REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
31-03-2021	Final Report	31-Jul-2015 - 31-Dec-2020

4. TITLE AND SUBTITLE
Final Report: Deep Models of Compositionality and Context

5a. CONTRACT NUMBER
W911NF-15-1-0462

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
Stanford University
3160 Porter Drive
Suite 100
Stanford, CA 94304 -8445

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)
67831-CS-DRP.29

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			Percy Liang
UU	UU	UU	UU		19b. TELEPHONE NUMBER 650-723-7683

RPPR Final Report
as of 07-Apr-2021

Agency Code: 21XD

Proposal Number: 67831CSDRP **Agreement Number: W911NF-15-1-0462**
INVESTIGATOR(S):

Name: Percy Liang
Email: pliang@cs.stanford.edu
Phone Number: 6507237683
Principal: Y

Organization: **Stanford University**
Address: 3160 Porter Drive, Stanford, CA 943048445
Country: USA

DUNS Number: 009214214 EIN: 941156365
Report Date: 31-Jan-2021 Date Received: 31-Mar-2021
Final Report for Period Beginning 31-Jul-2015 and Ending 31-Dec-2020
Title: Deep Models of Compositionality and Context
Begin Performance Period: 31-Jul-2015 **End Performance Period:** 31-Dec-2020
Report Term: 0-Other

Submitted By: Percy Liang Email: pliang@cs.stanford.edu
Phone: (650) 723-7683

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

STEM Degrees: **STEM Participants:**

Major Goals: Our main research goal is to develop techniques that allow humans to communicate with computers effectively to perform complex tasks. Primarily, this involves:

1. Create AI systems which allow humans to intuitively collaborate with diverse and context-aware text generation systems. Milestone: Build a system allowing humans to co-create short stories with a text generation system. [Completed]
2. Improve the capabilities of AI systems at adapting to new language and scenarios specified by humans. Milestone: Build a system that lets users use natural language and define new concepts to perform complex tasks in a Blocks World environment. [Completed]

See uploaded report addendum for more details.

Accomplishments: See uploaded report addendum for more details.

Training Opportunities: This program provided training opportunities for numerous undergraduates, graduate students, and post-doctoral researchers (see "Participants" for full list) on a variety of technical topics.

Students and postdocs had the opportunity to present their work at conferences such as NAACL, ACL, NeurIPS, EMNLP, CVPR, AAI, SIGNLL, and CoNLL, as well as numerous workshops.

Results Dissemination: Nothing to Report

Honors and Awards: Nothing to Report

Protocol Activity Status:

Technology Transfer: Nothing to Report

PARTICIPANTS:

Participant Type: Graduate Student (research assistant)
Participant: Dor Arad

Person Months Worked: 13.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

Participant Type: Graduate Student (research assistant)
Participant: Anusha Balakrishnan
Person Months Worked: 10.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

Participant Type: Graduate Student (research assistant)
Participant: Nikhil Xie Bhattasali
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: Christopher James Cundy
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: Nathan Dass
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: Bogac Kerem Goksel
Person Months Worked: 8.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

Participant Type: Graduate Student (research assistant)

RPPR Final Report
as of 07-Apr-2021

Participant: Kelvin Gu
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: Yi Peng He
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: Peter Miller Henderson
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: John William Hewitt
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: Dan Iter
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: Woncheol Jeong
Person Months Worked: 2.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

RPPR Final Report
as of 07-Apr-2021

Participant Type: Graduate Student (research assistant)
Participant: Robin Jia
Person Months Worked: 8.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

Participant Type: Graduate Student (research assistant)
Participant: Pratyusha Kalluri
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: Michael Robert Kayser
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: Urvashi Khandelwal
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: Fereshte Khani
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: Dae Hyun Kim
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: Pang Wei Koh
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: John Dietrich Lawson
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: Mina Lee
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: Konstantin Lopyrev
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: Yiwei Luo
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: Isabel Victoria Papadimitrou
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: Ashwin Pradeep Paranjape
Person Months Worked: 7.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

Participant Type: Graduate Student (research assistant)
Participant: Panupong Pasupat
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: Reid Pryzant
Person Months Worked: 10.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

