ARL-TR-9656 • MAR 2023

DEVCOM ARMY RESEARCH LABORATORY

Bot Language (Summary Technical Report, Oct 2016–Sep 2021)

by Matthew Marge, Claire Bonial, Stephanie Lukin, and Clare Voss

NOTICES

Disclaimers

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.

ARL-TR-9656 ● MAR 2023

Bot Language (Summary Technical Report, Oct 2016–Sep 2021)

Matthew Marge, Claire Bonial, Stephanie Lukin, and Clare Voss
DEVCOM Army Research Laboratory

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED (From - To)
March 2023	Summary Technical Report	October 2016–September 2021

4. TITLE AND SUBTITLE
Bot Language (Summary Technical Report, Oct 2016–Sep 2021)

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Matthew Marge, Claire Bonial, Stephanie Lukin, and Clare Voss

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
DEVCOM Army Research Laboratory
ATTN: FCDD-RLC-IT
Adelphi, MD 20783

8. PERFORMING ORGANIZATION REPORT NUMBER
ARL-TR-9656

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release: distribution unlimited.

13. SUPPLEMENTARY NOTES
ORCID IDs: Matthew Marge, 0000-0001-7672-779X; Claire Bonial, 0000-0002-3154-2852; Stephanie Lukin, 0000-0001-8761-167X; Clare Voss, 0000-0001-5023-6474

14. ABSTRACT
This report provides a comprehensive summary of the contributions made as part of the Bot Language project, a 5-year US Army Combat Capabilities Development Command Army Research Laboratory-led initiative in partnership with researchers at the University of Southern California's Institute for Creative Technologies and Carnegie Mellon University. In particular, this report describes accomplishments funded under the project "Naturalistic Behavior for Shared Understanding and Explanation with Intelligent Systems." The goal of this research was to provide more natural ways for people to communicate with robots using language. Our vision was to enable robots to engage in a back-and-forth dialogue with human teammates where robots can provide status updates and ask for clarification where appropriate. To this end, we conducted a phased progression of four experiments where human participants were giving navigation instructions to a remotely located robot, while the robot's dialogue and navigation processes were initially controlled by human experimenters. Over the course of the experiments, automation was progressively introduced until dialogue processing was completely driven by a classifier trained on the data collected in previous experiments.

15. SUBJECT TERMS
dialogue, human–robot interaction, human factors, and natural language processing, Humans in Complex Systems, Military Information Sciences, Summary Technical Report, STR

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Matthew Marge
Unclassified	Unclassified	Unclassified	UU	46	19b. TELEPHONE NUMBER (Include area code) 301-394-5787

Contents

List of Figures

List of Tables

Acknowledgments

The authors thank all the contributors to the project, including those from US Army Combat Capabilities Development Command (DEVCOM) Army Research Laboratory (ARL) (Anthony L Baker, David Baran, Austin Blodgett, Arthur William Evans III, Ashley Foots, Felix Gervits, Jason Gregory, Cory J Hayes, Susan G Hill, Reginald Hobbs, Taylor Hudson, Brandon Perelman, Kimberly A Pollard, John G Rogers III, and Douglas Summers-Stay); University of Southern California's Institute for Creative Technologies (Principal Investigators [PIs]: David Traum, Ron Arstein, Jill Boberg, Carla Gordon, Anton Leuski, and Volodymyr Yanov); and Carnegie Mellon University (PIs: Louis-Philippe Morency, Elif Bozkurt, and Chirag Raman).

The authors also thank our student contributors: Mitchell Abrams (2019–2021, now PhD student and SMART Scholar at Tufts University), Carlos Sanchez Amaro (2017–2018, now Quality Engineer at Howmet Aerospace), Lucia Donatelli (2018–2019, now Postdoctoral Fellow at Saarland University), Jessica Ervin (2018, now Software Engineer at FactSet), Felix Gervits (2017, now Computer Scientist at ARL), Cassidy Henry (2016–2018, now PhD student and SMART Scholar at the University of Maryland), Taylor Hudson (2020–2021, now Jr Computational Linguist at ARL), Darius Jefferson II (2016, now Computer Scientist at ARL), Brecken Keller (2018–2020, now Associate Developer at Crum and Forster), Elia Martin (2018–2019), and Pooja Moolchandani (2016–2018, now Data Analytics Consultant at KPMG).

The authors also thank Brendan Byrne, Taylor Cassidy, John J Morgan, and Adam Wiemerslage for their past contributions to the project.

Finally, the authors also acknowledge the otherwise invisible, invaluable, behind-the-scenes support and sustained commitment to this project from our branch and division chiefs: V Melissa Holland, Stuart Young, and Barbara Broome.

Executive Summary

This report provides a comprehensive summary of the contributions made as part of the Bot Language project, a 5-year initiative led by the US Army Combat Capabilities Development Command Army Research Laboratory in partnership with researchers at the University of Southern California's Institute for Creative Technologies and Carnegie Mellon University. In particular, this report describes accomplishments funded under the project "Naturalistic Behavior for Shared Understanding and Explanation with Intelligent Systems." The goal of this research is to provide more natural ways for people to communicate with robots using language. Our vision is to enable robots to engage in a back-and-forth dialogue with human teammates where robots can provide status updates and ask for clarification where appropriate. To this end, we conducted a phased progression of four experiments where human participants gave navigation instructions to a remotely located robot, while the robot's dialogue and navigation processes were initially controlled by human experimenters. Over the course of the experiments, automation was progressively introduced until dialogue processing was completely driven by a classifier trained on the data collected in previous experiments.

The novel contributions of the Bot Language project include 1) this multiphased approach to collecting unconstrained natural language as training data for machine learning algorithms to support conversational interactions, 2) the corpora of dialogue and robot data collected and curated into the SCOUT Corpus (Situated Corpus of Understanding Transactions), 3) a series of fully automated, proof-of-concept systems that show the technical promise of the approach taken, 4) the algorithms created as part of the project that now form the basis for the Army's Joint Understanding and Dialogue Interface capability enabling conversational interactions between Soldiers and autonomous systems, and 5) innovations in the semantics of instructions in human–robot dialogue through the Dialogue-AMR (Abstract Meaning Representation) formalism.

1. Introduction

The focus of this research is to make Soldier–agent interactions, especially with embodied agents such as robots, both safe and more effective by employing dialogue as a mode of communication. Dialogue, specifically back-and-forth verbal conversation using natural language, offers many benefits over traditional graphical user interfaces. Among these, dialogue enables agents to prompt a human teammate for clarification if a directive is unclear, and also to provide status updates as tasks are completed. Natural language dialogue can help achieve the vision of intelligent agents serving as teammates alongside Soldiers by offering an intuitive unconstrained mode of communication as used by Soldiers today in completing missions.

With the goal of collecting natural conversations with intelligent agents, we wanted an experimental approach that would enable us to address the following questions: 1) How can *agents* communicate effectively as teammates with humans to accomplish shared tasks? and 2) How can the *protocol for exchanges* elicit the natural diversity of communication strategies from *humans,* as they instruct agents such as robots, in a form that agents can use? To answer these questions, we worked with researchers at the University of Southern California's Institute for Creative Technologies (USC ICT), an Army University Affiliated Research Center, to determine experimentally how methods from the development of intelligent virtual humans could be adapted for robots. While physical robotic platforms motivated our main task, we aimed to identify methods that generalize to a variety of software agents that can benefit from dialogue.

In USC ICT's SimSensei[1] project, researchers used a methodology we call Data-driven "Wizard-of-Oz" (DWoZ) to observe how humans would chat with what they believed to be an autonomous virtual avatar. In reality, the avatar they saw on their screen was controlled by human "wizard" experimenters. In collaboration with USC ICT, our goal was to assess if these contributions could be extended to autonomous systems, namely ground robots, to support collaborative search and navigation tasks with human teammates. This project, sponsored under the US Army Combat Capabilities Development Command (DEVCOM) Army Research Laboratory (ARL) funding line "Naturalistic Behavior for Shared Understanding and Explanation with Intelligent Systems," and known externally as the "Bot Language" project, consisted of a series of experiments that executed the vision of multiphased experimentation where wizards standing in for artificial intelligence (AI) components were "automated away" in later phases. The operational hypothesis was that dialogue systems for physical agents like mobile robots could be trained from DWoZ-based dialogue collection.

The novel contributions of this research to the fields of dialogue, human–robot interaction, human factors, and natural language processing were the following:

- A multiphased, empirical approach to collecting training data for machine learning algorithms that support conversational interaction with intelligent agents that refer to the physical world (e.g., mobile robots) (Sections 4 and 5)
- A corpus of dialogue and robot data (Situated Corpus of Understanding Transactions [SCOUT]) that serves as a basis for informing intelligent agents on how to respond to human teammates in collaborative search and navigation tasks (Section 6.1)
- A series of fully automated, end-to-end, proof-of-concept systems developed over the course of the research that show the technical promise of natural conversational interactions with intelligent agents using the DWoZ approach (Section 6.2)
- Algorithms created as part of the project that now form the basis for the Army's Joint Understanding and Dialogue Interface (JUDI) capability enabling conversational interactions between Soldiers and autonomous systems (Section 6.3)
- A set of novel annotation schemes that model the structure, content, and semantics of dialogue exchanges between participants that instruct intelligent agents and wizard experimenters that control the robot's behavior (Section 6.4)

The remainder of this report is organized as follows. Section 2 provides a basic overview in related work. Section 3 relates prior research and pre-pilot studies conducted in advance of this project to the selected configuration of the DWoZ design. Section 4 overviews the task and experimental setup. High-level descriptions of the experiments and their findings are provided in Section 5. Finally, a discussion about the impact of the project can be found in Section 6, with metrics in Section 7, and concluding thoughts in Section 8.

