# REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188



| 1. REPORT DATE (DD-MM-YYYY) | 2. REPORT TYPE | 3. DATES COVERED (From - To) |
|---|---|---|
| 02-04-2021 | Final Report | 31-Jul-2015 - 31-Dec-2020 |

**4. TITLE AND SUBTITLE**
Final Report: Cognitively Coherent Human-Computer Communication

**5a. CONTRACT NUMBER**
W911NF-15-1-0461

**5b. GRANT NUMBER**

**5c. PROGRAM ELEMENT NUMBER**

**6. AUTHORS**

**5d. PROJECT NUMBER**

**5e. TASK NUMBER**

**5f. WORK UNIT NUMBER**

**7. PERFORMING ORGANIZATION NAMES AND ADDRESSES**
University of Illinois - Urbana - Champaign
c/o Office of Sponsored Programs
1901 S. First Street, Suite A
Champaign, IL 61820 -7406

**8. PERFORMING ORGANIZATION REPORT NUMBER**

**9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)**
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

**10. SPONSOR/MONITOR'S ACRONYM(S)**
ARO

**11. SPONSOR/MONITOR'S REPORT NUMBER(S)**
67832-CS-DRP.39

**12. DISTRIBUTION AVAILIBILITY STATEMENT**
Approved for public release; distribution is unlimited.

**13. SUPPLEMENTARY NOTES**
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

**14. ABSTRACT**

**15. SUBJECT TERMS**

| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 15. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| a. REPORT | b. ABSTRACT | c. THIS PAGE | | | Julia Hockenmaier |
| UU | UU | UU | UU | | 19b. TELEPHONE NUMBER 000-000-0000 |

# RPPR Final Report
as of 07-Apr-2021

Agency Code: 21XD

Proposal Number: 67832CSDRP **Agreement Number: W911NF-15-1-0461**
**INVESTIGATOR(S):**


**Name:** Julia Hockenmaier
**Email:** juliahmr@illinois.edu
**Phone Number:** 0000000000
**Principal:** Y

Organization: **University of Illinois - Urbana - Champaign**
Address: c/o Office of Sponsored Programs, Champaign, IL 618207406
Country: USA
DUNS Number: 041544081 EIN: 376000511


**Report Date:** 31-Jan-2021 Date Received: 02-Apr-2021
**Final Report** for Period Beginning 31-Jul-2015 and Ending 31-Dec-2020
**Title:** Cognitively Coherent Human-Computer Communication
**Begin Performance Period:** 31-Jul-2015 **End Performance Period:** 31-Dec-2020
**Report Term:** 0-Other

Submitted By: Julia Hockenmaier Email: juliahmr@illinois.edu
Phone: (000) 000-0000

**Distribution Statement:** 1-Approved for public release; distribution is unlimited.

**STEM Degrees:** 3 **STEM Participants:** 6

**Major Goals:** The objective of DARPA's CwC program is to facilitate and accelerate progress toward symmetric communication between humans and computers. In support of this goal, we propose to study and develop new capabilities for machines to learn to understand human language in context, and to interact with humans in ways that account for the domain, the task at hand and the knowledge shared between the machine and the human. The goal of our multi-university team is to combine Natural Language Processing (NLP), Machine Learning (ML), and Knowledge Representation and Reasoning (KRR) techniques into new technologies that support the CwC mission. Our project "Cognitively Coherent Human-Computer Communication" will explore ways for an interactive intelligent system to achieve cognitive coherency, concentrating on robust communication with humans through incrementally adapting its understanding of the human's language, its communication abilities and, consequently, its ability to support domain specific reasoning. Our key innovations are: (1) A new knowledge representation that builds on and extends Dowty's prototype view in multiple ways to cover a broad range of semantics bearing predicates, and to represent deeper context including topical and grounding information; (2) A flexible computational framework to support joint inference and learning that is representation independent, and is capable of leveraging mutually constraining information provided by different knowledge sources; (3) An interactive, response based learning protocol to understand human language while interacting in the context of a task; the protocol is supported by contextual distributional learning, domain grounding, and advice taking, and significantly extends psycholinguistics ideas and current advanced machine learning methodologies; and (4) Incorporated within the learning protocol is a framework for learning the domain model, that facilitates generating useful questions or advice requests in support of naturally communicating with the human. Together, they promise a computational approach to learning language while interacting with a human about a task, taking advice and planning to act and respond.

**Accomplishments:** Our objectives were to:
Create interactive agents for the collaborative building task in Minecraft that communicate successfully (give or execute instructions).
To achieve this objective, we have
Collected a human-human dialog dataset (509 conversations and game logs) for building collaborative agents in Minecraft (ACL 2019)
Built a baseline neural Architect system that can generate a response given the current game context whenever it is triggered, and devised and performed human evaluation on our generation task. Human evaluation shows that this system generates fully or partially correct instructions close to half the time (ACL 2019).
Built a neural Builder system that can execute instructions in their dialogue context. Even though this system has no access to any kind of prior domain knowledge (including spatial relations, names of structures, etc.), it shows a good understanding of instructions that involve create simple structures like rows or floating blocks. It also seems to

understand concepts like a “gap” between blocks, and displays plausible interpretations of instructions such as “and now do the same on the other side” that refer both to complex action sequences that were executed earlier in the game (“the same”) and to spatial relations that depend not only on the current built structure, but also the builder’s current perspective (“the other side”)

Built a rule-based Builder system that can execute simple instructions and ask a limited variety of clarification questions. This frontend of this system is a rule based parser to convert natural language instructions into a predicate-logic format that can be interpreted by the planning module. This parser uses a list of regular expressions and templates. The planning (problem solving) module translates these predicate logic representations into executable instructions to the Minecraft agent, and identifies missing or conflicting information. A dialog manager controls the overall flow of the interaction and generates utterances (e.g. clarification questions to request missing information).

Design a language comprehension module that converts natural language instructions into a formal representation that can be interpreted by the planner. Instructions may be given across multiple dialog turns, sharing context and mentioned entities across turns.

Design and implement an annotation schema for the Illinois Minecraft dialogues that is capable of capturing spatial semantics and pragmatics and grounding it all to the quantified space. Annotations are used as training data for the AI Architect and Builder.

Develop a Human-AI collaborative problem-solving architecture.

Develop a learning framework that allows the agent to learn and assimilate new concepts via interaction with its human collaborator. To achieve this:

We also worked on learning and using continuous representation of spatial relations. We looked at the following task - given two entities, predict the spatial relationship between these two entities in the context of an image. This is related to the task of predicting the location of one entity given another entity and the spatial relationship between them. We plan to develop a unified model to perform both tasks jointly.

We developed a framework that actively queries human expert for additional knowledge to improve efficiency and efficacy in hierarchical planning (AAMAS 2018). This framework initiates a 2-way dialog between human and agent, reduces burden on humans to provide all relevant knowledge upfront in CwC blocks world apparatus and proved some theoretical guarantees (Knowledge-Based Systems, 2019).

We developed a Human-AI collaboration framework that allows human experts to communicate additional knowledge in the form of inductive bias when learning rich concepts (Frontiers in Robotics and AI 2020, CoDS-COMAD 2020, AAAI 2020).

Collaborative systems need to faithfully model uncertainty and structure for explainability. The downside of such powerful representation is the computational complexity of counting satisfied instances. We have developed approximation techniques for efficient counting in human machine communication (AAAI 2019 & ProbProg 2018).

Develop a deeper understanding of the ability of existing systems to understand spatial relations and approaches to improving this using a richer representation and better modeling.

We developed a more expressive representation for spatial relationships called spatial configuration. We showed the alignment of elements of the spatial configuration with the new roles introduced in spatial AMR. Published at LREC 2020.

We developed a multi-modal model to understand implicit and explicit spatial relationships in language. We showed that pre-trained language models can give a strong prior to predict spatial relationships between objects, but by itself is not sufficient. We showed how combining image representations with the contextual representation of words can outperform either approach in isolation, even in the zero-shot setting. Published at LREC 2020.

We demonstrated that instruction following systems are very sensitive to perturbations in either modalities: language or world state. We characterized the expectations from an instruction following system and designed perturbations to show existing systems fail on these expectations. We then proposed an adversarial data augmentation framework that improves model performance on the perturbed examples, while preserving the standard performance on the original test set. Published at NAACL-HLT 2021

To better understand the challenge of spatial relations, we closely investigated the Blocks World dataset (Bisk et al, NAACL-HLT 2016). We showed that curriculum driven data augmentation and knowledge-free auxiliary tasks can help improve model performance. Submitted to ACL 2021.