Participant Type: Graduate Student (research assistant)
Participant: Peng Qi
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: Aditi Raghunathan
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: Pranav Samir Rajpurkar
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: Shiori Sagawa
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: Abigail Elizabeth See
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: Kai Sheng Tai
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: Alexander Tamkin
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: Pratiksha Ranjit Thaker
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: Sida Wang
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: Hao Wu
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: Jian Zhang
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: Yuhui Zhang
Person Months Worked: 2.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

Participant Type: Other Professional
Participant: Jason Elliot Bolton
Person Months Worked: 15.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)
Participant: Christopher James Donahue
Person Months Worked: 13.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)
Participant: Tatsunori Benjamin Hashimoto
Person Months Worked: 4.00 **Funding Support:**
Project Contribution:
International Collaboration:

International Travel:
National Academy Member: N
Other Collaborators:

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)
Participant: He He
Person Months Worked: 14.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

Participant Type: Undergraduate Student
Participant: Alexander Uhunoma Iyabor
Person Months Worked: 2.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

Participant Type: PD/PI
Participant: Percy Shuo Liang
Person Months Worked: 4.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

Participant Type: Co PD/PI
Participant: Daniel S Jurafsky
Person Months Worked: 4.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

Participant Type: Co PD/PI
Participant: Christopher David Manning
Person Months Worked: 5.00 **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:** 5-Submitted
Journal: arXiv
Publication Identifier Type: Other Publication Identifier: arXiv:1905.08836v1
Volume: Issue: First Page #:
Date Submitted: 8/30/19 12:00AM Date Published:
Publication Location:
Article Title: Sample Efficient Text Summarization Using a Single Pre-Trained Transformer
Authors: Urvashi Khandelwal, Kevin Clark, Dan Jurafsky, ?ukasz Kaiser
Keywords: Language Model
Abstract: Language model (LM) pre-training has resulted in impressive performance and sample efficiency on a variety of language understanding tasks. However, it remains unclear how to best use pre-trained LMs for generation tasks such as abstractive summarization, particularly to enhance sample efficiency. In these sequence-to-sequence settings, prior work has experimented with loading pre-trained weights into the encoder and/or decoder networks, but used non-pre-trained encoder-decoder attention weights. We instead use a pre-trained decoder-only network, where the same Transformer LM both encodes the source and generates the summary. This ensures that all parameters in the network, including those governing attention over source states, have been pre-trained before the fine-tuning step. Experiments on the CNN/Daily Mail dataset show that our pre-trained Transformer LM substantially improves over pre-trained Transformer encoder-decoder networks in limited-data settings.
Distribution Statement: 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support: **Y**

CONFERENCE PAPERS:

Publication Type: Conference Paper or Presentation **Publication Status:** 5-Submitted
Conference Name: Association for Computational Linguistics (ACL)
Date Received: 01-Sep-2016 Conference Date: 11-Aug-2016 Date Published: 11-Aug-2016
Conference Location: Berlin, Germany
Paper Title: Learning language games through interaction
Authors: Sida I. Wang, Percy Liang, Chris Manning
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 5-Submitted
Conference Name: Association for Computational Linguistics (ACL)
Date Received: 01-Sep-2016 Conference Date: 08-Aug-2016 Date Published: 08-Aug-2016
Conference Location: Berlin, Germany
Paper Title: Data recombination for neural semantic parsing
Authors: Robin Jia, Percy Liang
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 5-Submitted
Conference Name: Association for Computational Linguistics (ACL)
Date Received: 01-Sep-2016 Conference Date: 08-Aug-2016 Date Published: 08-Aug-2016
Conference Location: Berlin, Germany
Paper Title: A Persona-Based Neural Conversation Model
Authors: Jiwei Li, Michel Galley, Chris Brockett, Jianfeng Gao, Bill Dolan
Acknowledged Federal Support: **N**

