the participant. The system also exploits the dialogue structure annotations (section 2.2) as features. Error analysis has demonstrated that this preliminary system, by simply learning an association between an input string and a particular set of executed actions, fails to generalize to unseen, novel input instructions (e.g, *Turn left 100 degrees,* as opposed to a more typical number of degrees like 90), and is unable to interpret instructions with respect to the current physical surroundings (e.g., the destination of *Move to the door on the left* needs to be interpreted differently depending where the robot is facing).

Our proposed domain-specific AMRs from section 5 are intended as a replacement for the classifier functionality of the current preliminary dialogue system, allowing a much richer representation of the semantics of actions, including allowing previously unseen values, and compositional construction of referring expressions. A downstream dialogue manager component will be

[7] We plan to eventually model our graph-to-graph transformation on work by (Liu et al., 2015) for abstractive summarization with AMR, though in the opposite direction.

able to perform slot-filling dialogue (Xu and Rudnicky, 2000) including clarification of missing or vague descriptions and, if all required parameters are present, will use the domain-specific AMR for robot execution.

7 Related Work

7.1 Semantic Representation

There is a long-standing tradition of research in semantic representation within NLP, AI, as well as theoretical linguistics and philosophy (see Schubert (2015) for an overview). Thus, there are a variety of options that could be used within dialogue systems for NLU. However, for many of these representations, there are no existing automatic parsers, limiting their feasibility for larger-scale implementation. An exception is combinatory categorical grammar (CCG) (Steedman and Baldridge, 2011); CCG parsers have been incorporated in some current dialogue systems (Chai et al., 2014). Although promising, CCG parses closely mirror the input language, so systems making use of CCG parses still face the challenge of a great deal of linguistic variability that can be associated with a single intent. Universal Conceptual Cognitive Annotation (UCCA) (Abend and Rappoport, 2013), which also abstracts away from syntactic idiosyncrasies, and its corresponding parser (Hershcovich et al., 2017) merits future investigation.

7.2 NLU in Dialogue Systems

Broadly, the architecture of task-oriented spoken dialogue systems includes i) automatic speech recognition (ASR) to recognize an utterance, ii) an NLU component to identify the user's intent, and iii) a dialogue manager to interact with the user and achieve the intended task (Bangalore et al., 2006). The meaning representation within such systems has, in the past, been predefined frames for particular subtasks (e.g., flight inquiry), with slots to be filled (e.g., destination city) (Issar and Ward, 1993). In such approaches, the meaning representation was crafted for a specific application, making generalizability to new domains difficult if not impossible. Current approaches still model NLU as a combination of intent and dialogue act classification and slot tagging, but many have begun to incorporate recurrent neural networks (RNNs) and some multi-task learning for both NLU and dialogue state tracking (Hakkani-Tür et al., 2016; Chen et al., 2016), the latter of which allows the system to take advantage of information from the discourse context to achieve improved NLU. Substantial challenges to these systems include working in domains with intents that have a large number of possible values for each slot and accommodation of out-of-vocabulary slot values (i.e. operating in a domain with a great deal of linguistic variability).

7.3 Speech Act Taxonomies for Dialogue

Speech acts have been used as part of the meaning representation of task-oriented dialogue systems since the 1970s (e.g., Bruce, 1975; Cohen and Perrault, 1979; Allen and Perrault, 1980). For a summary of some of the earlier work in this area, see Traum (1999). There have been a number of widely used speech act taxonomies, including an ISO standard (Bunt et al., 2012), however these often have to be particularized to the domain of interest to be fully useful. Our approach with speech act types and subtypes representing a kind of semantic frame is perhaps most similar to the *dialogue primitives* of Hagen and Popowich (2000). Combining these types with fully compositional AMRs will allow flexible expressiveness, inferential power and tractable connection to robot action.

8 Conclusions

This paper has proposed refinements for AMR to encode information necessary for situated human-robot dialogue. Specifically, we elaborate 36 templates specific to situated dialogue that capture i) tense and aspect information; ii) speech acts; and iii) spatial parameters for robot execution. These refinements come after evaluating the coverage of existing AMR for a corpus of human-robot dialogue elicited from tasks related to search-and-rescue and reconnaissance. We also manually annotated 641 in-domain gold standard AMRs in order to evaluate and retrain existing AMR parsers, JAMR and CAMR, for performance on dialogue data. Future work will continue to annotate situated dialogue data and assess the performance of both a graph-to-graph transformation algorithm and an existing statistical classifier for eventual, autonomous human-robot collaboration. We plan to make our AMR-annotated data publicly available; please contact the authors if you would like access to it beforehand.