# RPPR Final Report
as of 07-Apr-2021

**Training Opportunities:** At Illinois, Computer Science PhD students Anjali Narayan-Chen and Prashant Jayannavar have been fully supported by this grant. They have both become experts in grounded dialogue, and neural models for generation and execution. They presented our papers at ACL 2019 and 2020. Anjali Narayan-Chenhas now graduated.
At the Univeristy of Pennsylvania, Soham Dan, a CIS PhD student has been supported by this grant and has learnt about several important topics under the umbrella of human-computer interaction - formal spatial semantics, continuous representation of relations, semantic parsing and extended semantic role labeling.
At Colorado, Julia Bonn, MA LING, has become an expert in the linguistics literature that deals with spatial relations, and also on related work in the NLP community such as ISO Space-ML and the Tulane spatial relations annotation project. She brings her expertise in AMR annotation and multi-sentence AMR annotation to the project, and has been training Kristin Wright-Bettner in those skills, as well as Michael Regan. She is now a Linguistics PhD student at Colorado. Julia and Ghazaleh were trained on VerbNet by Susan Brown, Research faculty at Colorado. Jon Cai, a CS PhD student, was trained in neural net techniques for AMR parsing, with guidance from James Gung, a senior CS PhD student. James Gung graduated with a PhD in CS in May 2020, and developed the VerbNet parser. Kevin Stowe and Ghazaleh Kazeminejad contributed to the development of the Spatial Relations Ontology, collaborating with Julia Bonn. Kevin Stowe graduated with a PhD in LinguisticsJuly, 2019.
At UTD and WSU, Graduate (PhD) students Mayukh Das (UTD), Harsha Kokel (UTD) and Rakibul Islam (WSU) have immensely benefitted from this grant on improving their fundamentals on human-machine collaborative problem solving. They have significantly improved their understanding of the communication issues when collaboratively solving planning tasks. Mayukh's entire PhD was funded from this grant and his thesis is fully supported by the findings from this grant (graduated 2019 Fall). A majority of Rakib's thesis has also been funded from this grant. The grant helped kickstart Harsha's thesis and she has already a AAAI paper from this grant.

**Results Dissemination:** Invited Talks that reference CwC:

-- Martha Palmer
The Blocks World Redux
Invited Keynote, Combined Workshop on Spatial Language Understanding and Grounded Communication for Robotics, June 6, 2019
Natural Language Seminar, U of Southern California - ISI - NL Seminar, May 16, 2019.
Linguistics Colloquium, University of New Mexico, March 15, 2019
CS Colloquium, University of Colorado, December 6, 2018
Artificial Intelligence Forum, University of Texas, November 26, 2018.
CSE Graduate Seminar, Texas A&M University, November 28, 2018.

-- Julia Hockenmaier
Stony Brook University, September 2019
LTI Colloquium, Carnegie Mellon University, April 2020
Invited Keynote, ICML Workshop on Learning in Open Worlds, 2020
Invited Keynote, EMNLP Workshop on Spatial Language Understanding, 2020
Department of Linguistics, Gothenburg University, March 2021

-- Sriraam Natarajan
Electrical Engineering and Computer Science Colloquium - Oregon State University, 2017.
Machine Learning Seminar, University of Texas at Austin, 2018
CS Colloquium, University of British Columbia, 2018
CS Colloquium, Washington State University, 2018
Invited Speaker, CoDS-COMAD 2019
Key Note Speaker, AI for Social Good, 2019

We have presented our research findings and given several talks on Human-Allied AI, our group's perspective on Human-AI collaboration. Mayukh Das presented talks at the UT Dallas Outreach program inaugural address in 2017 and 2018 as well as at KCAP '18 and AAAI '19, Honolulu, Hawaii. Sriraam Natarajan presented at several venues including PEROT museum (reaching out non-scientific audience), NSF Balkan Workshop in Serbia, Central Texas Symposium, IIT Madras, invited talk at ACM CoDS-COMAD conference, AI for Social Good Conference and Dagstuhl seminar to name a few.f
In an attempt to reach out to a wider audience, most of whom may not be AI practitioners, we created newsfeeds and blogs discussing our research on Human-in-the-loop AI which we believe is a major component of the vision of the CwC project. This allows us to present scientific progress in this regard in the form of simplified non peer-reviewed articles, that can reach the interested communities who may or may not have formal training in the field of AI, or Computer Science in general.

**Honors and Awards:** Sriraam Natarajan received the Intel Faculty award for Human-Machine Collaboration and Turvo faculty award for developing human-in-the-loop planning systems.

Martha Palmer became an AAAI Fellow in 2020

**Protocol Activity Status:**

**Technology Transfer:** Nothing to Report

**PARTICIPANTS:**

**Participant Type:** PD/PI
**Participant:** Julia Hockenmaier
**Person Months Worked:** 5.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N

Other Collaborators:

**Participant Type:** Co PD/PI
**Participant:** Sriraam Natarajan
**Person Months Worked:** 5.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

**Participant Type:** PD/PI
**Participant:** Dan Roth
**Person Months Worked:** 5.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

**Participant Type:** Co PD/PI
**Participant:** Martha Palmer
**Person Months Worked:** 2.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

**Participant Type:** Co PD/PI
**Participant:** Janardhan Rao Doppa
**Person Months Worked:** 5.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

**Participant Type:** Graduate Student (research assistant)
**Participant:** Anjali Yuan Narayan-Chen
**Person Months Worked:** 15.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

**Participant Type:** Graduate Student (research assistant)
**Participant:** Prashant Jayannavar
**Person Months Worked:** 15.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:

National Academy Member: N
Other Collaborators:

**Participant Type:** Graduate Student (research assistant)
**Participant:** Soham Dan
**Person Months Worked:** 15.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

**Participant Type:** Graduate Student (research assistant)
**Participant:** Colin Graber
**Person Months Worked:** 12.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

**Participant Type:** Graduate Student (research assistant)
**Participant:** Harsha Kokel
**Person Months Worked:** 15.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

**Participant Type:** Graduate Student (research assistant)
**Participant:** Mayukh Das
**Person Months Worked:** 15.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

**Participant Type:** Graduate Student (research assistant)
**Participant:** Rakibul Islam
**Person Months Worked:** 15.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

**Participant Type:** Faculty
**Participant:** Susan Windisch Brown
**Person Months Worked:** 4.00 **Funding Support:**
Project Contribution:
International Collaboration:

International Travel:
National Academy Member: N
Other Collaborators:

**Participant Type:** Graduate Student (research assistant)
**Participant:** Julia Bonn
**Person Months Worked:** 9.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

**Participant Type:** Graduate Student (research assistant)
**Participant:** Kristin Wright-Bettner
**Person Months Worked:** 2.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

**Participant Type:** Graduate Student (research assistant)
**Participant:** Michael Regan
**Person Months Worked:** 2.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

**Participant Type:** Graduate Student (research assistant)
**Participant:** Jenette Preciado
**Person Months Worked:** 3.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

**Participant Type:** Graduate Student (research assistant)
**Participant:** Jon Cai
**Person Months Worked:** 1.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

**Participant Type:** Graduate Student (research assistant)
**Participant:** James Gung
**Person Months Worked:** 3.00 **Funding Support:**
Project Contribution:

International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

**Participant Type:** Graduate Student (research assistant)
**Participant:** James Gung
**Person Months Worked:** 1.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

**Participant Type:** Graduate Student (research assistant)
**Participant:** Kevin Stowe
**Person Months Worked:** 6.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

**Participant Type:** Graduate Student (research assistant)
**Participant:** Katie Conger
**Person Months Worked:** 3.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

**Participant Type:** Graduate Student (research assistant)
**Participant:** Ghazaleh Kazeminejad
**Person Months Worked:** 3.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

**ARTICLES:**

# RPPR Final Report
as of 07-Apr-2021

**Publication Type:** Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
**Journal:** Knowledge-Based Systems
Publication Identifier Type: DOI Publication Identifier: 10.1016/j.knosys.2018.11.028
Volume: 165 Issue: First Page #: 219
Date Submitted: 10/1/19 12:00AM Date Published: 2/1/19 6:00AM
Publication Location:
**Article Title:** Planning with actively eliciting preferences
**Authors:** Mayukh Das, Phillip Odom, Md. Rakibul Islam, Janardhan Rao (Jana) Doppa, Dan Roth, Sriraam Natara
**Keywords:** Active preference elicitationHuman-in-the-loopPlanningHTNHuman–agent interaction
**Abstract:** Planning with preferences has been employed extensively to quickly generate high-quality plans. However, it may be difficult for the human expert to supply this information without knowledge of the reasoning employed by the planner. We consider the problem of actively eliciting preferences from a human expert during the planning process. Specifically, we study this problem in the context of the Hierarchical Task Network (HTN) planning framework as it allows easy interaction with the human. We propose an approach where the planner identifies when and where expert guidance will be most useful and seeks expert's preferences accordingly to make better decisions. Our experimental results on several diverse planning domains show that the preferences gathered using the proposed approach improve the quality and speed of the planner, while reducing the burden on the human expert.
**Distribution Statement:** 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support: **Y**