RPPR Final Report
as of 07-Apr-2021

Publication Type: Conference Paper or Presentation **Publication Status:** 0-Other
Conference Name: North American Chapter of the Association for Computational Linguistics: Human Language Technologies
Date Received: 01-Sep-2016 Conference Date: 16-Jun-2016 Date Published: 16-Jun-2016
Conference Location: San Diego, CA
Paper Title: A Diversity-Promoting Objective Function for Neural Conversation Models
Authors: Jiwei Li, Michel Galley, Chris Brockett, Jianfeng Gao, Bill Dolan
Acknowledged Federal Support: **N**

Publication Type: Conference Paper or Presentation **Publication Status:** 5-Submitted
Conference Name: Association for Computational Linguistics 2017
Date Received: 31-Aug-2017 Conference Date: 30-Jul-2017 Date Published:
Conference Location: Vancouver, Canada
Paper Title: Naturalizing a Programming Language via Interactive Learning
Authors: Sida I. Wang, Samuel Ginn, Percy Liang, Christoper D. Manning
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 5-Submitted
Conference Name: Association for Computational Linguistics 2017
Date Received: 31-Aug-2017 Conference Date: 30-Jul-2017 Date Published:
Conference Location: Vancouver, Canada
Paper Title: Learning Symmetric Collaborative Dialogue Agents with Dynamic Knowledge Graph Embeddings
Authors: He He, Anusha Balakrishnan, Mihail Eric, Percy Liang
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 0-Other
Conference Name: NIPS 2018
Date Received: 30-Aug-2019 Conference Date: 02-Dec-2018 Date Published:
Conference Location: Montreal, Canada
Paper Title: A Retrieve-and-Edit Framework for Predicting Structured Outputs
Authors: Tatsunori B. Hashimoto, Kelvin Guu, Yonaten Oren, Percy Liang
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 0-Other
Conference Name: CVPR 2019
Date Received: 30-Aug-2019 Conference Date: 16-Jun-2019 Date Published:
Conference Location: Long Beach, CA
Paper Title: GQA: A New Dataset for Real-World Visual Reasoning and Compositional Question Answering
Authors: Drew A. Hudson, Chris Maning
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 0-Other
Conference Name: NAACL 2019
Date Received: 30-Aug-2019 Conference Date: 02-Jun-2019 Date Published:
Conference Location: Minneapolis, MN
Paper Title: Generating Puns with Surprise
Authors: He He, Nanyun Peng, Percy Liang
Acknowledged Federal Support: **Y**

RPPR Final Report
as of 07-Apr-2021

Publication Type: Conference Paper or Presentation **Publication Status:** 0-Other
Conference Name: NAACL 2019
Date Received: 30-Aug-2019 Conference Date: 02-Jun-2019 Date Published:
Conference Location: Minneapolis, MN
Paper Title: Unifying Human and Statistical Evaluation for Text Generation
Authors: Hugh Zhang, Tatsunori B. Hashimoto, Percy Liang
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 0-Other
Conference Name: EMNLP 2019
Date Received: 30-Aug-2019 Conference Date: 03-Nov-2019 Date Published:
Conference Location: Hong Kong, China
Paper Title: Distributionally Robust Language Modeling
Authors: Tatsunori B. Hashimoto, Shiori Sagawa, Yonaten Oren, Kelvin Guu, Percy Liang
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 3-Accepted
Conference Name: CoNLL 2019
Date Received: 30-Aug-2019 Conference Date: 03-Nov-2019 Date Published:
Conference Location: Hong Kong, China
Paper Title: Do Massively Pretrained Language Models Make Better Storytellers?
Authors: Abigail See, Aneesh Pappu, Rohun Saxena and Akhila Yerukola
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 5-Submitted
Conference Name: AAAI Conference on Artificial Intelligence
Date Received: 10-Sep-2020 Conference Date: 07-Feb-2020 Date Published: 01-Jan-2001
Conference Location: New York, NY
Paper Title: Automatically neutralizing subjective bias in text
Authors: Reid Pryzant, Richard Martinez, Nathan Dass, Sadao Kurohashi, Dan Jurafsky, Diyi Yang
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 5-Submitted
Conference Name: Association for Computational Linguistics
Date Received: 10-Sep-2020 Conference Date: 05-Jul-2020 Date Published:
Conference Location: Virtual
Paper Title: Enabling Language Models to Fill in the Blanks
Authors: Chris Donahue, Mina Lee, Percy Liang
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 5-Submitted
Conference Name: Association for Computational Linguistics
Date Received: 10-Sep-2020 Conference Date: 05-Jul-2020 Date Published:
Conference Location: Virtual
Paper Title: Pretraining with Contrastive Sentence Objectives Improves Discourse Performance of Language Models
Authors: Dan Iter, Kelvin Guu, Larry Lansing, Dan Jurafsky
Acknowledged Federal Support: **Y**