2. Related Work

2.1 Dialogue for Human–Robot Interaction

Although natural language-based interaction has been explored extensively in the field of human–robot interaction,[2,3] the primary focus has been on creating algorithms for automatic processing of one direction of communication (i.e., interpreting human instructions or producing responses), but not both directions at the same time.

For the goal of interpreting human instructions, researchers followed the methodology of corpus-based robotics,[4] where experiments collect verbal or written instructions. The core algorithms rely on computational techniques for natural language understanding (e.g., Kruijff et al.[5] and Williams et al.[6]) and symbol grounding that maps language to symbolic representations used for task planning (e.g., Tellex et al.[7] and Hemachandra et al.[8]).

Limited effort has been done to create response generation algorithms for robot-to-human communications beyond templates written by system developers. Some focused on ways robots can explain tasks[9] and paths[10,11] to people using natural language. Meanwhile, others focused on clarification algorithms[12] about objects and asking for help with collaborative tasks.[13]

Several dialogue-based interfaces were developed for mobile robots, such as DIARC[14] and TeamTalk,[15] but most rely on handcrafted templates to select responses or synthetic training data. Our work builds upon this related work by investigating empirical methods to human–robot dialogue collection, which balances eliciting robot-directed natural language from participants while maintaining a tractable data set that can be used for training a dialogue system.

2.2 Wizard-of-Oz Methodology

The Wizard-of-Oz (WoZ) design methodology has been used for many years in human–computer interaction research, including research involving natural language interfaces, to inform design specifications for technologies that have not yet been implemented. WoZ's costs are limited to the costs of wizard experimenters standing in for future technologies and provides a very malleable way to alter system functionality—the only requirement being to change the policies that a wizard would follow. WoZ has been used for simulating dialogue interfaces,[16] such as for human–robot interaction.[17] Wizards also can play a role in collecting dialogue clarification strategies.[18] Similarly, our research partners at USC ICT have used WoZ to collect verbal and nonverbal behaviors to train algorithms for a virtual

human therapist in the SimSensei project.[1] Our work expands on these methods by exploring multimodal communication strategies when the robot and human are not co-present and must complete tasks together as a team where information such as location, visual content, and dialogue would be exchanged.

3. Background

Prior to the start of this project, the Army invested in research efforts for human–robot and human–agent natural language communication. One such project was the 5-year Army Research Office (ARO)-funded Multidisciplinary University Research Initiative (MURI) titled SUBTLE (Situation Understanding Bot through Language and Environment).[19,20] Following this successful effort, the question arose as to how the results might transition into ARL for further research. While the SUBTLE MURI yielded software integrated into a proof-of-concept system, several capabilities were intentionally left out of the system design, including automatic speech recognition and dialogue management software. Speech recognition was deemed insufficiently robust at the time but evolving at such a rapid pace that, in practical terms, the researchers decided others in the future could incorporate it into the system. They also determined that incorporating dialogue management was not feasible, as it was contingent on another capability they needed to develop first—a syntactic parser-plus-semantic analyzer module that would interpret natural language commands. As a consequence of these two design decisions, there was no practical way to collect spoken-language dialogue data sets.

Following the SUBTLE MURI, ARL researchers conducted two preliminary studies (i.e., pre-pilots) to determine how best to record, identify, and track the dialogue, video, and light detection and ranging (LIDAR) information that was explicitly shared by, or indirectly available to, two members of a human–robot team when conducting a collaborative search task. The result of lessons learned in those efforts led to the WoZ configuration of the Bot Language project. In the first pre-pilot, volunteers were enlisted to be the direction giver (i.e., "Commander") or the direction follower (i.e., "Robot Navigator") controlling a remotely located mobile robot.[21] Only the Robot Navigator could "see" for the robot, via its onboard video camera and LIDAR sent back to the Robot Navigator's display. The Robot Navigator was instructed to act as though they were situated in the robot's position and to obey the Commander. The Robot Navigator was to consider the robot's actions as their own, and to consider available video and LIDAR point cloud feeds as their own perceptions. The Commander and Robot Navigator communicated by text chat on their computers.

Follow-on discussions with volunteers indicated both that relying on the text chat communication was too slow for real-time navigation of the robot and that limiting the Commander to text-only information from the Robot Navigator was frustrating and left them "in the dark." To address these problems, in the second study, two changes were made to speed up the communication and expand the Commander's understanding of the robot's environment. The Commander and Robot Navigator participants now spoke to each other, though their roles remained the same with the Robot Navigator doing both the robot navigation and dialogue handling. The transmission of LIDAR map data and video stream was allowed to pass uninterrupted from the robot's sensors to the Commander under various conditions controlling for visual information that both could see. Participants were generally able to use both image and map data in conjunction with dialogue to build enough shared understanding to communicate about the environment and accomplish the exploration tasks at hand.[22] These changes led to dialogue speed-up, as expected. However, with this increased level of engagement came an unintended side effect: volunteers participating as the Robot Navigator found it exceedingly challenging to shift their attention between navigating the robot with image and LIDAR information and participating in the dialogue that required attention to coordinate turn-taking in conversation with the Commander.[23] This challenge spurred the decision already under consideration to "split" the Robot Navigator's stand-in role as robot to two wizards—one for dialogue management and one for navigation (Fig. 1). Section 4 describes the revised approach used for the Bot Language experiments.

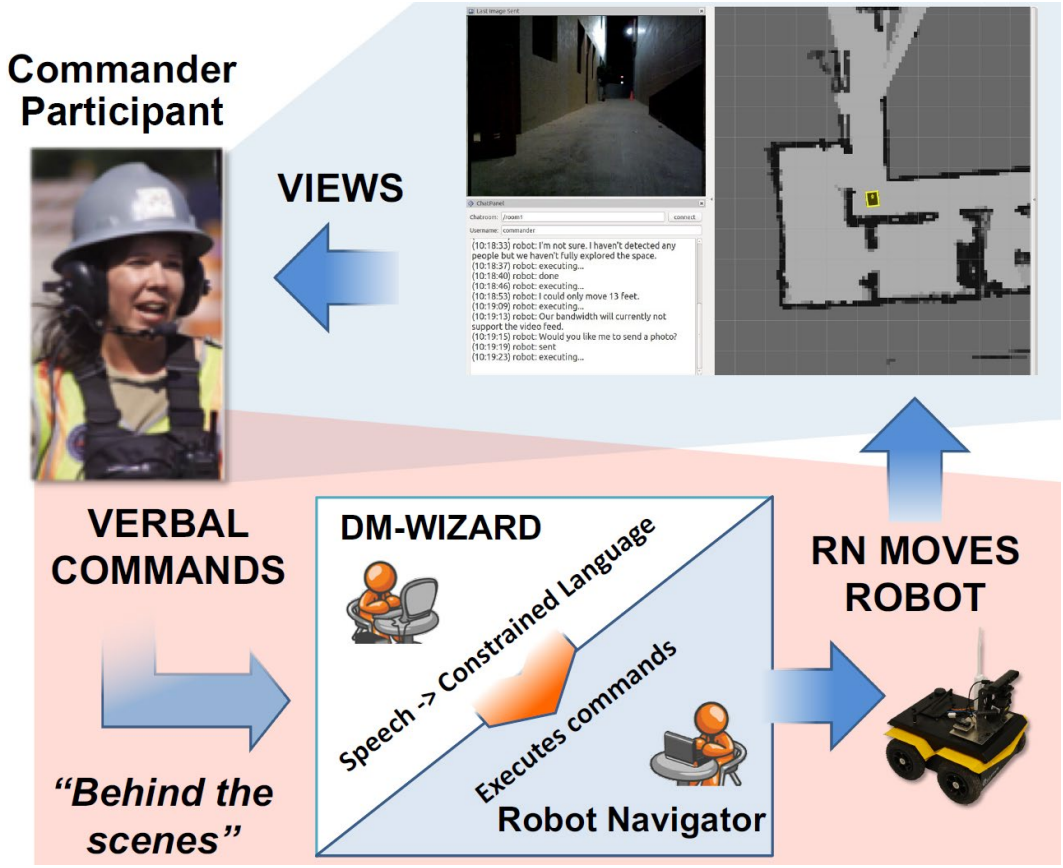

Fig. 1 The Commander issues verbal instructions to the robot, whose capabilities are performed by two wizards standing in for the respective abilities of dialogue management and robot navigation. (Original figure from Marge et al.[24])

4. Approach

This section overviews the task and experiment setup. For additional details, please see Marge et al.[24]

4.1 Task Domain

In this research, a human teammate called the "Commander" instructs a remotely located robot to explore a building. The rate of information exchange was restricted to simulate a low-bandwidth environment similar to the contested networking situations that Soldiers are expected to experience in the future battlefield. The Commander was expected to perform a building reconnaissance, that is, assess the structure of the building based on information sent from the robot, decide on the robot's actions, and determine if the building has been recently occupied. To accomplish this, the Commander would speak to the robot (e.g., "turn left 90 degrees," or "go through the door"). Direct teleoperation was not permitted due

to the low-bandwidth scenario. Unlike many of today's modes of human–robot interaction, the Commander could neither observe the robot directly nor see a video feed from its camera. Instead, the robot could respond in natural language and provide information that can feasibly be sent over a low-bandwidth network—a live map with the robot's location and an occupancy grid, and images captured from the robot's camera that can be sent upon request (see Fig. 1, upper right).