**Publication Type:** Journal Article Peer Reviewed: **Publication Status:** 4-Under Review
**Journal:** Journal of Artificial Intelligence Research (JAIR)
Publication Identifier Type: Publication Identifier:
Volume: Issue: First Page #:
Date Submitted: Date Published:
Publication Location:
**Article Title:** Human-in-the-loop Learning Framework via Tree-based Ensembles: Algorithms and Interpretability.
**Authors:** Das, M., Wu, Y., Khot, T., Kersting, K., & Natarajan, S
**Keywords:** human in the look learning
**Abstract:** A human-in-the-loop learning framework that leverages the inherent strengths of tree-based unsupervised ensembles to maximize the discovery of anomalies via label feedback from human analyst. The tree-based structure allows for improved interpretability resulting in two-way communication between human analyst and computer for complex problem solving.
**Distribution Statement:** 2-Distribution Limited to U.S. Government agencies only; report contains proprietary info
Acknowledged Federal Support:

**CONFERENCE PAPERS:**

**Publication Type:** Conference Paper or Presentation **Publication Status:** 1-Published
**Conference Name:** Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics
Date Received: 01-Oct-2019 Conference Date: 29-Jul-2019 Date Published: 29-Jul-2019
Conference Location: Florence, Italy
**Paper Title:** Collaborative Dialogue in Minecraft
**Authors:** Anjali Narayan-Chen, Prashant Jayannavar, Julia Hockenmaier
Acknowledged Federal Support: **Y**

# RPPR Final Report
as of 07-Apr-2021

**Publication Type:** Conference Paper or Presentation **Publication Status:** 1-Published
**Conference Name:** Twenty-Eighth International Joint Conference on Artificial Intelligence {IJCAI-19}
Date Received: 01-Oct-2019 Conference Date: 10-Aug-2019 Date Published: 10-Aug-2019
Conference Location: Macao, China
**Paper Title:** Randomized Greedy Search for Structured Prediction: Amortized Inference and Learning
**Authors:** Chao Ma, F A Rezaur Rahman Chowdhury, Aryan Deshwal, Md Rakibul Islam, Janardhan Rao Doppa,
Acknowledged Federal Support: **Y**

**Publication Type:** Conference Paper or Presentation **Publication Status:** 1-Published
**Conference Name:** Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, {IJCAI-19}
Date Received: 01-Oct-2019 Conference Date: 10-Aug-2019 Date Published: 10-Aug-2019
Conference Location: Macao, China
**Paper Title:** Learning and Inference for Structured Prediction: A Unifying Perspective
**Authors:** Aryan Deshwal , Janardhan Rao Doppa, Dan Roth
Acknowledged Federal Support: **Y**

**Publication Type:** Conference Paper or Presentation **Publication Status:** 1-Published
**Conference Name:** AAAI 2019
Date Received: 01-Oct-2019 Conference Date: 27-Jan-2019 Date Published: 17-Jul-2019
Conference Location: Honolulu, Hawaii
**Paper Title:** Fast Relational Probabilistic Inference and Learning: Approximate Counting via Hypergraphs
**Authors:** Das, M., Dhami, D.S., Kunapuli, G., Kersting, K., & Natarajan, S
Acknowledged Federal Support: **Y**

**Publication Type:** Conference Paper or Presentation **Publication Status:** 1-Published
**Conference Name:** AAMAS 2018
Date Received: 01-Oct-2019 Conference Date: 10-Jul-2018 Date Published: 10-Jul-2018
Conference Location: Stockholm, Sweden
**Paper Title:** Preference-Guided Planning: An Active Elicitation Approach
**Authors:** Mayukh Das, Philip Odom, Md. Rakibul Islam, Janardhan Rao Doppa, Dan Roth, Sriraam Natarajan
Acknowledged Federal Support: **Y**

**Publication Type:** Conference Paper or Presentation **Publication Status:** 1-Published
**Conference Name:** Combined Workshop on Spatial Language Understanding (SpLU) & Date: June 6, 2019
Grounded Communication for Robotics (RoboNLP), non-archival poster presentation.
Date Received: 02-Oct-2019 Conference Date: 04-Jun-2019 Date Published: 04-Jun-2019
Conference Location: Minneapolis, MN
**Paper Title:** From Spatial Relations to Spatial Configurations
**Authors:** Soham Dan, Parisa Kordjamshidi, Julia Bonn, Archna Bhatia, Martha Palmer, and Dan Roth
Acknowledged Federal Support: **Y**

**Publication Type:** Conference Paper or Presentation **Publication Status:** 1-Published
**Conference Name:** Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics
Date Received: 01-Apr-2021 Conference Date: 05-Jul-2021 Date Published:
Conference Location: Online
**Paper Title:** Learning to execute instructions in a Minecraft dialogue
**Authors:** Prashant Jayannavar, Anjali Narayan-Chen, Julia Hockenmaier
Acknowledged Federal Support: **Y**

# RPPR Final Report
as of 07-Apr-2021

**Publication Type:** Conference Paper or Presentation **Publication Status:** 1-Published
**Conference Name:** Proceedings of the First Workshop on Language Grounding for Robotics
Date Received: 01-Apr-2021 Conference Date: 01-Aug-2017 Date Published: 01-Aug-2017
Conference Location: Vancouver, Canada
**Paper Title:** Towards Problem Solving Agents that Communicate and Learn
**Authors:** Anjali Narayan-Chen, Colin Graber, Mayukh Das, Md Rakibul Islam, Soham Dan, Sriraam Natarajan, Ja
Acknowledged Federal Support: **Y**

**Publication Type:** Conference Paper or Presentation **Publication Status:** 1-Published
**Conference Name:** 12th Conference on Language Resources and Evaluation
Date Received: 01-Apr-2021 Conference Date: 01-May-2020 Date Published: 01-May-2020
Conference Location: Marseille, France
**Paper Title:** Spatial AMRs: Expanded Spatial Annotation in the Context of a Grounded Minecraft Corpus
**Authors:** Julia Bonn, Martha Palmer, Zheng Cai, Kristin Wright-Bettner
Acknowledged Federal Support: **Y**

**Publication Type:** Conference Paper or Presentation **Publication Status:** 1-Published
**Conference Name:** Proceedings of the 12th Language Resources and Evaluation Conference
Date Received: 01-Apr-2021 Conference Date: 01-May-2020 Date Published: 01-May-2010
Conference Location: Marseille, France
**Paper Title:** From Spatial Relations to Spatial Configurations
**Authors:** Soham Dan, Parisa Kordjamshidi, Julia Bonn, Archna Bhatia, Zheng Cai, Martha Palmer, Dan Roth
Acknowledged Federal Support: **Y**

**Publication Type:** Conference Paper or Presentation **Publication Status:** 1-Published
**Conference Name:** 12th Conference on Language Evaluation and Resources
Date Received: 01-Apr-2021 Conference Date: 02-May-2020 Date Published: 02-May-2020
Conference Location: Marseille France
**Paper Title:** Understanding Spatial Relations through Multiple Modalities.
**Authors:** Dan S, He H, Roth D
Acknowledged Federal Support: **Y**

**Publication Type:** Conference Paper or Presentation **Publication Status:** 1-Published
**Conference Name:** CODS COMAD 2021 & 8th ACM IKDD CODS and 26th COMAD
Date Received: 01-Apr-2021 Conference Date: 03-Jan-2021 Date Published: 03-Jan-2021
Conference Location: Bangalore India
**Paper Title:** Human-Guided Learning of Column Networks;#x23;x3a#x3b; Knowledge Injection for Relational Deep Learning
**Authors:** Das, M., Dhami, D., Yu, Y., Kunapuli, G., ; Natarajan, S
Acknowledged Federal Support: **Y**

**Publication Type:** Conference Paper or Presentation **Publication Status:** 1-Published
**Conference Name:** Proceedings of the 2016 SIAM International Conference on Data Mining
Date Received: 01-Apr-2021 Conference Date: 05-May-2016 Date Published: 05-May-2016
Conference Location: Miami, FL
**Paper Title:** Scaling Lifted Probabilistic Inference and Learning Via Graph Databases
**Authors:** Das, M., Wu, Y., Khot, T., Kersting, K., ; Natarajan, S
Acknowledged Federal Support: **Y**

# RPPR Final Report
as of 07-Apr-2021

**Publication Type:** Conference Paper or Presentation **Publication Status:** 1-Published
**Conference Name:** Workshop on Human In the Loop Learning (HILL) 2019.
Date Received: 01-Apr-2021 Conference Date: 14-Jun-2019 Date Published: 14-Jun-2019
Conference Location: Long Beach, CA
**Paper Title:** Knowledge-augmented Column Networks: Guiding Deep Learning with Advice,
**Authors:** Das, M., & Dhami, D.S., & Yu, Y., & Kunapuli, G., & Natarajan, S.
Acknowledged Federal Support: **Y**