RPPR Final Report
as of 07-Apr-2021

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: TACL 2015
Date Received: Conference Date: 21-Jun-2015 Date Published:
Conference Location: Ischia Island (Italy)
Paper Title: Imitation learning of agenda-based semantic parsers
Authors: Jonathan Berant, Percy Liang
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 0-Other
Conference Name: EMNLP 2016
Date Received: Conference Date: 01-Nov-2016 Date Published:
Conference Location: Austin, Texas
Paper Title: Deep Reinforcement Learning for Dialogue Generation
Authors: Jiwei Li, Will Monroe, Alan Ritter, Michel Galley, Jianfeng Gao, Dan Jurafsky
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 0-Other
Conference Name: SIGNLL 2016
Date Received: Conference Date: 11-Aug-2016 Date Published:
Conference Location: Berlin, Germany
Paper Title: Compression of Neural Machine Translation Models via Pruning
Authors: Abigail See, Minh-Thang Luong, Christopher D Manning
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: ACL 2017
Date Received: Conference Date: 30-Jul-2017 Date Published:
Conference Location: Vancouver, CA
Paper Title: Naturalizing a programming language via interactive learning
Authors: Sida I Wang, Samuel Ginn, Percy Liang, Christopher D Manning
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: EMNLP 2017
Date Received: Conference Date: 07-Sep-2017 Date Published:
Conference Location: Copenhagen, Denmark
Paper Title: Neural Net Models of Open-domain Discourse Coherence
Authors: Jiwei Li, Dan Jurafsky
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: EMNLP 2017
Date Received: Conference Date: 07-Sep-2017 Date Published:
Conference Location: Copenhagen, Denmark
Paper Title: Adversarial Learning for Neural Dialogue Generation
Authors: Jiwei Li, Will Monroe, Tianlin Shi, Sébastien Jean, Alan Ritter, Dan Jurafsky
Acknowledged Federal Support: **Y**

RPPR Final Report
as of 07-Apr-2021

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: ACL 2018
Date Received: Conference Date: 15-Jul-2018 Date Published:
Conference Location: Melbourne, Australia
Paper Title: Generating Sentences by Editing Prototypes
Authors: Kelvin Guu, Tatsunori Hashimoto, Yonatan Oren, Percy Liang
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: ACL 2018
Date Received: Conference Date: 15-Jul-2018 Date Published:
Conference Location: Melbourne, Australia
Paper Title: Sharp Nearby, Fuzzy Far Away: How Neural Language Models Use Context
Authors: Urvashi Khandelwal, He He, Peng Qi, Dan Jurafsky
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: NAACL 2018
Date Received: Conference Date: 02-Jun-2018 Date Published:
Conference Location: New Orleans, LA
Paper Title: Delete, Retrieve, Generate: A Simple Approach to Sentiment and Style Transfer
Authors: Juncen Li, Robin Jia, He He, Percy Liang
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: EMNLP 2018
Date Received: Conference Date: 02-Nov-2018 Date Published:
Conference Location: Brussels, Belgium
Paper Title: Decoupling Strategy and Generation in Negotiation Dialogues
Authors: He He, Derek Chen, Anusha Balakrishnan, Percy Liang
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: EMNLP 2019
Date Received: Conference Date: 03-Nov-2019 Date Published:
Conference Location: Hong Kong, China
Paper Title: Integrating Text and Image: Determining Multimodal Document Intent in Instagram Posts
Authors: Julia Kruk, Jonah Lubin, Karan Sikka, Xiao Lin, Dan Jurafsky, Ajay Divakaran
Acknowledged Federal Support: **Y**

DARPA CwC Final Report Addendum

Stanford University

March 2021

The DARPA Communicating with Computers (CwC) program enabled numerous research contributions from our team at Stanford University. Our earlier contributions were motivated primarily by the *Blocks World* use case, and later contributions by the *Collaborative Composition* use case. This document provides technical details of our primary accomplishments under each use case as well as how these accomplishments contributed to the goals of the CwC program. Additionally, we highlight several other contributions and artifacts that were enabled by this program.