Acknowledgments

We are grateful to anonymous reviewers for their feedback and to Jessica Ervin for her contributions to the early stages of this research. The second author was sponsored by the U.S. Army Research Laboratory (ARL) under the Advanced Research Technology, Inc. contract number W911QX-18-F-0096; the fifth and sixth authors were sponsored by ARL under contract number W911NF-14-D-0005.

References

Omri Abend and Ari Rappoport. 2013. Universal Conceptual Cognitive Annotation (UCCA). In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 228–238.

James F Allen and C Raymond Perrault. 1980. Analyzing intention in utterances. *Artificial Intelligence*, 15(3):143–178.

John Langshaw Austin. 1975. *How to Do Things with Words*, 2nd edition. Harvard University Press and Oxford University Press.

Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. 2013. Abstract Meaning Representation for sembanking. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pages 178–186.

Srinivas Bangalore, Dilek Hakkani-Tür, and Gokhan Tur. 2006. Introduction to the special issue on spoken language understanding in conversational systems. *Speech Communication*, 48(3–4):233–238.

Bertram C. Bruce. 1975. Generation as a social action. In *Theoretical Issues in Natural Language Processing*, pages 64–67.

Harry Bunt, Jan Alexandersson, Jae-Woong Choe, Alex Chengyu Fang, Koiti Hasida, Volha Petukhova, Andrei Popescu-Belis, and David Traum. 2012. ISO 24617-2: A semantically-based standard for dialogue annotation. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 430–437.

Shu Cai and Kevin Knight. 2013. Smatch: an evaluation metric for semantic feature structures. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, volume 2, pages 748–752.

Joyce Y. Chai, Lanbo She, Rui Fang, Spencer Ottarson, Cody Littley, Changsong Liu, and Kenneth Hanson. 2014. Collaborative effort towards common ground in situated human-robot dialogue. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, pages 33–40. ACM.

Yun-Nung Chen, Dilek Hakkani-Tür, Gokhan Tur, Jianfeng Gao, and Li Deng. 2016. End-to-end memory networks with knowledge carryover for multi-turn spoken language understanding. In *Interspeech 2016*, pages 3245–3249.

Philip R Cohen and C Raymond Perrault. 1979. Elements of a plan-based theory of speech acts. *Cognitive science*, 3(3):177–212.

Lucia Donatelli, Michael Regan, William Croft, and Nathan Schneider. 2018. Annotation of tense and aspect semantics for sentential AMR. In *Proceedings of the Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions (LAW-MWE-CxG-2018)*, pages 96–108.

Jeffrey Flanigan, Sam Thomson, Jaime Carbonell, Chris Dyer, and Noah A Smith. 2014. A discriminative graph-based parser for the Abstract Meaning Representation. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1426–1436.

Eli Hagen and Fred Popowich. 2000. Flexible speech act based dialogue management. In *1st SIGdial Workshop on Discourse and Dialogue*, pages 131–140.

Dilek Hakkani-Tür, Gokhan Tur, Asli Celikyilmaz, Yun-Nung Chen, Jianfeng Gao, Li Deng, and Ye-Yi Wang. 2016. Multi-domain joint semantic frame parsing using bi-directional RNN-LSTM. In *Interspeech 2016*, pages 715–719.

Daniel Hershcovich, Omri Abend, and Ari Rappoport. 2017. A transition-based directed acyclic graph parser for UCCA. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1127–1138.

Sunil Issar and Wayne Ward. 1993. CMU's robust spoken language understanding system. In *Third European Conference on Speech Communication and Technology (Eurospeech 93)*, pages 2147–2150.

Anton Leuski and David Traum. 2011. NPCEditor: Creating virtual human dialogue using information retrieval techniques. *AI Magazine*, 32(2):42–56.

Fei Liu, Jeffrey Flanigan, Sam Thomson, Norman Sadeh, and Noah A Smith. 2015. Toward abstractive summarization using semantic representations. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*.