4.2 Multi-Wizard Setup

As mentioned in Section 2, we use the WoZ methodology to collect the necessary training data in our experiments. In the initial phases of the project, there were two wizards. Each wizard takes the role of what were ultimately separate modules in an autonomous system. A "Dialogue Manager Wizard" (DM-Wizard) would listen to the Commander's verbal instructions and respond using a chat window to send status updates and clarification questions. As long as the instruction was executable in the current context, the DM-Wizard would send a constrained version of the instruction to a second wizard called the "Robot Navigator Wizard" (RN-Wizard). The RN-Wizard would directly teleoperate the robot without the Commander's knowledge. As the RN-Wizard moved the robot, they would convey status updates verbally back to the DM-Wizard, which would then be passed back to the Commander as status updates. See Fig. 1 for an illustration of this setup.

4.3 Multiphased Approach

This project takes a multiphased approach to creating automated dialogue capabilities for autonomous systems; in initial phases, wizards play the role of components that are automated in later phases. These phases can be summarized in Table 1. Experiments with human subjects served as a way to not only collect training data for the automated components but served as milestones to measure our progress. By convention, we ordered experiments numerically. Experiment 1 served as our initial exploratory phase where we wanted to collect the full range of communications that may take place in the task domain. The DM-Wizard would type responses to Commander instructions in real time based on a set of guidelines that were established during experiment piloting. The guidelines specified that executable instructions were those that contained both a clear instruction and a specified endpoint. Thus, open-ended instructions equivalent to verbal teleoperation (e.g., "move forward") were not acceptable. The DM-Wizard faced several challenges in Experiment 1: not only did they have to respond as quickly as possible to both the Commander and RN-Wizard, but to physically type these messages with as few typographical errors as possible to avoid impacting the Commander's perception of autonomy.

Table 1 Testing scenarios over time. Columns indicate progression of testing scenario experimentation and development; rows represent scenario components. (Original table from Marge et al.[28]).

	Exp 1 [25]	Exp 2 [26]	Exp 3 completed 2018	Exp 4 completed 2019	ScoutBot [27]	MultiBot [28]
Dialogue Processing	wizard + typing	wizard + button presses	wizard + button presses	ASR + auto-DM	ASR + auto-DM	ASR + auto-DM
Robotic Behaviors	wizard + joystick	wizard + joystick	wizard + joystick	wizard + joystick	finite state machine	auto-assign via TBS
Robot(s)	1 physical	1 physical	1 simulated	1 simulated	1 simulated	2 simulated
Environment	indoors + real building	indoors + real building	indoors + sim building	indoors + sim building	indoors + sim building	outdoors + sim buildings

Notes: DM: Dialogue Management; ASR: Automatic Speech Recognition; TBS: Tactical Behavior Specification.

The Experiment 1 data were analyzed to identify DM-Wizard messages and response templates that balance tractability for an autonomous system to use and full coverage of responses to what participants were likely to say in the task domain. This included strategies the DM-Wizard could use to ask for clarification and recover from problematic instructions. In Experiment 2, this set of messages was incorporated into a click-button graphical interface (see Fig. 2), substantially reducing typing and composition effort by the DM-Wizard. The graphical interface provided greater uniformity in responses.

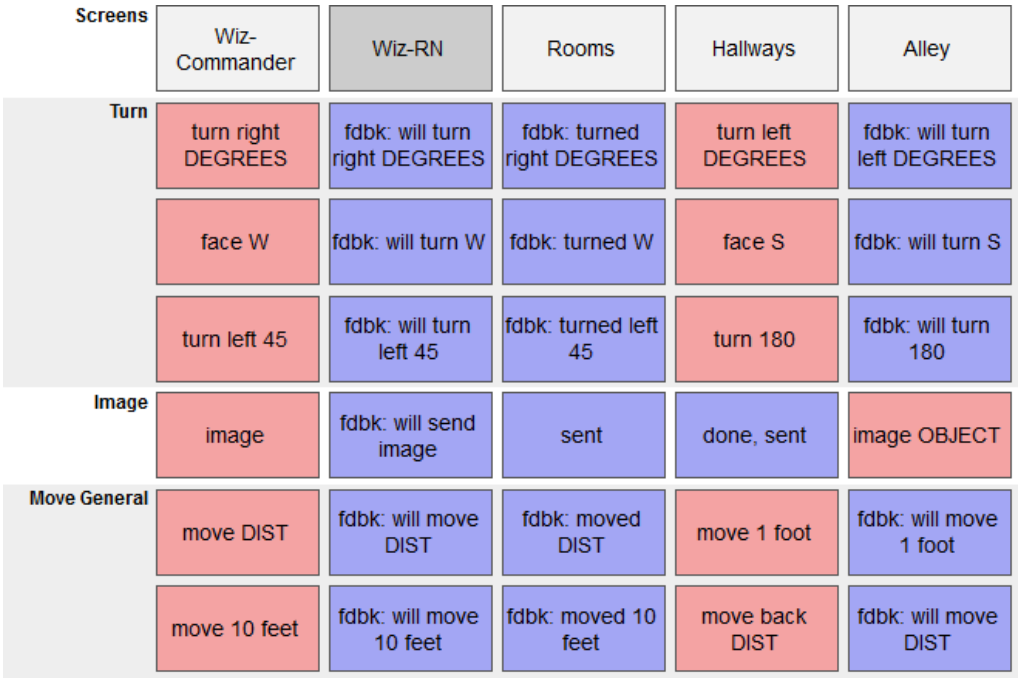

Fig. 2 Excerpt of a Wizard graphical interface for handling commands and composing replies to participants. Blue buttons reply to the Commander participant, while red buttons route messages to the RN-Wizard that teleoperates the robot. CAPS indicate text-input slots.

The first two experiments were done in a real-world indoor environment with a physical robot (Clearpath Robotics Jackal). While this resulted in realistic and useful data for analysis, the pace of data collection was limited to the availability of the physical space used in data collection. Previous work showed that instruction-giving in virtual and real-world environments are similar,[29] so we explored the option of a virtual environment setup in Experiment 3. Like the first two experiments, the Commander would have access to the same visual information (e.g., a live map, pictures from the robot's camera, and text dialogue), only it was simulated in the Gazebo high-fidelity virtual simulator.[30] This included the physics of objects and views from a virtual version of the robot and its camera. From the Commander's perspective, the study was equivalent to the previous two, with the exception that the images from the camera were virtually rendered. Shifting the experiment setup to simulation offered several benefits: 1) there was no longer a dependency on the availability of physical space or robot platforms, 2) technical problems were far less likely in a controlled virtual environment, and 3) data collection could be done in parallel at multiple experiment sites. During this time, an additional collaborator (Carnegie Mellon University) joined the team to help in analyzing emotional expression and face pose data from participants during the study. Beyond this difference, the experiment followed the protocol and experiment design of the first two studies.

Data collection during Experiment 3 resulted in sufficient training data to train an automated component that replaced the DM-Wizard. Instead of a human experimenter listening to the Commander's speech and routing messages back and to the RN-Wizard, in Experiment 4 a dialogue system provided these capabilities. This final study had only the RN-Wizard as a wizard experimenter to teleoperate the robot in response to instructions provided to it by the dialogue system. The system itself was divided into two main components: 1) a speech recognition component that would interpret speech in real time from the Commander, and 2) a dialogue manager classifier that would determine the Commander's intent from their speech using training data collected in the previous studies. Additional technical details regarding the classifier can be found in Gervits et al.[31] The system used in this study was the precursor to JUDI, used by ARL's Artificial Intelligence for Maneuver and Mobility (AIMM) Essential Research Program (ERP).

5. Data Collection Experiments

In all experiments, the participant (Commander) conducts a collaborative search and navigation task with a robot teammate to assess the structure of a house-like environment and locate objects of interest.

5.1 Experiment Design and Method

Participants first answered a questionnaire to collect demographic information (see Appendix A for a list of all surveys). They were then seated at a computer monitor and fitted with a headset microphone and a keyboard configured with a push-to-talk button. Participants were also provided a list of the robot's capabilities (see Appendix B), shown a photo of the robot (either real or virtual, depending on the experiment configuration), and given a worksheet outlining the tasks and a pen to take notes. An experimenter walked through the heads-up interface used by participants to monitor the robot and its environment (see Fig. 1, upper right). Participants were not informed that the robot was controlled by wizard experimenters.