1 Collaborative Composition: WRITING WITH AI

Since 2018, our research efforts focused on the Collaborative Composition use case. The culmination of these efforts was WRITING WITH AI, a system enabling seamless human-computer co-creation for creative writing of short stories, which we will describe in detail below. The WRITING WITH AI system builds on insights from numerous contributions made by our CwC team to the related tasks of editing human text [Guu et al., 2018, Hashimoto et al., 2018, He et al., 2019, Pryzant et al., 2020] and text generation [Li et al., 2018, Khandelwal et al., 2018, Kruk et al., 2019, Hashimoto et al., 2019, Zhang et al., 2019, Iter et al., 2020].

1.1 WRITING WITH AI

WRITING WITH AI is a system which assists humans in creative writing for short stories by providing dynamic suggestions to a user based on their current focus (Figure 1). The design of this system was motivated by the observation that existing tools for story co-creation required humans to learn a new and unfamiliar format for creative writing, e.g., taking turns with a computer to contribute individual sentences. We felt that forcing users to learn an entirely new paradigm was

Figure 1: Screenshot of the WRITING WITH AI interface.

at tension with several CwC hallmarks: that humans should be able to communicate comfortably (H9), learn to use the system quickly (H72), and that the system should not distract users (H32). Hence, we aspired to build a system which dynamically offered suggestions to users as they typed, allowing users to have help available when they need it but otherwise not impeding them.

Guided by this principle, we designed an interface **allowing for intuitive and unobtrusive co-creation**: a simple text box and a list of suggestions which autopopulate as the user types. Additionally, to provide an initial spark of inspiration, we provide the user with a simple prompt (e.g. "The moon is haunted"), similar to the model popularized by the /R/WRITINGPROMPTS community.[1]. To increase relevance to a user's current focus, the system provides suggestions based on the location of the user's cursor. The suggestions themselves come from two underlying language models, one which generates new text and one which paraphrases existing text, which will be described in more detail below.

A major goal of WRITING WITH AI is to be able to **support non-linear writing patterns of writing**. Humans often skip around when they write stories, updating earlier parts of the story based on later parts. However, most text generation models do not facilitate such workflows as they can only generate text at the end of

[1] https://reddit.com/r/WritingPrompts

a story, forcing humans to write in a rigid left-to-right manner. To allow for more flexibility, WRITING WITH AI builds on our previous research on infilling [Donahue et al., 2020], allowing for variable-length text generation at any user-specified point(s) in the story. This added flexibility mirrors CwC hallmarks regarding effective use of context (H55) and using all available information to respond (H53).

Another goal of our system is to **encourage more users to write stories by making the process more efficient and enjoyable**. Writing a short story is time consuming and difficult, and writers block can often . Users on communities like /R/WRITINGPROMPTS help others overcome these barriers by providing topical suggestions, and we believe that having a computer act as an interactive writing partner could make the process even more engaging (H17) and enjoyable (H28).

In addition to making the process more efficient and enjoyable, we also want our system to **enable users to write stories that they never could have written alone**. In the domain of creative writing, the unusual suggestions often associated with text generation models may be a feature rather than a bug, inspiring new ideas in humans (H34) and steering them to more interesting and creative results (H27).

To achieve these ambitious goals, we first needed to train a text generation system capable of providing high-quality suggestions to users at any point in their ongoing story. In Donahue et al. [2020], we proposed a strategy to add this functionality to existing text generation systems (which could only consider text before the point of generation), combining the quality of their generated text with the added flexibility of being able to "infill", i.e., generate text at any point in a document.