Stephanie M. Lukin, Felix Gervits, Cory J. Hayes, Anton Leuski, Pooja Moolchandani, John G. Rogers, III, Carlos Sanchez Amaro, Matthew Marge,

Clare R. Voss, and David Traum. 2018. ScoutBot: A Dialogue System for Collaborative Navigation. In *Proceedings of ACL 2018, System Demonstrations*, pages 93–98, Melbourne, Australia.

Chunchuan Lyu and Ivan Titov. 2018. Amr parsing as graph prediction with latent alignment. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, pages 397–407.

Christopher Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven Bethard, and David McClosky. 2014. The Stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 55–60.

Matthew Marge, Claire Bonial, Brendan Byrne, Taylor Cassidy, A. William Evans, Susan G. Hill, and Clare Voss. 2016. Applying the Wizard-of-Oz technique to multimodal human-robot dialogue. In *RO-MAN 2016: IEEE International Symposium on Robot and Human Interactive Communication*.

Matthew Marge, Claire Bonial, Ashley Foots, Cory Hayes, Cassidy Henry, Kimberly Pollard, Ron Artstein, Clare Voss, and David Traum. 2017. Exploring variation of natural human commands to a robot in a collaborative navigation task. In *Proceedings of the First Workshop on Language Grounding for Robotics*, pages 58–66.

Tim O'Gorman, Michael Regan, Kira Griffitt, Ulf Hermjakob, Kevin Knight, and Martha Palmer. 2018. AMR beyond the sentence: The multi-sentence AMR corpus. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 3693–3702.

Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. The proposition bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31(1):71–106.

Penman Natural Language Group. 1989. The Penman user guide. *Technical report, Information Sciences Institute*.

Lenhart K Schubert. 2015. Semantic representation. In *AAAI*, pages 4132–4139.

John Rogers Searle. 1969. *Speech acts: An essay in the philosophy of language*. Cambridge University Press.

Mark Steedman and Jason Baldridge. 2011. Combinatory categorial grammar. In Robert Borsley and Kersti Börjars, editors, *Non-Transformational Syntax: A Guide to Current Models*, chapter 5, pages 181–224. Wiley-Blackwell.

David Traum, Cassidy Henry, Stephanie Lukin, Ron Artstein, Felix Gervits, Kimberly Pollard, Claire Bonial, Su Lei, Clare Voss, Matthew Marge, Cory Hayes, and Susan Hill. 2018. Dialogue structure annotation for multi-floor interaction. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, pages 104–111, Miyazaki, Japan. European Language Resources Association (ELRA).

David R. Traum. 1999. Speech acts for dialogue agents. In Anand Rao and Michael Wooldridge, editors, *Foundations of Rational Agency*, pages 169–201. Kluwer.

David R. Traum and Staffan Larsson. 2003. The information state approach to dialogue management. In Jan van Kuppevelt and Ronnie W. Smith, editors, *Current and new directions in discourse and dialogue*, pages 325–353. Springer.

Chuan Wang, Nianwen Xue, and Sameer Pradhan. 2015. A transition-based algorithm for AMR parsing. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 366–375.

Wei Xu and Alexander I. Rudnicky. 2000. Task-based dialog management using an agenda. In *Proceedings of the 2000 ANLP/NAACL Workshop on Conversational systems-Volume 3*, pages 42–47.

Appendix

Type	Subtype	Example
Command	Move	Move forward 5 feet
	Turn	Turn left 90 degrees
	Send-Image	Take a picture
	Repeat	Do that again
	Cancel	Cancel that
	Stop	Ok stop here
Assert	Move	I will move forward 5 feet
	Turn	I turned right 90 degrees
	Send-Image	Sent
	Do	Executing...
	Confirm	Correct
	Scene	I see two doorways ahead
	Ability	I can't manipulate objects
	Map	The table is 2 feet away
	Task	Calibration complete
Request	Wait	Please wait
	Confirm	I'll go as far as I can, ok?
	Clarify	Can you describe it another way?
	Instruct	What should we do next?
Question	Ability	Can you speak Arabic?
	Scene	Have you seen any shoes?
	Map	How far are you from wall?
Express	Greet	Hello!
	Thank	Thanks for the help!
	Good	Good job!
	Mistake	Woops!
	Sorry	Sorry!

Table 4: Listing of Speech Act Types and Subtypes (actions), with example utterances. Note that each subtype corresponds to a unique augmented AMR template. 27 subtypes are listed here; the Assert-Task subtype has several subtypes of its own, which are omitted here.