Participants then completed a training session where they would practice interacting with the robot in a simplified version of the task (specifically, this environment was a narrow corridor with a few rooms to explore with the robot). Once they were comfortable with the task, participants proceeded to the two main trials. These two trials took place at different starting locations in a house-like environment, with the order of starting points counterbalanced across participants. Each trial also had different tasks, such as tallying doorways, shovels, shoes, or whether the space was recently occupied. These trials lasted until participants self-reported that they finished the task, or 20 min had elapsed, whichever came first. After each trial, the participant reported on findings to the experimenter. In all the following experiments, a participant only participated in an experiment run once.

5.2 Experiment 1: Free Response Mode

5.2.1 Method Summary

Experiment 1 followed the general experiment design, with a Commander participant and two wizard experimenters (DM-Wizard and RN-Wizard). The DM-Wizard typed responses manually using a keyboard into chat windows to both the Commander (responses and clarifications) and RN-Wizard (tasks to complete), so we call this experiment Free Response Mode. Experimentation took place at Adelphi Laboratory Center (ALC).

5.2.2 Participation

Ten people participated in Experiment 1. There were 8 male and 2 female participants, and the mean age was 44 (min = 28, max = 58).

5.2.3 Key Findings

Experiment 1 sought to examine the feasibility of our setup: can we collect useful robot-directed dialogue data with human experimenter stand-ins for components that will ultimately be automated? While our setup did show that collecting useful dialogue, speech, and robot data was possible, there were limitations to the consistency of the responses generated by the DM-Wizard. First, responses were slow. This was because not only did the DM-Wizard have to think on-the-spot about how to respond, they also needed to type carefully so as not to introduce typographical errors, as mentioned in Section 4. Second, since every response was manually typed, there was naturally some divergence. This divergence needed to be manually analyzed to identify the core patterns in how a robot should respond in similar situations.

Upon completion of Experiment 1, we identified a need to automate some of the DM-Wizard labor to reduce latency at responding to a Commander's instructions, and also to reduce variation in the syntactic structure of responses. Semantically, many forms of a response may be appropriate, but without a detailed analysis, there are challenges in gaining value from all collected data. These responses are key to training a dialogue system: a dialogue system will need to know not only the kind of situations to expect, but also how to respond to those situations. Consistent responses are better for creating a smoother distribution of training data for an automated system.

The completion of Experiment 1 also provided an opportunity to analyze the form and content of instructions directed to the robot. We found that, in general, participants preferred to include metric content (e.g., move forward 3 ft) over references to the physical environment, such as landmarks (e.g., move to the box in front of you), in their instructions. However, we observed that participants gravitated to using more landmark-based instructions over time. This difference was statistically significant: when comparing instructions from what the first trial participants did to the second, they used more landmarks in their instructions in the later trial. This result suggests that as participants built experience with the robot and the natural language interface, they were comfortable giving it more "human-like" instructions as opposed to instructions that would be equivalent to verbal teleoperation of a robot. Additional discussion can be found in Marge et al.[25]

5.3 Experiment 2: Structured Response Mode (Real-World)

5.3.1 Method Summary

Experiment 2 followed the same experiment design as Experiment 1, with the exception of the DM-Wizard using a graphical interface to produce messages to both the Commander and RN-Wizard. We call this experiment Structured Response Mode (Real-World). Experimentation took place at ALC.

5.3.2 Participation

Ten people participated in Experiment 2. There were 5 male and 5 female participants, and the mean age was 42 (min = 18, max = 58).

5.3.3 Key Findings

The use of a graphical interface to automate some of the DM-Wizard's labor, while providing them the freedom to select and occasionally insert context-specific content into responses, was a substantial improvement in our experimental setup. We leveraged a process already in place with our collaboration partners at USC ICT that accomplished a similar feat with building a virtual human,[32] but added innovations specific to the challenges in human–robot interaction. Generally, these innovations relate to identifying content patterns in the instructions and responses as they relate to the robot's current task and surroundings. Details on this process can be found in Bonial et al.[26]

Following creation of an annotation scheme to measure the content of language instructions and dialogue exchanges, we compared the quality of data collected in Experiment 2 to Experiment 1. Not only did we find that the graphical interface significantly improved the pace of dialogue during trials, we also found the data itself improved the accuracy of an algorithm that derives intent from natural language instructions. Despite the limitations of a click-button interface compared to free-form natural language to produce responses to Commander instructions, the interface itself maintained good coverage of situations in the task domain. All results are reported in Marge et al.[24]

5.4 Experiment 3: Structured Response Mode (Virtual)

5.4.1 Method Summary

In contrast to previous experiments, Experiment 3 shifted the setup to a virtual rendering of the real-world environment. Like previous experiments, Experiment 3 also had a Commander participant, DM-Wizard, and RN-Wizard. The DM-Wizard

used a graphical interface similar to Experiment 2. The virtual environment was designed to replicate a 1–1 mapping of objects from the first two experiments and their locations. The task domain was equivalent to previous studies. We call this experiment Structured Response Mode (Virtual). Experimentation took place at ALC, Aberdeen Proving Ground, and ARL-West (located in Los Angeles, California).

5.4.2 Participation

In Experiment 3, 63 people participated. There were 23 male and 40 female participants, and the mean age was 42 (min = 18, max = 70).

5.4.3 Key Findings

Experiment 3 assessed whether the shift to a virtual setup would accelerate the pace of data collection, while still collecting useful training data. Indeed, this was a success: in a matter of three months, we collected an order of magnitude more data from 63 participants compared to previous experiments that saw only 10 participants take part per study. The shift to a virtual setup afforded us flexibility in where to conduct the study and reduced a burden on equipment and staff needs required to conduct the study—specifically, a robotics expert was no longer required in case of technical difficulties. Additionally, a shift to virtual setup provided the benefit of collecting data from multiple populations of participants: the local communities of northern Maryland, the Washington DC Metro area, and the greater Los Angeles Metro area. In contrast to previous studies, multiple experiments could be run per day, as there was no dependency on the availability of a physical space or functional robot.

5.5 Experiment 4: Automated Response Mode

5.5.1 Method Summary

The final study, Experiment 4, leveraged data collected in previous experiments to use a fully automated conversational interface that replaced the DM-Wizard role. This interface included real-time speech and dialogue processing (with push-to-talk for speech endpointing). An RN-Wizard was still included, as the focus of the research was evaluating the robustness of dialogue processing as opposed to introducing unanticipated robot navigation errors. The task domain was equivalent to previous studies. We call this experiment Automated Response Mode. Experimentation took place at ARL-West.

5.5.2 Participation

Ten people participated in Experiment 4. There were 7 male and 3 female participants, and the mean age was 29 (min = 21, max = 48).

5.5.3 Key Findings

Since Experiment 4 was the first opportunity to evaluate an automated system, it was unclear whether participants would be able to successfully complete the task due to unforeseen errors or poor coverage by the automated system. Similar to previous studies, participants were able to complete tasks with the robot. An analysis of the data collected showed the amount of Commander instructions was similar to the pace of collection in previous experiments (see Table 2). While occasionally there were inaccuracies in the system's behavior, participants were able to use established conversational strategies (e.g., issuing a cancel request) that would allow the interaction to continue without further disruption.

Table 2 Corpus statistics for SCOUT

Statistics	Exp 1	Exp 2	Exp 3	Exp 4	Total
Participants	10	10	63	10	93
Hours of audio	10	10	63	10	93
Transcripts	30	30	188	30	278
Participant utterances	3,267	1,425	13,218	3,151	21,061
Participant words	20,174	10,224	81,202	13,392	124,992
Images taken	835	565	3,694	691	5,785

6. Impact

This section summarizes the resulting research products of Bot Language experimentation and its impact on the Army and international community.

6.1 SCOUT Corpus

The collection of language and robot data from Experiments 1–4 comprise SCOUT, the Situated Corpus of Understanding Transactions. SCOUT contains participants' speech and all messages from the DM-Wizard (in Experiment 4, the automated Dialogue Manager) and RN-Wizard, and logs of all robot messages, maps, and navigation commands. The contents of SCOUT are summarized in Table 2.

The corpus consists of data from 93 participants, whose audio recordings total approximately 1 h per participant. Time-aligned transcriptions display all the interlocutors involved in the dialogue (i.e., the Commander participant, the DM-Wizard, and the RN-Wizard), in their respective communication floors. Table 3 shows a sample exchange. One transcript is created per session the participant

completed (i.e., training and main trials), for a total of 278 dialogues, containing 21,061 participant utterances and 124,992 participant words.

Table 3 Time-aligned dialogue transcript excerpt

0	Left floor		Right floor	
ID no.	Commander	DM→Commander	DM→RN	RN
1	Continue to move down the hallway	. . .	. . .	. . .
2	. . .	How far should I continue down the hallway? Would you like me to reach something that you see or go a certain distance?	. . .	. . .
3	Go down the end of the hallway where it appears to be a clock on the wall	. . .	. . .	. . .
4	. . .	. . .	Move to the end of the hallway where you are facing the clock on the wall.	. . .
5	. . .	Executing	. . .	. . .
6	Take a picture	. . .	. . .	. . .
7	. . .	. . .	Image	. . .
8	. . .	. . .	. . .	Done
9	. . .	Sent	. . .	. . .