To get a sense of whether infilling would be interesting and intuitive to users in a creative writing context, we first built a proof-of-concept interface which allowed people to rewrite an existing story by selecting parts of it to replace with generated text (Figure 2).[2] We conducted a pilot study with labmates, and users seemed to enjoy the experience overall. However, in addition to complaints about how the interface was clunky and unnatural, a common complaint was that the system only allowed users to generate new text, rather than rewrite text they had already written.

To address this, we began researching strategies for paraphrasing text. We focused on the simplest case, replacing individual words in context with other appropriate words. In Lee et al. [2021], we collected a large dataset consisting of human judgements of the contextual acceptability of substitute words in particular contexts. We then fine-tuned a language model called BERT [Devlin et al., 2018] on this paraphrashing task. The resultant model was integrated alongside our infilling model in the WRITING WITH AI system, where users were presented with replacements for their currently-selected word in addition to generated sentence

[2] `https://chrisdonahue.com/ilm/`

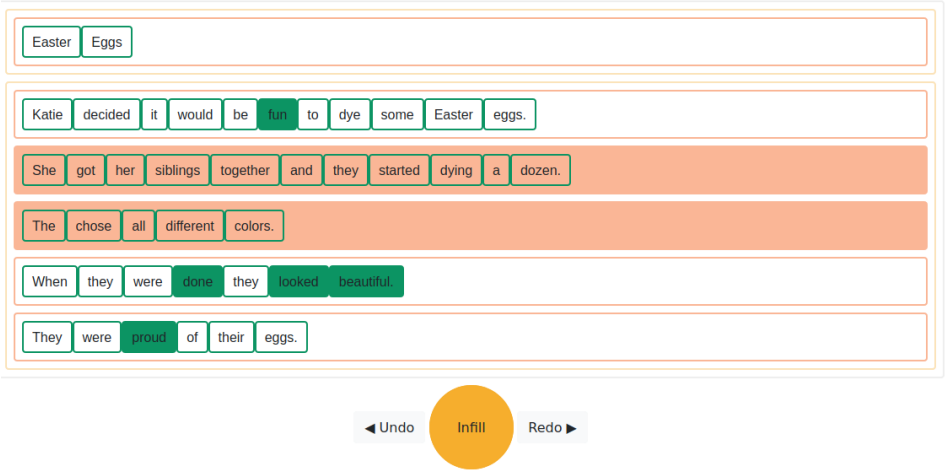

Figure 2: Screenshot of our initial interface which explored text infilling for co-creation of short stories.

completions from our infilling model (Figure 1).

To gauge the effectiveness of our system, we conducted a study with 100 participants on Amazon Mechanical Turk. Participants were required to interact with the system for at least five minutes, and were asked to write a story in response to one of four randomly-selected prompts:

1. (*Animal*) You awake one morning and find an unusual animal in your bed.
2. (*Book*) A librarian notices small changes in patrons who check out a particular book.
3. (*Rock*) You begin to suspect that your pet rock is actually alive.
4. (*Well*) You drop a coin down a well and hear someone say "Ow!"

After the study, users were asked to fill out a survey regarding the experience.

In Table 1, we show one example per prompt of stories co-created by users and our WRITING WITH AI system during the user study.[3] Overall, despite no explicit requirements to do so, $> 90\%$ of users added suggestions from our system to their story, and system suggestions comprised a total of 38% of the words present in final stories. While users only accepted about 4% of the total suggestions from

[3]We acknowledge and thank Dr. Cheryl Clark and the MITRE team for assistance with analyzing the logs from our WRITING WITH AI user study.