**Publication Type:** Conference Paper or Presentation **Publication Status:** 1-Published
**Conference Name:** Proceedings of the Inaugural International Conference on Probabilistic Programming
Date Received: 01-Apr-2021 Conference Date: 04-Oct-2018 Date Published: 02-Apr-2018
Conference Location: Boston, MA
**Paper Title:** Approximate Counting for Fast Inference and Learning in Probabilistic Programming
**Authors:** Das, M., & Dhami, D.S., & Kunapuli, G., & Kersting, K., & Natarajan, S
Acknowledged Federal Support: **Y**

**Publication Type:** Conference Paper or Presentation **Publication Status:** 1-Published
**Conference Name:** ICML Workshop on Human in the Loop Learning, 2020.
Date Received: 01-Apr-2021 Conference Date: 01-Aug-2020 Date Published: 01-Aug-2020
Conference Location: online
**Paper Title:** GLAD: GLocalized Anomaly Detection via Active Feature Space Suppression
**Authors:** Islam, M. R., Das, S., & Doppa, J. R., & Natarajan, S.
Acknowledged Federal Support: **Y**

**Publication Type:** Conference Paper or Presentation **Publication Status:** 1-Published
**Conference Name:** Human-Machine Collaborative Learning, AAAI Workshops, 2017
Date Received: 01-Apr-2021 Conference Date: 05-Feb-2017 Date Published: 05-Feb-2017
Conference Location: San Francisco, CA
**Paper Title:** Active Preference Elicitation for Planning, Human-Machine Collaborative Learning,
**Authors:** Das, M., Islam, M.R., Doppa, J.R., Roth, D., & Natarajan, S
Acknowledged Federal Support: **Y**

**Publication Type:** Conference Paper or Presentation **Publication Status:** 1-Published
**Conference Name:** Workshop on Statistical Relational AI (StarAI) 2020
Date Received: 01-Apr-2021 Conference Date: 07-Feb-2020 Date Published: 07-Feb-2020
Conference Location: New York, NY
**Paper Title:** One-Shot Induction of Generalized Logical Concepts via Human Guidance
**Authors:** Das, M., & Ramanan, N., & Doppa, J.R., & Natarajan, S
Acknowledged Federal Support: **Y**

**Publication Type:** Conference Paper or Presentation **Publication Status:** 1-Published
**Conference Name:** K-CAP 2017: Knowledge Capture Conference
Date Received: 01-Apr-2021 Conference Date: 01-Dec-2017 Date Published: 01-Dec-2017
Conference Location: Austin TX USA
**Paper Title:** User Friendly Automatic Construction of Background Knowledge
**Authors:** Hayes, A.L., & Das, M., & Odom, P., & Natarajan, S.
Acknowledged Federal Support: **Y**

# RPPR Final Report
as of 07-Apr-2021

**Publication Type:** Conference Paper or Presentation **Publication Status:** 1-Published
**Conference Name:** Proceedings of the First International Workshop on Designing Meaning Representations
Date Received: 01-Apr-2021 Conference Date: 01-Aug-2019 Date Published: 01-Aug-2019
Conference Location: Florence, Italy
**Paper Title:** VerbNet Representations: Subevent Semantics for Transfer Verbs
**Authors:** Susan Windisch Brown, Julia Bonn, James Gung, Annie Zaenen, James Pustejovsky, Martha Palmer
Acknowledged Federal Support: **Y**

**Publication Type:** Conference Paper or Presentation **Publication Status:** 1-Published
**Conference Name:** Proceedings of the 27th International Conference on Computational Linguistics
Date Received: 01-Apr-2021 Conference Date: 02-Aug-2018 Date Published: 01-Aug-2018
Conference Location: Santa Fe, NM
**Paper Title:** Automatically Extracting Qualia Relations for the Rich Event Ontology
**Authors:** Ghazaleh Kazeminejad, Claire Bonial, Susan Windisch Brown, Martha Palmer
Acknowledged Federal Support: **Y**

**Publication Type:** Conference Paper or Presentation **Publication Status:** 1-Published
**Conference Name:** Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)
Date Received: 01-Apr-2021 Conference Date: 01-May-2018 Date Published: 01-May-2018
Conference Location: Miyazaki, Japan
**Paper Title:** Integrating Generative Lexicon Event Structures into VerbNet
**Authors:** Susan Windisch Brown, James Pustejovsky, Annie Zaenen, Martha Palmer
Acknowledged Federal Support: **Y**

**Publication Type:** Conference Paper or Presentation **Publication Status:** 1-Published
**Conference Name:** Interoperability for Semantic Annotation (ISA-12)
Date Received: 01-Apr-2021 Conference Date: 23-May-2016 Date Published: 23-May-2018
Conference Location: Portoroz, Slovenia
**Paper Title:** Integrating VerbNet and GL Predicative Structure
**Authors:** James Pustejovsky, Martha Palmer, Annie Zaenen, and Susan Brown.
Acknowledged Federal Support: **Y**

**Publication Type:** Conference Paper or Presentation **Publication Status:** 1-Published
**Conference Name:** Interoperability for Semantic Annotation (ISA-12)
Date Received: 01-Apr-2021 Conference Date: 23-May-2016 Date Published: 23-May-2016
Conference Location: Portoroz, Slovenia
**Paper Title:** A Lexically-Informed Upper Level Event Ontology,
**Authors:** Claire Bonial, Susan Brown and Martha Palmer,
Acknowledged Federal Support: **Y**

**Publication Type:** Conference Paper or Presentation **Publication Status:** 1-Published
**Conference Name:** Proceedings of the 58th Annual Conference of the Association for Computational Linguistics (ACL)
Date Received: 01-Apr-2021 Conference Date: 05-Jul-2020 Date Published: 05-Jul-2020
Conference Location: online
**Paper Title:** Structured Tuning for Semantic Role Labeling
**Authors:** Tao Li, Parth Anand Jawale, Martha Palmer and Vivek Srikumar
Acknowledged Federal Support: **Y**

# RPPR Final Report
as of 07-Apr-2021

**Publication Type:** Conference Paper or Presentation **Publication Status:** 1-Published
**Conference Name:** Proceedings of the 23rd Conference on Computational Natural Language Learning #x28;CoNLL#x29;
Date Received: 01-Apr-2021 Conference Date: 06-Nov-2019 Date Published: 06-Nov-2019
Conference Location: Hong Kong
**Paper Title:** Linguistic Analysis Improves Neural Metaphor Detection
**Authors:** Kevin Stowe, Sarah Moeller, Laura Michaelis, Martha Palmer,
Acknowledged Federal Support: **Y**

**DISSERTATIONS:**

**Publication Type:** Thesis or Dissertation
**Institution:** University of Illinois at Urbana-Champaign
Date Received: Completion Date: 12/1/20 10:56PM
**Title:** Towards collaborative dialogue in Minecraft
**Authors:** Anjali Yuan Narayan-Chen
Acknowledged Federal Support: **Y**

**Publication Type:** Thesis or Dissertation
**Institution:** University of Texas at Dallas
Date Received: 01-Apr-2021 Completion Date: 12/1/19 6:00AM
**Title:** Human-Allied Efficient and Effective Learning in Noisy Domains
**Authors:** Mayukh Das
Acknowledged Federal Support: **Y**

# Communicating with Computers
# Final Report

*PIs: Janardhan Rao Doppa (Washington State University), Julia Hockenmaier (University of Illinois), Sriraam Natarajan (University of Texas Dallas), Martha Palmer (University of Colorado), Dan Roth (University of Pennsylvania)*

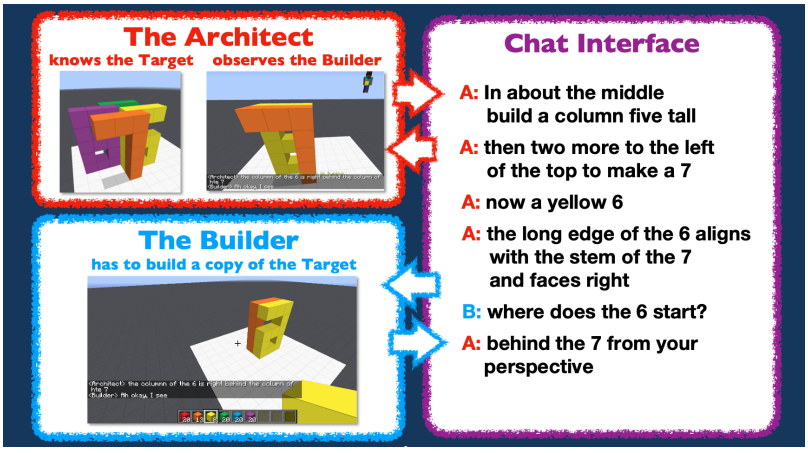
## 1 Introduction

Our team worked on the Blocksworld scenario. Although we initially focused on a 2D version of this scenario (RoboNLP '17), we later moved to fully embodied agents in the virtual 3D environment provided by the Minecraft gaming platform. To developed agents for this scenario, we first collected a dataset of humans performing our task. We used this dataset to train neural agents and to extend a prior semantic annotation scheme to cover the linguistic phenomena that arise in this domain (including spatial relations and dialogue information).
We also developed models for Human AI communication for concept learning and classification as well as more general models to understand spatial relations.