Notes: DM: Dialogue Manager Wizard; RN: Robot Navigator Wizard.

SCOUT additionally includes several types of annotations regarding the form and content of the dialogues. For more information on these annotations, see Marge et al.[25] for instruction-level annotations and Traum et al.[33] for dialogue structure annotations. Still-frame pictures from every time the participant requested the robot to take an image of the environment have been extracted from the SCOUT robot data, yielding a collection of 5,785 images that are interleaved with and critical to the unfolding dialogue.

To date, ARL has distributed SCOUT to research partners at four universities—two international and two domestic—with one involved in ARL's AI and Autonomy for Multi-Agent Systems (ArtIAMAS) Cooperative Agreement, who are studying linguistic patterns and are investigating the feasibility of novel neural network algorithms for human–robot dialogue with the data. These collaborations have oriented members of the research community toward challenges of interest to the Army. To date, these collaborators have published papers on modal expressions[34] and neural dialogue algorithms.[35]

6.2 Prototype Systems

As the Bot Language experiments were piloted, launched, and the data analyzed, team members conducted software "sprints" aimed at demonstrable proof-of-concept systems that show the technical promise of real-time dialogue with autonomous systems. These systems are summarized in the final two columns of Table 1. The first prototype, ScoutBot, was created to determine if the data collected in the initial experiments could be used to train a dialogue system to support collaborative navigation in a task domain similar to Experiments 1–4.[27] ScoutBot was the first human–robot dialogue system trained entirely off data collected using the DWoZ methodology. ScoutBot permitted users to issue verbal navigation instructions to a virtual Clearpath Robotics Jackal in an indoor environment. Notable technical demonstrations of ScoutBot were given to the Chief of Staff of the Army, General James McConville, and to attendees at the 2018 Annual Meeting of the Association for Computational Linguistics (ACL), the premier conference for research in natural language processing and computational linguistics. General McConville noted the technical promise of human–robot dialogue: "You could tell one robot to go here, another there, and another with certain sensors to go there."

A second prototype, called MultiBot, aimed to extend ScoutBot capabilities to support dialogue interaction with multiple robotic platforms, and in a different task domain.[28] By combining dialogue with advanced robotic behaviors (e.g., Tactical Behavior Specifications[36]) originally created as part of the ARL Robotics Collaborative Technology Alliance, MultiBot could interpret goal-based instructions (scout route bravo) to an aerial-ground team based on each robot's capabilities in a search task. Notable technical demonstrations of MultiBot were given to ARL Director Dr Patrick Baker and to attendees at the 2019 Annual Meeting of the North American Chapter of the Association for Computational Linguistics (NAACL), a top-tier conference for research in natural language processing and computational linguistics. As a successor to ScoutBot, MultiBot demonstrated the generalizability of the Bot Language project's technical contributions to now enabling dialogue processing between one human and a team of mobile robots.

6.3 JUDI

In partnership with USC ICT, the prototype ScoutBot dialogue system was organized into an application to support conversational interaction with autonomous systems known as JUDI, the Joint Understanding and Dialogue Interface. JUDI combines spoken language interaction with the full capabilities of the ARL Autonomy Stack

that comprise navigation and perception algorithms for robotic platforms. JUDI leverages elements originally created in USC ICT's Virtual Human Toolkit[37] for spoken language and dialogue processing algorithms. JUDI also features tight integration with offline speech recognition provided by the Kaldi open-source speech recognition toolkit.[38] In contrast to many of today's conversational systems, JUDI does not require a cloud connection to function.

As a primary mode of Soldier–agent interaction for agents that use the ARL Autonomy Stack, JUDI provides a "heads-up, hands-free" mode of communication that allows Soldiers to keep their heads up and hands free for other tasks. Experiments with fielding JUDI are underway as part of the AIMM ERP.

6.4 Dialogue-AMR

Although the dialogue structure annotations of SCOUT provide information critical to dialogue systems on which different utterances are related and what relations exist between them, the annotations do not provide a markup of the semantic content of participant instructions. Because we hypothesize that such a semantic representation would be valuable in mapping the natural language instructions to the autonomous system's set of executable behaviors and to the real-world objects mentioned in those behaviors (i.e., symbol grounding), we explored how to develop a semantic representation suitable for human–robot dialogue.

We first evaluated the suitability of existing semantic representations, and opted to explore the strengths and weaknesses of Abstract Meaning Representation (AMR) to support both natural language understanding and symbol grounding in human–robot dialogue.[39] AMR is a formalism for sentence semantics that abstracts away from some syntactic idiosyncrasies.[40] 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., `(r / robot)`). AMR provides an appropriate level of abstraction for natural language understanding in our human–robot dialogue application. As the goal of AMR research is to capture core facets of meaning unrelated to surface structure, the same underlying concept realized alternatively as a noun (a left turn), verb (turn to the left), or light verb construction (make a left turn) are all represented by identical AMRs. This is well-suited to our setup: the agent (e.g., a robot) has a limited number of executable behaviors it can perform, and any user utterance needs to be mapped to a simple yet structured representation that the agent can understand. In turn, the agent only needs to communicate back to the user regarding those same concepts. Thus, the AMR formalism smooths away many syntactic and lexical features that are unimportant to the agent. Existing AMR parsers can be used to obtain an initial interpretation of

a user utterance, making the interpretation process easier than parsing natural language text directly into an agent-oriented representation.

However, we also found that AMR has several gaps in information crucial to human–robot dialogue: 1) the *illocutionary force* of the speaker (i.e., what the speaker is trying to do with their utterance in a conversational context, such as give a command or ask a question); 2) *tense* or the time of instructed actions (i.e., will it happen in the future or has it already happened); and 3) *aspect* or the completion status of instructed actions (i.e., is it complete, ongoing, or not started yet). The formalism that we developed, Dialogue-AMR, addresses each of these gaps in a novel way that captures these desired three features along with the semantic, propositional content of instructions, within one computer-readable representation.[41]

We have since shown that leveraging both AMR and Dialogue-AMR in an intelligence architecture for agents can reduce the amount of training data needed and increase computation time and efficiency for grounding the natural language instructions (i.e., deriving correspondences between the objects mentioned and the robot's sensory perceptions, and selecting action primitives and parameters for the behaviors) (see Howard et al.[42] for the grounding approach). We also demonstrated that we can extend Dialogue-AMR, as well as the automatic pipeline involving an AMR parser and a conversion system that outputs Dialogue-AMR, to novel domains, such as the blocks-world building domain of Minecraft, with very little training data from the new domain while maintaining high accuracy in the representation.[43] Encouraged by these promising results, ongoing work includes continuing to explore how to leverage AMR and Dialogue-AMR for situated human–robot dialogue for different collaborative tasks as well as interactions with multiple humans and agents.

7. Metrics

As of the publication of this report, the Bot Language project has produced 16 refereed conference and symposium publications, 12 refereed workshop publications, 2 technical reports, and 3 published abstracts.

7.1 Refereed Conference and Symposium Publications

1) Bonial C, Abrams M, Traum D, Voss C. Builder, we have done it: evaluating & extending Dialogue-AMR NLU pipeline for two collaborative domains. In: Proceedings of the 14th International Conference on Computational Semantics (IWCS); Association for Computational Linguistics; 2021. p. 173–183.

2) Bonial C, Donatelli L, Abrams M, Lukin SM, Tratz S, Marge M, Artstein R, Traum D, Voss C. Dialogue-AMR: Abstract Meaning Representation for dialogue. In: Proceedings of the 12th International Conference on Language Resources and Evaluation; European Language Resources Association; 2020. p. 684–695.

3) Hayes C, Marge M. Towards preference learning for autonomous ground robot navigation tasks. In: Proceedings of AI-HRI; Artificial Intelligence-Human–Robot Interaction Symposium; 2020.

4) Abrams M, Bonial C, Donatelli L. Graph-to-graph meaning representation transformations for human–robot dialogue. In: Proceedings of the Society for Computation in Linguistics; Society for Computation in Linguistics; 2020. p. 250–253.

5) Bonial C, Donatelli L, Ervin J, Voss CR. Abstract Meaning Representation for human–robot dialogue. In: Proceedings of the Society for Computation in Linguistics; Society for Computation in Linguistics; 2019. p. 236–246.

6) Marge M, Bonial C, Lukin S, Hayes C, Foots A, Artstein R, Henry C, Pollard K, Gordon C, Gervits F, Leuski A, Hill SG, Voss CR, Traum D. Balancing efficiency and coverage in human–robot dialogue collection. In: Proceedings of AI-HRI; Artificial Intelligence-Human–Robot Interaction Symposium; 2018.

7) Pollard KA, Lukin SM, Marge M, Foots A, Hill SG. How we talk with robots: eliciting minimally-constrained speech to build natural language interfaces and capabilities. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting; Vol. 62; Human Factors and Ergonomics Society; 2018. p. 160–164.