Prompt	Co-created story
Animal	"Mommy?" I asked gently, shivering slightly. The blanket starts shifting as I start to reach for the lamp. "Gross" I think to myself. Something wet is on my blanket and there is a small hump under my covers. "Maybe I need to just take out the trash or something" I muse. I turn the light on to see a little grey hunk of skin standing on my lap. "I need to talk to you" the grey mass says. I am flustered. "How are you talking?!" I reply. "There's no time for that. We need your help..."
Book	She wonders what is in the book that causes them to change. After some deep contemplation, she decides to read the book for herself. Right from the start of the first chapter, she's drawn into the story as if she were actually in it. Feeling a sense of fascination, as the sounds and smells come alive, she realizes that she actually IS in the story! "This is amazing!" she cried out. She began to investigate her surroundings exploring every corner and searching for every detail. She finds a rather odd looking rock and pushes it over, revealing a mysterious black hole in the ground. Unable to see into the hole, she leans into it to get a better look, and is suddenly sucked into the black hole! Everything fades to blackness and when she finally wakes up, she thinks to herself 'That's weird" as she finds herself back in the library.
Rock	"Oh wow! How is this even possible", I pondered before leaning back and letting my legs fall on the cold floor. I get up, pick up the rock and set it on a chair, "so do you feel hungry?, I don't know any rock food...", I said.
Well	I wonder what or who is down in that well. I definitely want my coin back and will go to any length to get it because I think someone in down there collecting the coins. I said to my friend "I will go and find it, I will stop at nothing until I get my coin back." I realize the well is gross as I reach out to grab the coin. My friend said that he wanted to help me. I responded and said, "Will you take my coin?"

Table 1: Examples of co-created stories by users from our Mechanical Turk study and WRITING WITH AI, one for each of the four prompts. Highlighted text was generated by our system verbatim, underlined text was generated by our system and then modified by users before adding to the final story.

the system, this is not cause for concern as our system was designed such that suggestions are passive and can be ignored when the user does not need assistance.

Analysis of the logs shows extensive evidence that users write stories in non-linear fashions, with users commonly backtracking to earlier parts of their story to edit or re-write text. In addition, more than 25% of the suggestions from our system that made it into the final stories were inserted in the middle of the story, suggesting that our system does indeed support such non-linear patterns.

From a qualitative perspective, more than half of users responded to our survey that the task was more enjoyable with the system than it would have been otherwise, and that the system was easy-to-use and intuitive. Moreover, 75% of users reported that the system helped them come up with new ideas that they could not have come up with alone. Finally, despite the participants of this survey being recruited from Mechanical Turk rather than communities focused on creative writing, most users stated they would like to engage with the system again. We found this to be encouraging evidence that our system was making the overall experience of writing short-stories more engaging, even for people who might have never done so otherwise.

In addition to positive feedback, users reported some drawbacks of the system. While a majority of users found the task more enjoyable with the system, less than half found the task to be more efficient. Additionally, users reported that generated text was sometimes grammatically incorrect or inconsistent with the surrounding text, which agrees with our anecdotal experience of using the system. Nevertheless, some users still found such suggestions beneficial, as they provided (in the words of one user) "words or ideas I found worth introducing to the story."

Overall, the results from our user study indicate that WRITING WITH AI helped humans perform the task of creative writing, contributing not only low-level generated text but also high-level ideas which shaped the overall narrative. In See et al. [2019], we reported that the type of large-scale language models used in WRITING WITH AI have numerous shortcomings preventing them from writing compelling stories by themselves. Hence, we believe that WRITING WITH AI is a concrete step towards one of the primary goals of the CwC program: systems where humans can collaborate with computers to overcome each other's shortcomings, accomplish a shared task more effectively than either could do alone.

2 Blocks World: SHRDLURN and VOXELURN

Before 2018, our research efforts focused on the Blocks World use case. In particular, we focused on building systems which were capable of adapting to the changing needs of human users. Human adaptation and accommodation are funda-

mental aspects of bidirectional communication. Even today, most natural language interfaces on computers or phones are trained once and deployed, and users must live with their limitations. Our work on interactive language games [Wang et al., 2016, 2017] demonstrates that humans can perform complex tasks cooperatively with machine learning systems only if the model is capable of adapting to new language and use cases specified by humans. Both of these works rely on a technique called *semantic parsing* [Zettlemoyer and Collins, 2005, Liang and Potts, 2015] which converts natural language into formal language; the CwC program also enabled our team to discover core improvements to semantic parsing [Berant and Liang, 2015, Jia and Liang, 2016], hopefully leading to improved collaborative systems downstream.