## 2 Research Contributions

### 2.1 University of Illinois

**Minecraft as a virtual platform for BlocksWorld scenarios**
Minecraft is a popular virtual gaming platform, where avatars navigate a 3D environment and can place block-like materials in a discrete grid. Minecraft is becoming increasingly popular for AI research, especially in the context of reinforcement learning for navigation, and Microsoft Research has developed an API to facilitate AI research with Minecraft agents.




**The Minecraft Collaborative Building Task and the Minecraft Dialogue Corpus**
The Illinois group focused on leveraging the virtual gaming platform Minecraft to facilitate the creation of systems that can collaborate and communicate with each other to solve BlocksWorld-like tasks. To focus on the communication and

collaboration aspects, we defined the so-called Minecraft Collaborative Building Task. As shown in the figure on the left, this is a two-player game where one player (the Architect) is shown a 3D target structure constructed out of building blocks, and has to instruct the other player (the Builder) to construct the same structure via chat. The Architect can observe the Builder, but cannot place any blocks. The Builder can chat with the Architect (e.g. to ask clarification questions), but needs to be able to execute these instructions. Both players need different capabilities: the Architect needs to be able to compare the target with the currently built structure to identify potential mistakes and next steps, and then verbalize these as instructions, while the Builder needs to be able to understand the Architect's instructions, execute them when possible, and ask appropriate clarification questions otherwise.
To perform this research, we first had to extend Microsoft Research's Project Malmo API to allow us to collect natural language dialogue and game logs between human players, and to deploy agents in an interactive setting.
Then, to understand how people perform this task, and to obtain training and evaluation data for supervised learning of agents that can perform either role, we collected and released the Minecraft Dialogue corpus, a corpus of 500+human-human dialogues and game logs for this task. (ACL '19). The figure above shows a snippet of such a dialogue between two human players. As can be seen, the Architect provide multi-utterance instructions, and players refer to substructures (in this case, numbers such as 6 or 7 or the "*stem of the 7*", or more generally, columns, rows, edges), use spatial relations (that often depend on the Builder's current perspective, or that refer to absolute positions in the build region (the white square on the ground), as in "the middle". Human Builders commonly ask clarification questions ("*where does the 6 start?*"), which require them to identify when instructions are vague or unclear.
This task forms the baseline of a NeurIPS '21 competition proposal led by Julia Kiseleva et a. at Microsoft Research.

**Neural baseline models for the Architect and Builder**
This dataset then allowed us to develop and train neural models for the Architect and Builder in a supervised fashion. Although we ultimately wish to develop fully interactive agents that can successfully engage with a human player long enough to complete an entire target structure, our initial efforts in both cases focused on developing models that could produce adequate utterances (for the Architect) or sequences of block placements and removals (for the Builder) in the context of a human-human game where the human Architect said something or where the human Builder placed or removed blocks. We see this as an important first step towards full interactivity.
Given the pervasive success of so called end-to-end neural models in many areas of NLP over the last few years, we were also particularly interested to investigate how well such models (which are trained on raw data, without any intermediate symbolic representations) can perform on these challenging tasks, especially given the relatively small size of our training data.

The main challenge in developing models for the Architect is the need to compare the built to the target structure. Our first baseline models (ACL '19) defined a simple "block counters" approach that relied on identifying the number of blocks (globally, or locally, for each cell neighboring and including the last cell where a block was placed) that need to be placed or

removed to obtain the target. While this gave us better results than more complex approaches (or ignoring the game state entirely), this leads only to a moderately high accuracy of spatial and color terms (as determined by automatic evaluation) and to a similarity moderate accuracy of instructions as determined by human evaluation. We now believe that our dialog data may be too complex (and too limited in size) to train Architect models from scratch, and are pursuing a curriculum learning approach where we automatically generate a large number of simple instructions (without a dialog context) to train an initial model.

While the Builder needs to be able to ask clarification questions (and often engages in chit-chat with the Architect in our human-human games), the ability to know when an instruction can either be executed in multiple ways or possibly not at all in a current game state (and hence necessitates either a clarification request or a confirmation) requires an execution model. In our ACL'20 paper, we therefore first focused on the task of predicting action sequences (block placements and removals) in game states where the human Builder placed or removed blocks. We novel 3D convolutional neural network architecture to represent the world state and predict the next action (placing or removing a block in a particular grid cell). Automatic evaluation as well as human inspection showsthat this model produces many plausible interpretations of instructions that contain references to complex shapes ("put blue blocks in the gap") or structures ("make it float"), or that require an understanding of prior actions and utterances ("now do the same on the other side").

**Unpublished work and work in progress:**
We also developed a synthetic task and dataset for shape recognition and localization in Minecraft -- given an arbitrary set of blocks, we predict what shapes exist in it and (in the localization model) identify their block locations. We can train highly accurate neural models on these tasks, and work in progress shows that the ability to localize shapes gives a further increase in performance on the Builder's action prediction task.
Other work in progress includes a reranking models for the Builder. The neural model for the aforementioned BAP task is based on a recurrent decoder network that generates/decodes action sequences using greedy decoding. Though beam search decoding yields worse performance, we discovered that there are high quality sequences in the beam that are not ranked high enough by the BAP model. This motivated us to explore a neural re-ranker model that would re-rank sequences in the beam for better performance.
We are also working on curriculum learning for the Builder and the Architect, where we train on simulated data that mimics the Minecraft game logs but are much simpler, in addition to the original Minecraft data. We are also starting to begin work on reinforcement learning for both tasks, as well as on models that can predict when the Builder should speak.
And although we showed We are also working towards a better Architect model using CNNs.

## 2.2 University of Colorado

### Language Comprehension, Annotation, and Parsing

With the goal of creating interactive agents capable of communicating with human participants in collaborative structure building tasks, we designed a language comprehension system that

uses Abstract Meaning Representation annotation as an intermediary layer through which to convert spatial dialogue into language-independent spatial predicates for downstream systems. Our products include a new Spatial Conceptualization, a new inventory of spatial PropBank/AMR rolesets, a new expanded AMR schema (Spatial AMR), an annotated Spatial AMR corpus taken from UIUC's Minecraft Dialogue Corpus, a state of the art custom AMR parser, and an AMR translator that converts AMR graphs into a modified first order logic format. (LREC 2020, LRE and IWCS submissions 2021).

**Spatial Conceptualization**
The Spatial Conceptualization places fine-grained spatial semantic properties into categories. Each category is associated with a set of schematic predicates, designed by Jerry Hobbs for this project, that serve as stackable semantic entailments for the new spatial PB/AMR rolesets. Once annotation with the new rolesets is complete, annotations can be broken down into these entailments and then repackaged into the planner-specific predicates used downstream. Based on the language observed in the Minecraft Dialogue Corpus, we identify categories including location, orientation, configuration, extent, direction, region, topology, dimensionality, scale, and frame of reference.

The Spatial Conceptualization grew out of our early collaboration with SIFT in creating spatially relevant ECIs for their Ecipedia. In addition to Mark Burstein, David McDonald, and Rusty Bobrow at SIFT, we collaborated with Jerry Hobbs, Jonathan Gordon, James Pustejovsky, Parisa Kordjamshidi, and Archna Bhatia throughout our development of this portion of the project.

In a related collaboration with James Pustejovsky also in support of our early ECI expansion goals, we overhauled VerbNet's semantic representations so that they better align with Generative Lexicon structures. The new representations use an updated set of semantic predicates (dovetailing with Jerry Hobbs's predicates and the Spatial Conceptualization where applicable), presented in an improved Generative Lexicon-compatible subevent structure schema. Much like the predicate entailments associated with each new spatial roleset, these new representations allow for clearer and more decomposable semantics that improve parsability; they also are much more effective at demonstrating the semantic relationships between VerbNet classes that the previous schema. (DMR Workshop at ACL 2019).

**Spatial AMR expansion & PropBank/AMR Rolesets**
We created Spatial AMR, an expanded AMR schema, to allow for more detailed, grounded coverage of spatial concepts within the already established general domain AMR framework. AMR is attractive as an annotation schema because of its user-friendliness, broad general semantic coverage, and ability to track intersentential coreference and implicit arguments. Spatial AMR adds more fine-grained semantic tools as well as pragmatic tools that allow multimodal contextual information about the environment to be incorporated into the graphs.