8) Lukin SM, Gervits F, Hayes CJ, Leuski A, Moolchandani P, Rogers JG III, Amaro CS, Marge M, Voss CR, Traum D. ScoutBot: a dialogue system for collaborative navigation. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics–System Demonstrations; Association for Computational Linguistics; 2018. p. 93–98.

9) Lukin S, Pollard K, Bonial C, Marge M, Henry C, Artstein R, Traum D, Voss C. Consequences and factors of stylistic differences in human–robot dialogue. In: Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue; Association for Computational Linguistics; 2018. p. 110–118.

10) Traum D, Henry C, Lukin S, Artstein R, Gervits F, Pollard K, Bonial C, Lei S, Voss C, Marge M, Hayes C, Hill S. Dialogue structure annotation for multifloor interaction. In: Calzolari N, Choukri K, Cieri C, Declerck T, Goggi S, Hasida K, Isahara H, Maegaard B, Mariani J, Mazo H, Moreno A, Odijk J, Piperidis S, Tokunaga T, editors. Proceedings of the Eleventh International Conference on Language Resources and Evaluation; European Language Resources Association; 2018. p. 104–111.

11) Henry C, Lukin S, Pollard KA, Bonial C, Foots A, Artstein R, Voss CR, Traum D, Marge M, Hayes CJ, Hill SG. The Bot Language project: moving towards natural dialogue with robots. In: Proceedings of the SoCalNLP Symposium; SoCal NLP; 2018.

12) Moolchandani P, Hayes CJ, Marge M. Evaluating robot behavior in response to natural language. In: Companion of the 2018 ACM/IEEE International Conference on Human–Robot Interaction; Association for Computing Machinery; 2018. p. 197–198.

13) Bonial C, Marge M, Foots A, Gervits F, Hayes CJ, Henry C, Hill SG, Leuski A, Lukin SM, Moolchandani P, Pollard KA, Traum D, Voss CR. Laying down the yellow brick road: development of a Wizard-of-Oz interface for collecting human–robot dialogue. In: Proceedings of the AAAI Fall Symposium on Natural Communication for Human–Robot Collaboration; arXiv; 2017.

14) Marge M, Bonial C, Pollard KA, Artstein R, Byrne B, Hill SG, Voss C, Traum D. Assessing agreement in human–robot dialogue strategies: a tale of two wizards. In: Traum D, Swartout W, Khooshabeh P, Kopp S, Scherer S, Leuski A, editors. Intelligent Virtual Agents; Springer International Publishing; 2016. p. 484–488.

15) Marge M, Bonial C, Byrne B, Cassidy T, Evans AW, Hill SG, Voss C. Applying the Wizard-of-Oz technique to multimodal human–robot dialogue. In: Proceedings of IEEE RO-MAN; arXiv; 2016.

16) Cassidy T, Voss C, Summers-Stay D. Turn-taking in commander–robot navigator dialog (Video Abstract). In: Proceedings of the AAAI Spring Symposium Series; Association for the Advancement of Artificial Intelligence; 2015.

7.2 Refereed Workshop Publications

1) Bonial C, Abrams M, Baker AL, Hudson T, Lukin SM, Traum D, Voss CR. Context is key: annotating situated dialogue relations in multi-floor dialogue. In: Proceedings of the 25th Workshop on the Semantics and Pragmatics of Dialogue; SEMDIAL; 2021.

2) Gervits F, Leuski A, Bonial C, Gordon C, Traum D. A classification-based approach to automating human–robot dialogue. In: Marchi E, Siniscalchi SM, Cumani S, Salerno VM, Li H, editors. Increasing Naturalness and Flexibility in Spoken Dialogue Interaction: 10th International Workshop on Spoken Dialogue Systems; Springer Singapore; 2021. p. 115–127.

3) Bonial C, Donatelli L, Lukin SM, Tratz S, Artstein R, Traum D, Voss C. Augmenting Abstract Meaning Representation for human–robot dialogue. In: Proceedings of the First International Workshop on Designing Meaning Representations; Association for Computational Linguistics; 2019. p. 199–210.

4) Lukin SM, Bonial C, Voss CR. Visual understanding and narration: a deeper understanding and explanation of visual scenes. In: Proceedings of the Workshop on Shortcomings in Vision and Language (SiVL); arXiv; 2019.

5) Lukin S, Hobbs R, Voss C. A pipeline for creative visual storytelling. In: Proceedings of the First Workshop on Storytelling (StoryNLP); Association for Computational Linguistics; 2018. p. 20–32.

6) Henry C, Gordon C, Traum D, Lukin SM, Pollard KA, Artstein R, Bonial C, Voss CR, Foots A, Marge M. Faster pace in human–robot dialogue leads to fewer dialogue overlaps. In: Proceedings of the NAACL Workshop on Widening NLP; Association for Computational Linguistics; 2018.

7) Hayes CJ, Marge M, Stump E, Bonial C, Voss C, Hill SG. Towards learning user preferences for remote robot navigation. In: Proceedings of the RSS 2018 Workshop on Models and Representations for Human–Robot Communication; University of Rochester; 2018.

8) Bonial C, Lukin SM, Foots A, Henry C, Marge M, Pollard KA, Artstein R, Traum D, Voss CR. Human–robot dialogue and collaboration in search and navigation. In: Proceedings of the Annotation, Recognition and Evaluation of Actions (AREA) Workshop of the 2018 Language Resources and Evaluation Conference (LREC); European Language Resources Association; 2018.

9) Marge M, Bonial C, Foots A, Hayes C, Henry C, Pollard K, Artstein R, Voss C, Traum D. 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; Association for Computational Linguistics; 2017. p. 58–66.

10) Henry C, Moolchandani P, Pollard KA, Bonial C, Foots A, Artstein R, Hayes C, Voss CR, Traum D, Marge M. Towards efficient human–robot dialogue collection: moving Fido into the virtual world. In: Proceedings of the Workshop on Women and Underrepresented Minorities in Natural Language Processing (WiNLP); Association for Computational Linguistics; 2017.

11) Summers-Stay D, Cassidy T, Voss C. Joint navigation in commander/robot teams: dialog & task performance when vision is bandwidth-limited. In: Proceedings of the Third Workshop on Vision and Language; Dublin City University and the Association for Computational Linguistics; 2014. p. 9–16.

12) Voss C, Cassidy T, Summers-Stay D. Collaborative exploration in human–robot teams: What's in their corpora of dialog, video, & LIDAR messages? In: Proceedings of the EACL 2014 Workshop on Dialogue in Motion; Association for Computational Linguistics; 2014. p. 43–47.

7.3 Technical Reports

1) Bonial C, Traum D, Henry C, Lukin SM, Marge M, Artstein R, Pollard KA, Foots A, Baker AL, Voss CR. Dialogue structure annotation guidelines for Army Research Laboratory (ARL) human–robot dialogue corpus. DEVCOM Army Research Laboratory (US); 2019. Report No.: ARL-TR-8833.

2) Bonial C, Henry C, Artstein R, Marge M. Transcription guidelines for Army Research Laboratory (ARL) human–robot dialogue corpus. DEVCOM Army Research Laboratory (US); 2019. Report No.: ARL-TR-8832.

7.4 Published Abstracts

1) Hayes C, Marge M, Bonial C, Voss C, Hill SG. Team-centric motion planning in unfamiliar environments (Conference Presentation). In: Degraded Environments: Sensing, Processing, and Display; Vol. 10642; SPIE; 2018.

2) Marge M, Traum D, Voss CR, Hill SG. Towards natural dialogue with robots: ARL Bot Language. In: Proceedings of the NDIA Human Systems Conference; National Defense Industrial Association; 2018.

3) Marge M, Bonial C, Pollard KA, Henry C, Artstein R, Byrne B, Hill SG, Voss C, Traum D. Towards natural dialogue with robots: Bot Language. In: Proceedings of AI-HRI; Association for the Advancement of Artificial Intelligence; 2016.

7.5 Seminars and Presentations

The project resulted in 17 invited department-level seminars showcasing natural language human–robot interaction research:

1) Matthew Marge: University of Maryland, Georgetown University, University of Rochester, Naval Research Laboratory, Defence Science Technology Group (Australia), Tufts University, Bose Research, University of Gothenburg (Sweden), and Air Force Research Laboratory

2) Claire Bonial: Georgetown University, University of Maryland, University of Colorado Boulder, and Brown University

3) Stephanie Lukin: Loyola University Maryland and Disney Imagineering Research

4) Clare Voss: USC ICT and University of Illinois Urbana-Champaign

Finally, the project resulted in 28 presentations at international conference and workshop venues, and 3 keynote presentations at conferences:

1) Matthew Marge: 2019 AI-HRI Symposium

2) Claire Bonial, Stephanie Lukin, and Clare Voss: 2020 SemDial Workshop on the Semantics and Pragmatics of Dialogue

3) Stephanie Lukin: 2020 AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment

8. Conclusions

This report summarizes technical accomplishments from the Bot Language project, which aimed to provide more natural ways for people to communicate and interact with remotely located robots using natural language dialogue. The primary contribution of this body of work was a series of experiments that elicited dialogue-based communications between humans and robots and the associated data that

form a data set called SCOUT. In the early stages of this work, the natural language components were controlled by human experimenters using the WoZ methodology but were progressively automated in later experiments. A dialogue system prototype developed in this work formed the basis for the Army's JUDI capability that enables conversational interactions between Soldiers and autonomous systems. Another major contribution has been to the natural language processing community in the form of annotation schemes for modeling the structure, content, and semantics of dialogue-based exchanges from SCOUT. Today, these annotations are used to train machine learning algorithms to understand patterns in robot-directed natural language, improving how robots can make decisions when teaming with humans.

9. References

1. DeVault D, Artstein R, Benn G, Dey T, Fast E, Gainer A, Georgila K, Gratch J, Hartholt A, Lor-Lhommet M, et al. SimSensei Kiosk: a virtual human interviewer for healthcare decision support. In: Proceedings of AAMAS; 13th International Conference on Autonomous Agents and Multiagent Systems; 2014 Jan. p. 1061–1068.

2. Mavridis N. A review of verbal and non-verbal human–robot interactive communication. Rob Auton Syst. 2015;63:22–35.

3. Tellex S, Gopalan N, Kress-Gazit H, Matuszek C. Robots that use language. Annu Rev Control Robot Auton Syst. 2020;3:25–55.

4. Bugmann G, Klein E, Lauria S, Kyriacou T. Corpus-based robotics: a route instruction example. In: Proceedings of IAS-8; Intelligent Autonomous Systems; 2004 Mar 10–13; Amsterdam, the Netherlands. p. 96–103.

5. Kruijff GJM, Lison P, Benjamin T, Jacobsson H, Zender H, Kruijff-Korbayová I, Hawes N. Situated dialogue processing for human–robot interaction. In: Christensen HI, Kruijff GJM, Wyatt JL, editors. Cognitive Systems Monographs; Vol. 8; Springer; c2010. p. 311–364.

6. Williams T, Briggs G, Oosterveld B, Scheutz M. Going beyond literal command-based instructions: Extending robotic natural language interaction capabilities. In: Proceedings of AAAI; Association for the Advancement of Artificial Intelligence; 2015.

7. Tellex SA, Kollar TF, Dickerson SR, Walter MR, Banerjee A, Teller S, Roy N. Understanding natural language commands for robotic navigation and mobile manipulation. In: Proceedings of AAAI; Association for the Advancement of Artificial Intelligence; 2011.

8. Hemachandra S, Duvallet F, Howard TM, Roy N, Stentz A, Walter MR. Learning models for following natural language directions in unknown environments. In: Proceedings of ICRA; IEEE Robotics and Automation Society; 2015.

9. Foster ME, Giuliani M, Isard A, Matheson C, Oberlander J, Knoll A. Evaluating description and reference strategies in a cooperative human–robot dialogue system. In: Proceedings of IJCAI; International Joint Conferences on Artificial Intelligence; 2009.

10. Bohus D, Saw CW, Horvitz E. Directions robot: in-the-wild experiences and lessons learned. In: Proceedings of AAMAS; International Foundation for Autonomous Agents and Multiagent Systems; 2014.

11. Perera V, Selveraj SP, Rosenthal S, Veloso M. Dynamic generation and refinement of robot verbalization. In: Proceedings of IEEE RO-MAN; IEEE Robotics and Automation Society; 2016.

12. Deits R, Tellex S, Thaker P, Simeonov D, Kollar T, Roy N. Clarifying commands with information-theoretic human–robot dialog. J Hum–Robot Interact. 2013;2(2):58–79.

13. Knepper RA, Tellex S, Li A, Roy N, Rus D. Recovering from failure by asking for help. Auton Robots. 2015;39(3):347–362.

14. Scheutz M, Williams T, Krause E, Oosterveld B, Sarathy V, Frasca T. An overview of the distributed integrated cognition affect and reflection DIARC architecture. Cogn Archit. 2019;94:165–193.

15. Marge M, Rudnicky AI. Miscommunication detection and recovery in situated human–robot dialogue. ACM Transactions on Interactive Intelligent Systems (TiiS). 2019;9(1):1–40.

16. Fraser NM, Gilbert GN. Simulating speech systems. Comput Speech Lang. 1991;5(1):81–99.

17. Riek L. Wizard of Oz studies in HRI: a systematic review and new reporting guidelines. J Hum–Robot Interact. 2012;1(1).

18. Passonneau RJ, Epstein SL, Ligorio T, Gordon J. Embedded wizardry. In: Proceedings of SIGdial; Association for Computational Linguistics; 2011.

19. Brooks D, Lignos C, Finucane C, Medvedev M, Perera I, Raman V, KressGazit H, Marcus M, Yanco H. Make it so: continuous, flexible natural language interaction with an autonomous robot. In: Proceedings of the Grounding Language for Physical Systems Workshop at the AAAI Conference on Artificial Intelligence; Association for the Advancement of Artificial Intelligence; 2012.

20. Marcus M, Kress-Gazit H, Yanco H, Brooks D, Lignos C, Finucane C, Lee K, Medvedev M, Perera I, Raman V. SUBTLE: situation understanding bot through language and environment. Army Research Office (US); 2016. Report No.: 52555-MA-MUR.19.

21. Voss C, Cassidy T, Summers-Stay D. Collaborative exploration in human–robot teams: What's in their corpora of dialog, video, & LIDAR messages? In: Proceedings of the EACL 2014 Workshop on Dialogue in Motion; Association for Computational Linguistics; 2014. p. 43–47.

22. Summers-Stay D, Cassidy T, Voss C. Joint navigation in commander/robot teams: dialog & task performance when vision is bandwidth-limited. In: Proceedings of the Third Workshop on Vision and Language; Dublin City University and the Association for Computational Linguistics; 2014. p. 9–16.

23. Cassidy T, Voss C, Summers-Stay D. Turn-taking in commander-robot navigator dialog (Video Abstract). In: Proceedings of the AAAI Spring Symposium Series; Association for the Advancement of Artificial Intelligence; 2015.

24. Marge M, Bonial C, Lukin S, Hayes C, Foots A, Artstein R, Henry C, Pollard K, Gordon C, Gervits F, Leuski A, Hill SG, Voss CR, Traum D. Balancing efficiency and coverage in human–robot dialogue collection. In: Proceedings of AI-HRI; Artificial Intelligence-Human–Robot Interaction Symposium; arXiv; 2018.

25. Marge M, Bonial C, Foots A, Hayes C, Henry C, Pollard K, Artstein R, Voss C, Traum D. 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; Association for Computational Linguistics; 2017. p. 58–66.

26. Bonial C, Marge M, Foots A, Gervits F, Hayes CJ, Henry C, Hill SG, Leuski A, Lukin SM, Moolchandani P, Pollard KA, Traum D, Voss CR. Laying down the yellow brick road: development of a Wizard-of-Oz interface for collecting human-robot dialogue. In: Proceedings of the AAAI Fall Symposium on Natural Communication for Human–Robot Collaboration; arXiv; 2017.

27. Lukin SM, Gervits F, Hayes CJ, Leuski A, Moolchandani P, Rogers JG III, Amaro CS, Marge M, Voss CR, Traum D. ScoutBot: a dialogue system for collaborative navigation. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics–System Demonstrations; Association for Computational Linguistics; 2018. p. 93–98.

28. Marge M, Nogar S, Hayes CJ, Lukin SM, Bloecker J, Holder E, Voss C. A research platform for multi-robot dialogue with humans. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations); Association for Computational Linguistics; 2019. p. 132–137.

29. Marge M, Rudnicky A. Comparing spoken language route instructions for robots across environment representations. In: Proceedings of SIGdial; Association for Computational Linguistics; 2010. p. 157–164.

30. Koenig N, Howard A. Design and use paradigms for Gazebo, an open-source multi-robot simulator. In: Proceedings of IROS; IEEE Robotics and Automation Society; 2004. p. 2149–2154.

31. Gervits F, Leuski A, Bonial C, Gordon C, Traum D. A classification-based approach to automating human–robot dialogue. In: Marchi E, Siniscalchi SM, Cumani S, Salerno VM, Li H, editors. Increasing Naturalness and Flexibility in Spoken Dialogue Interaction: 10th International Workshop on Spoken Dialogue Systems; Springer Singapore; 2021. p. 115–127.

32. Artstein R, Leuski A, Maio H, Mor-Barak T, Gordon C, Traum DR. How many utterances are needed to support time-offset interaction? In: Proceedings of FLAIRS; Florida Online Journals; 2015.

33. Traum D, Henry C, Lukin S, Artstein R, Gervits F, Pollard K, Bonial C, Lei S, Voss C, Marge M, Hayes C, Hill S. Dialogue structure annotation for multi-floor interaction. In: Calzolari N, Choukri K, Cieri C, Declerck T, Goggi S, Hasida K, Isahara H, Maegaard B, Mariani J, Mazo H, Moreno A, Odijk J, Piperidis S, Tokunaga T, editors. Proceedings of the Eleventh International Conference on Language Resources and Evaluation; European Language Resources Association; 2018. p. 104–111.