We added 200 new or expanded spatial PropBank rolesets to the PropBank/AMR lexical inventory that encode the semantics from the Spatial Configuration in the form the argument

structures and decomposable entailments. We also added 11 new general (not lemma-specific) semantic frames with corresponding roles (e.g., the frame have-orientation-91, which reifies the role :orientation) and new entity types (e.g. dimension-entity, slope-entity). The schema accommodates frame of reference annotation-- an essential step in converting between spatial instructions and grounded actions in the environment-- with the help of a new dummy AMR graph at the beginning of each dialogue. The dummy AMR defines the specific spatial frameworks that are referenced in a task, presenting them in the form of taggable nodes that are available for coreference annotation during the multi-sentence annotation pass. What this means is that for each utterance of a directed spatial expression like *to the right of*, we not only capture all of the entailments associated with *rightness* in general, but also whose right it is in this particular instance and how *rightness* in the context of that entity maps onto the environment at that moment in time. Our translator later converts this information into absolute coordinates.

**AMR annotation**

AMR annotation occurs in two passes-- first, a single sentence pass in which the surface form semantics of individual sentences are put into AMR graph form, and second, a multi sentence pass in which all of the single sentence AMR graphs are combined into a single document, all untagged arguments are reintroduced as potential implicit arguments, and coreference tracking between sentences is completed. Our Spatial AMR-annotated corpus includes single sentence annotation of all of the Minecraft Dialogue Corpus dialogues recorded from 3-30 through 4-14-- a total of 324 dialogues, 10,000 dialogue sentences, and an additional 15,000 automatically generated AMR graphs that encode Builder actions as sentences. The 324 dialogues annotated represent 60% of the 547 dialogues recorded by UIUC. Multi Sentence annotation was completed for 60 of these dialogues before an extended delay due to necessary retrofits to the annotation schema and multi-sentence AMR set-up tool. We are seeking additional funding for the needed additional annotation. We spawned a continuing annotation-trading collaboration with Claire Bonial, Clare Voss and David Traum at Army Research Labs which will eventually double our annotation numbers. This effort is currently delayed due to the ongoing COVID-19 pandemic.

## AMR parser

The automatic AMR parsers aim at providing real-time AMRs for adhoc downstream applications to extract structured semantics. We experimented with adapting several recent state-of-the-art AMR parsers as well as different training methods to encompass the spatial-relation dialogue domain. These include JAMR (Flanigan et al. 2016), an AMR parser from Lyu et al. (2018), and the JHU parser (Zhang et al. 2018). Among these we found the JHU parser with a data augmentation training method achieved the best performance and was the most robust. We were able to boost the F1 score to 72.9% by merging the Minecraft data with the newly released LDC 2017 data and introducing bias towards the Minecraft data. After post processing, we are able to reach a new best performance on the random split test set with a smatch score of 73%. Notice, this result does not include the trivial cases arising from the automatically generated AMR graphs (these trivial cases have 100% F1 under rule-based parsing), and are solely for descriptive purposes rather than expanding the content of the

conversational data through genuine, natural utterances (LREC 2020, LRE and IWCS submissions 2021).

### AMR translator

We further developed an AMR to logical format translator. This translator is used as an intermediary module that produces the needed input for the planner from the AMR graph. The planner takes in logical format tuples as its input to plan the configuration of the 3D Minecraft space and schedule the action sequence. However, AMR, as a formalism, does not provide this format directly. We adapted the approach described in Bos (2016) to translate AMR graphs into First Order Logic, which we then adapt to the final logical statements with the predicates expected by our planner. A preliminary result shows 71.08% of the logical terms of a 20 sentence test set are correctly mapped to our planner logic.

## 2.3 University of Pennsylvania:

**Semantic Parsing:**

We developed a semantic parser for the previously described Minecraft builder task. This builds on a semantic parser for a 2 dimensional blocks world environment. We designed a formal language to represent the spatial concepts and relations between objects. The semantic parser maps natural language instructions to this formal language, so that the planner can execute the semantic parse. We found that the parser, although rule-based, is able to handle a large number of instructions. It can also handle the addition of newly learned concepts in its vocabulary and parse future instructions involving that concept. However, since the instructions that can be handled are much simpler than the kind of instructions present in the Minecraft corpus, we realized the need to create a richer spatial representation within the formal language to handle more complex instructions (described in the following section). In parallel, the rule based semantic parser was integrated into the rest of the Minecraft builder pipeline to obtain a system that can take a natural language instruction, parse it into the formal language and execute the plan in the Minecraft world.

**Understanding Spatial Relations:**

1. Spatial Reasoning from language is essential for natural language understanding. Supporting it requires a representation scheme that can capture spatial phenomena encountered in language as well as in images and videos.Existing spatial representations are not sufficient for describing spatial configurations used in complex tasks. This work extends the capabilities of existing spatial representation languages and increases coverage of the semantic aspects that are needed to ground spatial meaning of natural language text in the world. Our spatial relation language is able to represent a large, comprehensive set of spatial concepts crucial for reasoning and is designed to support composition of static and dynamic spatial configurations. We integrate this language with the Abstract Meaning Representation (AMR) annotation schema and

present a corpus annotated by this extended AMR. To exhibit the applicability of our representation scheme, we annotate text taken from diverse datasets and show how we extend the capabilities of existing spatial representation languages with fine-grained decomposition of semantics and blend it seamlessly with AMRs of sentences and discourse representations as a whole.
Publication: Dan, S., Kordjamshidi, P., Bonn, J., Bhatia, A., Cai, Z., Palmer, M., and Roth, D., (2020). From Spatial Relations to Spatial Configurations. In Proceedings of the 12th Language Resources and Evaluation Conference (LREC) pp. 5855--5864.

2. Recognizing spatial relations and reasoning about them is essential in multiple applications including navigation, direction giving and human-computer interaction in general. Spatial relations between objects can either be explicit – expressed as spatial prepositions, or implicit – expressed by spatial verbs such as moving, walking, shifting, etc. Both these, but implicit relations in particular, require significant common sense understanding. In this work, we introduce the task of inferring implicit and explicit spatial relations between two entities in an image. We design a model that uses both textual and visual information to predict the spatial relations, making use of both positional and size information of objects and image embeddings. We contrast our spatial model with powerful language models and show how our modeling complements the power of these, improving prediction accuracy and coverage and facilitates dealing with unseen subjects, objects and relations.
Publication: Dan, S., He, H., & Roth, D. (2020, May). Understanding Spatial Relations through Multiple Modalities. In Proceedings of The 12th Language Resources and Evaluation Conference (LREC) (pp. 2368-2372).

3. Understanding and executing natural language instructions in a grounded domain is one of the hallmarks of artificial intelligence. In this work, we focus on instruction understanding in the blocks world domain and investigate the language understanding abilities of two top performing systems for this task. We aim to understand if the test performance of these models indicates an understanding of the spatial domain and of the natural language instructions relative to it, or whether they merely overfit spurious signals in the data-set. We formulate a set of expectations one might have from an instruction following model and concretely characterize the different dimensions of robustness such a model should possess. Despite decent test performance, we find that state-of-the-art models fall short of these expectations and are extremely brittle. We then propose a learning strategy that involves data augmentation and show through extensive experiments that the proposed learning strategy yields models that are competitive on the original test set while satisfying our expectations much better.
Publication: Dan, S., Zhou, M., Roth, D., Generalization in Instruction Following Systems, Accepted to NAACL-HLT 2021

4. Executing natural language instructions in a physically grounded domain requires a model that understands both spatial concepts such as left of and above, and the compositional language used to identify landmarks and articulate instructions relative to

them. In this work, we study instruction understanding in the blocks world domain. Given an initial arrangement of blocks and a natural language instruction, the system executes the instruction by manipulating selected blocks. The highly compositional instructions are composed of atomic components and understanding these components is a necessary step to executing the instruction. We show that while end-to-end training (supervised only by the correct block location)
fails to address the challenges of this task and performs poorly on instructions involving a single atomic component, knowledge-free auxiliary signals can be used to significantly improve performance by providing supervision for the instruction's components. Specifically, we generate signals that aim at helping the model gradually understand components of the compositional instructions, as well as those that help it better understand spatial concepts, and show their benefit to the overall task, especially when the training data is limited—which is usual in such tasks.
Publication: Dan,S., Han, X., Roth,D. Compositional Data and Task Augmentation for Instruction Following. (in submission to ACL 2021)

## 2.4 University of Texas Dallas and Washington State University

We developed a framework that can not only achieve cognitively coherent symmetric communication with human counterparts but can also leverage such communication to collaboratively learn to solve complex problems. We addressed this problem from three different perspectives -- (1) in the task of sequential decision-making including planning where the system communicates with the human and collaboratively solves the tasks, (2) in the traditional tasks of learning rich concepts and complex classification tasks where the human communication is used to provide a strong inductive/search bias to the learner, and (3) in the context of transfer and generalization, where human inputs are provided to the learning system for effective and efficient transfer across seemingly unrelated tasks.