34. Donatelli L, Lai K, Pustejovsky J. A two-level interpretation of modality in human–robot dialogue. In: Proceedings of COLING; International Committee on Computational Linguistics; 2020. p. 4222–4238.

35. Kawano S, Yoshino K, Traum D, Nakamura S. Dialogue structure parsing on multi-floor dialogue based on multi-task learning. In: Proceedings of the 1st RobotDial Workshop on Dialogue Models for Human–Robot Interaction; International Joint Conferences on Artificial Intelligence; 2021. p. 21–29.

36. Boularias A, Duvallet F, Oh J, Stentz A. Grounding spatial relations for outdoor robot navigation. In: Proceedings of ICRA; IEEE Robotics and Automation Society; 2015. p. 1976–1982.

37. Gratch J, Hartholt A, Dehghani M, Marsella S. Virtual humans: a new toolkit for cognitive science research. In: Proceedings of the Annual Meeting of the Cognitive Science Society; Cognitive Science Society; 2013.

38. Povey D, Ghoshal A, Boulianne G, Burget L, Glembek O, Goel N, Hannemann M, Motlicek P, Qian Y, Schwarz P, Silovsky J, Stemmer G, Vesely K. The Kaldi speech recognition toolkit. In: IEEE Workshop on Automatic Speech Recognition and Understanding; IEEE Signal Processing Society; 2011.

39. Bonial C, Donatelli L, Ervin J, Voss CR. Abstract Meaning Representation for human–robot dialogue. In: Proceedings of the Society for Computation in Linguistics; Society for Computation in Linguistics; 2019. p. 236–246.

40. Banarescu L, Bonial C, Cai S, Georgescu M, Griffitt K, Hermjakob U, Knight K, Koehn P, Palmer M, Schneider N. Abstract Meaning Representation for sembanking. In: Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse; Association for Computational Linguistics; 2013. p. 178–186.

41. Bonial C, Donatelli L, Abrams M, Lukin SM, Tratz S, Marge M, Artstein R, Traum D, Voss C. Dialogue-AMR: Abstract Meaning Representation for dialogue. In: Proceedings of the 12th International Conference on Language Resources and Evaluation; European Language Resources Association; 2020. p. 684–695.

42. Howard TM, Stump E, Fink J, Arkin J, Paul R, Park D, Roy S, Barber D, Bendell R, Schmeckpeper K, et al. An intelligence architecture for grounded language communication with field robots. In: Field Robotics; Field Robotics Publication Society; 2021.

43. Bonial C, Abrams M, Traum D, Voss C. Builder, we have done it: evaluating & extending dialogue-AMR NLU pipeline for two collaborative domains. In: Proceedings of the 14th International Conference on Computational Semantics (IWCS); Association for Computational Linguistics; 2021. p. 173–183.

Appendix A. Survey Questions

The surveys and questions that participants completed include the following:

Demographic Surveys:

- Age (free text box)
- Gender (Female, Male options only)
- Wear glasses or contacts
- Experience in military service
- Experience in video gaming
- Experience in working with automated voice systems (e.g., SIRI)
- Experience in working with robots and automation
- Color vision test (screening test; we required participants to pass this six-question test)
- Spatial orientation test
- Mini-IPIP (International Personality Item Pool)

Surveys taken before and after trials:

- Trust Perception Scale —HRI (human–robot interaction)
- NASA-TLX (Task Load Index)
- Robot Perception Survey — Robot Dominance/Humanlikeness/Knowledge scales

Appendix B. Robot Capabilities

These are, verbatim, the capabilities provided on a sheet to study participants:

The robot can take a photo of what it sees when you ask. The robot has certain capabilities but cannot perform these tasks on its own. The robot and you will act as a team.

Robot capabilities are:

- Robot listens to verbal instructions from you
- Robot responds in this text box (Experimenter points to instant messenger box on screen) or by taking action
- Robot will avoid obstacles
- Robot can take photos directly in front of it when you give it a verbal instruction
- Robot will know what some objects are, but not all objects
- Robot also knows:
    - Intrinsic properties like color and size of objects in the environment
    - Proximity of objects like where objects are relative to itself and to other objects
    - A range of spatial terms like to the right of, in front of, cardinal directions like N, S
- History: the Robot remembers places it has been
- Robot doesn't have arms and it cannot manipulate objects or interact with its environment except for moving throughout the environment
- Robot cannot go through closed doors and it cannot open doors, but it can go through doorways that are already open
- Robot can only see about knee height (~1.5 ft)

List of Symbols, Abbreviations, and Acronyms

ACL	Association for Computational Linguistics
AI	artificial intelligence
AIMM	Artificial Intelligence for Maneuver and Mobility
ALC	Adelphi Laboratory Center
AMR	Abstract Meaning Representation
ARL	Army Research Laboratory
ARO	Army Research Office
ArtIAMAS	AI and Autonomy for Multi-Agent Systems
ASR	Automatic Speech Recognition
CAPS	all uppercase
DEVCOM	US Army Combat Capabilities Development Command
DM	Dialogue Management
DM-Wizard	Dialogue Manager Wizard
DWoZ	Data-driven "Wizard-of-Oz"
ERP	Essential Research Program
HRI	human–robot interaction
JUDI	Joint Understanding and Dialogue Interface
LIDAR	light detection and ranging
MURI	Multidisciplinary University Research Initiative
NAACL	North American Chapter of the Association for Computational Linguistics
NASA	National Aeronautics and Space Administration
PI	Principal Investigator
RN-Wizard	Robot Navigator Wizard
SCOUT	Situated Corpus of Understanding Transactions

SUBTLE	Situation Understanding Bot through Language and Environment
TBS	Tactical Behavior Specification
USC ICT	University of Southern California's Institute for Creative Technologies
WoZ	Wizard of Oz

1 (PDF) DEFENSE TECHNICAL INFORMATION CTR
DTIC OCA

1 (PDF) DEVCOM ARL
FCDD RLB CI
TECH LIB

1 (PDF) DA HQ
DASA(R&T)

8 (PDF) USARMY AFC
L BROUSSEAU
A LINZ
K WADE
S BRADY
J REGO
T KELLY
E JOSEPH
B SESSLER

2 (PDF) DEVCOM HQ
FCDD ST
C SAMMS
M HUBBARD

87 (PDF) DEVCOM ARL
FCDD RLA
C BEDELL
B SADLER
A SWAMI
J ALEXANDER
M GOVONI
M WRABACK
H EVERITT
S KARNA
JF NEWILL
AM RAWLETT
SE SCHOENFELD
J CHEN
PJ FRANASZCZUK
C OLIVER
G BRILL
C TEETER
B PERELMAN
A SCHOFIELD
J LANE
FCDD RLA B
A WEST
FCDD RLA CA
J FOSSACECA
FCDD RLA CB
P GILLICH
FCDD RLA CC
C KRONINGER
FCDD RLA CD
J ROBINETTE
FCDD RLA CL
F FRESCONI
FCDD RLA CG
D STRATIS-CULLUM
FCDD RLA CV
M KWEON
FCDD RLA F
JR GASTON
FCDD RLA G
ML REED
FCDD RLA H
JJ SUMNER
FCDD RLA I
D WIEGMANN
FCDD RLA J
B PIEKARSKI
FCDD RLA L
RD DEL ROSARIO
FCDD RLA M
ES CHIN
FCDD RLA N
BM RIVERA
FCDD RLA P
WL BENARD
FCDD RLA PD
F FATEMI
FCDD RLA T
RZ FRANCART
FCDD RLA V
S SILTON
FCDD RLA W
TV SHEPPARD
FCDD RLB
J ZABINSKI
FCDD RLB D
JS ADAMS
FCDD RLB PE
B ASHFORD
FCDD RLB R
S STRANK
FCDD RLB RN
C QUIGLEY
FCDD RLB RW
P KHOOSHABEH
FCDD RLB S
K KAPPRA
J VETTEL
FCDD RLC IT
M MARGE
C BONIAL
S LUKIN
C VOSS
FCDD RLD
PJ BAKER

A KOTT
K JACOBS
FCDD RLD A
A MOGRO
FCDD RLD C
T KINES
FCDD RLD D
T ROSENBERGER
G LARKIN
R ZACHERY
FCDD RLD I
A FINCH
A LLOPIS-JEPSEN
D ROLL
S SHIDFAR
D KELLEY
FCDD RLD M
N ZANDER
R MURRAY
FCDD RLR
B HALPERN
S LEE
P REYNOLDS
K RASMUSSEN
FCDD RLR A
D STEPP
FCDD RLR C
LL TROYER
FCDD RLR DS
R FREED
FCDD RLR E
RA MANTZ
FCDD RLR EF
F GREGORY
FCDD RLR EG
H DE LONG
FCDD RLR EH
V MARTINDALE
FCDD RLR EI
SP IYER
FCDD RLR EL
J QIU
FCDD RLR EM
C VARANASI
FCDD RLR EN
JM COYLE
FCDD RLR EP
PM BAKER
FCDD RLR ET
B LOVE
FCDD RLR EV
J BARZYK
FCDD RLR EW
JD MYERS
FCDD RLR G
A SCRUGGS