**Human AI communication in sequential tasks:** As one of our first goals, we conceptualized and developed a robust Human-AI collaborative system that can assimilate human guidance/knowledge into a "problem-solving (AI) agent" (that solves sequential decision making tasks) in structured noisy environments. The system was designed in such a fashion that the problem/state spaces could be encoded to seamlessly align with the language and dialog modules. We adapted a hierarchical higher-order planner as the core problem solving engine and extended it with adaptive knowledge-guided decision making. This collaborative problem-solving agent has the capability to understand, quantify and measure "what-is-doesn't-know" (dearth of relevant information) and leverage that understanding to elicit "advice/knowledge" at the most relevant decision points from its human counterpart and potentially learn better solutions. Our publications (at AAMAS, AAAI and KBS) reinforce our claim and illustrate how our adaptive/active collaboration with human experts allow the problem-solving agent to discover more effective and efficient solutions. Along with many well-known problem-solving domains such as solitaire card games, our framework showed surprising but expected effectiveness on solving problems of building complex spatial structures

(blocks world and its variants), even using the Kinect-driven blocks-world apparatus provided by DARPA. Towards the end of the first phase of the CwC program we adapted the collaborative problem-solving/planning component for the 'builder' agent prototype described above that can communicate with human(s), understand the spatial structures that the human is trying to convey, generate solutions/plans (sequence of instructions/steps) by which the structure can be built and convey that back to the human. We conducted extensive human-subject evaluations and measured both quality of solutions generated by the builder agent under human-agent collaboration as well as the richness of bidirectional communication.

**Human AI communication for concept learning and classification:** In the 2nd phase, we have extended this framework to Minecraft, which allowed for considering a broader range of tasks with relaxed physical principles. This entailed updating the domain representation and translation protocols with the language/dialog modules. However, probably the most critical enhancement at this stage was the development of a learning framework in our problem-solver (decision-maker) which generalizes structural concepts from one or few instances described by the human architect and enhances its knowledge base progressively. The learning framework is also collaborative in nature since it elicits guidance from a human-teacher to balance the data/information sparsity and learns via a form of contrastive divergence (is the current model drifting too far when contrasted with known sample). This work was published in Frontiers in Robotics and AI journal. Parallely, to enhance the adaptability of our collaborative problem-solving/learning framework we proposed and developed Human-AI collaboration framework [Knowledge-augmented Column Networks (KCLNs)] for learning from low-level representations (ICML HILL workshop 2019; CODS-CoMAD 2021). Column Networks are `neuro-symbolic' deep models that can represent relational structures implicitly. K-CLN allows human experts to communicate additional knowledge and uses an attention mechanism to augment learning with expert knowledge. A similar approach for knowledge-induction was performed in the context of gradient-boosting (AAAI 2020). We have in addition developed interfaces for effective communication when learning to transfer across domains (for instance, when transferring a learned knowledge from baseball to finance).

# 3 Accomplishments

Our objectives were to:

- Create interactive agents for the collaborative building task in Minecraft that communicate successfully (give or execute instructions).
  To achieve this objective, we have
  a. Collected a human-human dialog dataset (509 conversations and game logs) for building collaborative agents in Minecraft (ACL 2019)
  b. Built a baseline neural Architect system that can generate a response given the current game context whenever it is triggered, and devised and performed human evaluation on our generation task. Human evaluation shows that this

system generates fully or partially correct instructions close to half the time (ACL 2019).

c. Built a neural Builder system that can execute instructions in their dialogue context. Even though this system has no access to any kind of prior domain knowledge (including spatial relations, names of structures, etc.), it shows a good understanding of instructions that involve create simple structures like rows or floating blocks. It also seems to understand concepts like a “gap” between blocks, and displays plausible interpretations of instructions such as “and now do the same on the other side” that refer both to complex action sequences that were executed earlier in the game (“the same”) and to spatial relations that depend not only on the current built structure, but also the builder’s current perspective (“the other side”)

d. Built a rule-based Builder system that can execute simple instructions and ask a limited variety of clarification questions. This frontend of this system is a rule based parser to convert natural language instructions into a predicate-logic format that can be interpreted by the planning module. This parser uses a list of regular expressions and templates. The planning (problem solving) module translates these predicate logic representations into executable instructions to the Minecraft agent, and identifies missing or conflicting information. A dialog manager controls the overall flow of the interaction and generates utterances (e.g. clarification questions to request missing information).

- Design a language comprehension module that converts natural language instructions into a formal representation that can be interpreted by the planner. Instructions may be given across multiple dialog turns, sharing context and mentioned entities across turns.
- Design and implement an annotation schema for the Illinois Minecraft dialogues that is capable of capturing spatial semantics and pragmatics and grounding it all to the quantified space. Annotations are used as training data for the AI Architect and Builder.
- Develop a Human-AI collaborative problem-solving architecture.
- Develop a learning framework that allows the agent to learn and assimilate new concepts via interaction with its human collaborator. To achieve this:
    a. We also worked on learning and using continuous representation of spatial relations. We looked at the following task - given two entities, predict the spatial relationship between these two entities in the context of an image. This is related to the task of predicting the location of one entity given another entity and the spatial relationship between them. We plan to develop a unified model to perform both tasks jointly.
    - We developed a framework that actively queries human expert for additional knowledge to improve efficiency and efficacy in hierarchical planning (AAMAS 2018). This framework initiates a 2-way dialog between human and agent, reduces burden on humans to provide all relevant knowledge upfront in CwC blocks world apparatus and proved some theoretical guarantees (Knowledge-Based Systems, 2019).
    - We developed a Human-AI collaboration framework that allows human experts to communicate additional knowledge in the form of inductive bias when learning

rich concepts (Frontiers in Robotics and AI 2020, CoDS-COMAD 2020, AAAI 2020).
    - Collaborative systems need to faithfully model uncertainty and structure for explainability. The downside of such powerful representation is the computational complexity of counting satisfied instances. We have developed approximation techniques for efficient counting in human machine communication (AAAI 2019 & ProbProg 2018).
- Develop a deeper understanding of the ability of existing systems to understand spatial relations and approaches to improving this using a richer representation and better modeling.
    - We developed a more expressive representation for spatial relationships called spatial configuration. We showed the alignment of elements of the spatial configuration with the new roles introduced in spatial AMR. Published at LREC 2020.
    - We developed a multi-modal model to understand implicit and explicit spatial relationships in language. We showed that pre-trained language models can give a strong prior to predict spatial relationships between objects, but by itself is not sufficient. We showed how combining image representations with the contextual representation of words can outperform either approach in isolation, even in the zero-shot setting. Published at LREC 2020.
    - We demonstrated that instruction following systems are very sensitive to perturbations in either modalities: language or world state. We characterized the expectations from an instruction following system and designed perturbations to show existing systems fail on these expectations. We then proposed an adversarial data augmentation framework that improves model performance on the perturbed examples, while preserving the standard performance on the original test set. Published at NAACL-HLT 2021
    - To better understand the challenge of spatial relations, we closely investigated the Blocks World dataset (Bisk et al, NAACL-HLT 2016). We showed that curriculum driven data augmentation and knowledge-free auxiliary tasks can help improve model performance. Submitted to ACL 2021.

# Training Opportunities

- At Illinois, Computer Science PhD students Anjali Narayan-Chen and Prashant Jayannavar have been fully supported by this grant. They have both become experts in grounded dialogue,  and neural models for generation and execution. They presented our papers at ACL 2019 and 2020. Anjali Narayan-Chen has now graduated.
- At the Univeristy of Pennsylvania, Soham Dan, a CIS PhD student has been supported by this grant and has learnt about several important topics under the umbrella of human-computer interaction - formal spatial semantics, continuous representation of relations, semantic parsing and extended semantic role labeling.

- At Colorado, Julia Bonn, MA LING, has become an expert in the linguistics literature that deals with spatial relations, and also on related work in the NLP community such as ISO Space-ML and the Tulane spatial relations annotation project. She brings her expertise in AMR annotation and multi-sentence AMR annotation to the project, and has been training Kristin Wright-Bettner in those skills, as well as Michael Regan. She is now a Linguistics PhD student at Colorado. Julia and Ghazaleh were trained on VerbNet by Susan Brown, Research faculty at Colorado. Jon Cai, a CS PhD student, was trained in neural net techniques for AMR parsing, with guidance from James Gung, a senior CS PhD student. James Gung graduated in May 2020, and developed the VerbNet parser. Kevin Stowe and Ghazaleh Kazeminejad contributed to the development of the Spatial Relations Ontology, collaborating with Julia Bonn. Kevin Stowe graduated July, 2019.
- At UTD and WSU, Graduate (PhD) students Mayukh Das (UTD), Harsha Kokel (UTD) and Rakibul Islam (WSU) have immensely benefitted from this grant on improving their fundamentals on human-machine collaborative problem solving. They have significantly improved their understanding of the communication issues when collaboratively solving planning tasks. Mayukh's entire PhD was funded from this grant and his thesis is fully supported by the findings from this grant (graduated 2019 Fall). A majority of Rakib's thesis has also been funded from this grant. The grant helped kickstart Harsha's thesis and she has already a AAAI paper from this grant.

# 4 Publications


Bonial, C., Brown, S. and Palmer, M. (2016, May) A Lexically-Informed Upper Level Event Ontology, ISA-12, In the Interoperability of Semantic Annotation (ISA-12), with LREC 2016, Portorož, Slovenia.

Bonn, J., Palmer, M., Cai, Z., and Wright-Bettner, K. (2020). Spatial AMR: Expanded Spatial Annotation in the Context of a Grounded Minecraft Corpus. In Proceedings of the 12th Language Resources and Evaluation Conference (LREC) pp. 4883--4892.

Bonn, J., Palmer, M., Cai, Z., and Wright-Bettner, K. (2021) Spatial AMR: extended annotation for spatially grounded corpora. *Language Resources and Evaluation.* Submitted for Publication. (in submission)

Brown, S. W., Pustejovsky, J., Zaenen, A., and Palmer, M. (2018, May) Integrating Generative Lexicon Event Structures into VerbNet. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation* (LREC 2018), Miyazaki, Japan.

Brown, S. W., Bonn, J., Gung, J., Zaenen, A., Pustejovsky, J. and Palmer, M., (2019, August) VerbNet Representations: Subevent Semantics for Transfer Verbs. In the *Proceedings of the 1st Designing Meaning Representations Workshop, DMR-2019*, held with ACL, Florence, Italy.

Cai, Z., Bonn, J., Martin, J.H., and Palmer, M. (2021). From AMR to Modal Logic: a Graph Pattern Solution. The 14th International Conference on Computational Semantics (IWCS). (in submission)

Dan, S., Kordjamshidi, P., Bonn, J., Bhatia, A., Cai, Z., Palmer, M., and Roth, D., (2020). From Spatial Relations to Spatial Configurations. In Proceedings of the 12th Language Resources and Evaluation Conference (LREC) pp. 5855--5864.

Dan, S., He, H., & Roth, D. (2020, May). Understanding Spatial Relations through Multiple Modalities. In Proceedings of The 12th Language Resources and Evaluation Conference (LREC) (pp. 2368-2372).

Dan, S., Zhou, M., Roth, D., Generalization in Instruction Following Systems, Accepted to NAACL-HLT 2021

Dan,S., Han, X., Roth,D. Compositional Data and Task Augmentation for Instruction Following. (in submission to ACL 2021)

Das, M. Human-Allied Efficient and Effective Learning in Noisy Domains, PhD Dissertation, UT Dallas, December 2019

Das, M., Dhami, D., Yu, Y., Kunapuli, G., & Natarajan, S. Human-Guided Learning of Column Networks: Knowledge Injection for Relational Deep Learning. COMAD/CODS 2021: 110-118

Das, M., Dhami, D.S., Yu, Y., Kunapuli, G., & Natarajan, S., *Knowledge-augmented Column Networks: Guiding Deep Learning with Advice*, Workshop on Human In the Loop Learning (HILL) 2019.

Das, M., Dhami, D.S., Kunapuli, G., Kersting, K., & Natarajan, S., *Fast Relational Probabilistic Inference and Learning Approximate Counting via Hypergraphs*, 33rd AAAI Conference on Artificial Intelligence (AAAI) 2019.

Das, M., & Ramanan, N., & Doppa, J.R., & Natarajan, S., One-Shot Induction of Generalized Logical Concepts via Human Guidance, Workshop on Statistical Relational AI (StarAI) 2020.

Das, M., Islam, M.R., Doppa, J.R., Roth, D., & Natarajan, S., *Active Preference Elicitation for Planning*, Human-Machine Collaborative Learning (@ AAAI) 2017.

Das, M., Odom, P., Islam, M.R., Doppa, J., Roth, D., & Natarajan, S., *Planning with actively eliciting preferences*, Knowledge-Based Systems 2018.

Das, M., Odom, P., Islam, M.R., Doppa, J., Roth, D., & Natarajan, S., *Preference- Guided Planning: An Active Elicitation Approach* , International Conference on Autonomous Agents and Multiagent Systems (AAMAS) 2018.

Das M, Ramanan N, Doppa JR, Natarajan S. Few-Shot Induction of Generalized Logical Concepts via Human Guidance. Front Robot AI. 2020;7:122.

Das, M., Wu, Y., Khot, T., Kersting, K., & Natarajan, S., *Scaling Lifted Probabilistic Inference and Learning Via Graph Databases*, SIAM International Conference on Data Mining (SDM) 2016.

Aryan Deshwal, Janardhan Rao Doppa, and Dan Roth, Learning and Inference for Structured Prediction: A Unifying Perspective in Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI), 2019

Green, M., Hargraves, O., Bonial, C., Chen, J., Clark, L., Palmer, M. (2017, June) VerbNet/OntoNotes-based sense annotation. In *The Handbook of Linguistic Annotations.* Ed. Pustejovsky, J. and Nancy Ide. Springer.

Hayes, A.L., Das, M., Odom, P., & Natarajan, S., *User Friendly Automatic Construction of Background Knowledge: Mode Construction from ER Diagrams*, Knowledge Capture Conference 2017.

Jayannavar, P., Narayan-Chen, A., & Hockenmaier, J. (2020, July). Learning to execute instructions in a Minecraft dialogue. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (pp. 2589-2602).

Li, T., Jawale, P.A., Palmer, M. and Srikumar, V. (2020, July) Structured Tuning for Semantic Role Labeling, In the *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (ACL 2020), virtual.

Chao Ma*, F A Rezaur Rahman Chowdhury*, Aryan Deshwal, Md Rakibul Islam, Janardhan Rao Doppa, and Dan Roth (*denotes equal contribution) Randomized Greedy Search for Structured Prediction: Amortized Inference and Learning in Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI), 2019

Kokel, H., Odom, P., Yang, S., & Natarajan, S. (2020). A Unified Framework for Knowledge Intensive Gradient Boosting: Leveraging Human Experts for Noisy Sparse Domains. Proceedings of the AAAI Conference on Artificial Intelligence, 34(04), 4460-4468.

Kumaraswamy, R., Ramanan, N., Odom, P. & Natarajan, S. Interactive Transfer Learning in Relational Domains. Künstl Intell 34, 181–192 (2020)

Narayan-Chen, A., Jayannavar, P., & Hockenmaier, J. (2019, July). Collaborative Dialogue in Minecraft. In Proceedings of the 57th Conference of the Association for Computational Linguistics (pp. 5405-5415).

Narayan-Chen, A. Towards collaborative dialogue in Minecraft, PhD Dissertation, University of Illinois at Urbana-Champaign, December 2020.

Narayan-Chen, A, Graber, C., Das, M., Islam, M.R., Dan S., Natarajan S., Doppa J., Hockenmaier J., Palmer M., Roth D. Towards Problem Solving Agents that Communicate and Learn Proceedings of the First Workshop on Language Grounding for Robotics (RoboNLP), 2017

Peterson, D.W., Palmer, M. (2018, February) Bayesian Verb Sense Clustering, In the *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18),* New Orleans.

Pustejovsky, J., Palmer, M., Zaenen, A., and Brown, S.. Integrating VerbNet and GL Predicative Structures. Interoperability for Semantic Annotation (ISA-12) 2016 with LREC-2016, Portoroz, Slovenia. May 23-28, 2016.

Stowe, K. and Palmer, M. (2018, June) Leveraging syntactic constructions for metaphor identification. In *Proceedings of the Workshop on Figurative Language Processing*, held with NAACL 2018, New Orleans, Louisiana.

Stowe, K., Moeller, S., Michaelis, L, Palmer, M., (2019, Nov) Linguistic Analysis Improves Neural Metaphor Detection, *In the Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL),* held with EMNLP, Hong Kong.

# Invited talks that referenced CwC:

Martha Palmer
*The Blocks World Redux*
Invited Keynote, Combined Workshop on Spatial Language Understanding and Grounded Communication for Robotics, June 6, 2019
Natural Language Seminar, U of Southern California - ISI - NL Seminar, May 16, 2019.
Linguistics Colloquium, University of New Mexico, March 15, 2019
CS Colloquium, University of Colorado, December 6, 2018
Artificial Intelligence Forum, University of Texas, November 26, 2018.
CSE Graduate Seminar, Texas A&M University, November 28, 2018.

Julia Hockenmaier
Stony Brook University, September 2019
LTI Colloquium, Carnegie Mellon University, April 2020
Invited Keynote, ICML Workshop on Learning in Open Worlds, 2020
Invited Keynote, EMNLP Workshop on Spatial Language Understanding, 2020
Department of Linguistics, Gothenburg University, March 2021

Sriraam Natarajan
Electrical Engineering and Computer Science Colloquium - Oregon State University, 2017.
Machine Learning Seminar, University of Texas at Austin, 2018
CS Colloquium, University of British Columbia, 2018
CS Colloquium, Washington State University, 2018
Invited Speaker, CoDS-COMAD 2019

Key Note Speaker, AI for Social Good, 2019