2.1 SHRDLURN

The first system we developed as part of this program was SHRDLURN [Wang et al., 2016].[4] In SHRDLURN, humans specify commands in natural language to a computer system regarding how to complete a complex task in a Blocks World environment. With this system, we endeavored to build a computer agent which could learn to complete complex tasks *from scratch*, i.e., with no pre-defined language. For computers to be able to communicate effectively with humans, it is important that they be capable of learning new language and instructions as specified on-the-fly by humans during an interaction.

As shown in Figure 3, the shared objective of the human and the computer is to transform a starting configuration of blocks into a goal configuration. The human knows the goal state but the computer does not, and the only action the human can take is entering a command in natural language. To facilitate this collaboration, SHRDLURN parses the user-specified natural language commands and produces a ranked list of possible interpretations according to its current model. The human scrolls through the list and chooses the intended one, simultaneously advancing the state of the blocks and providing feedback to the computer. Both the human and the computer wish to reach the goal state with as little scrolling as possible. For the computer to be successful, it has to learn the human's language quickly over the course of the game, so that the human can accomplish the goal more efficiently. Conversely, the human can also speed up progress by understanding and working around the limitations of the computer.

We model the computer as a semantic parser [Zettlemoyer and Collins, 2005, Liang and Potts, 2015], which maps natural language commands (e.g., 'remove red') into logical forms (e.g., REMOVE(WITH(RED))). The semantic parser has no

[4] `https://shrdlurn.sidaw.xyz/`

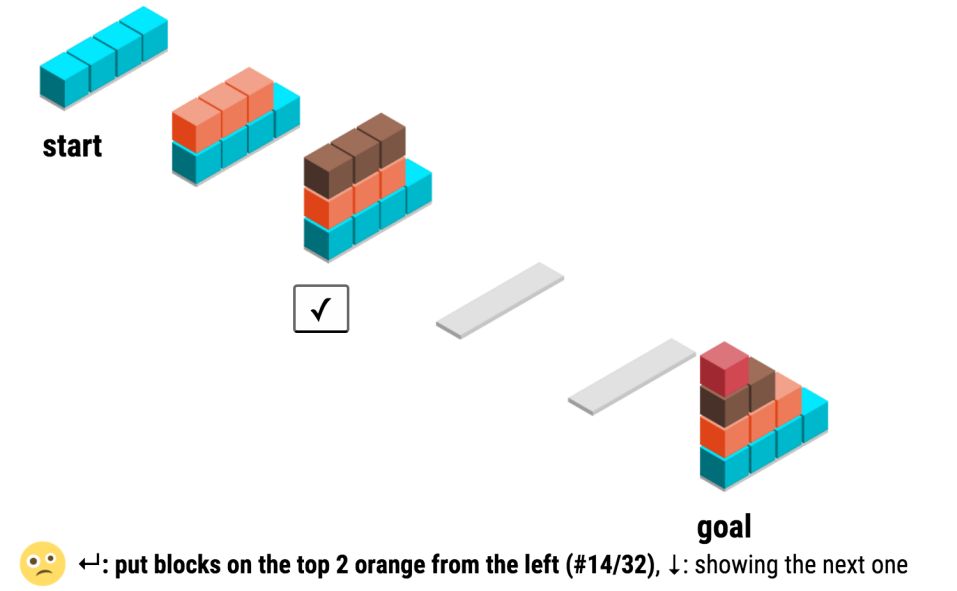

Figure 3: The SHRDLURN game: the objective is to transform the start state into the goal state. The human types in a natural language command, and the computer (which does not know the goal state) tries to interpret the command and perform the corresponding action. The computer initially knows nothing about the language, but through the human's feedback, learns the human's language while making progress towards the game goal.

seed lexicon and no annotated logical forms, so it just generates many candidate logical forms. From the human's feedback, it learn by adjusting the parameters corresponding to simple and generic lexical features. It is crucial that the computer learns quickly, or users are frustrated and the system is less usable.

To evaluate the effectiveness of the SHRDLURN system, we conducted a user study on Amazon Mechanical Turk (AMT). In this study, 100 workers piloted our system to complete the same task, requiring an average of 102 commands each to complete the task across an average time of around an hour. We were pleased to see that some users even found this simplistic setup to be engaging, as reflected in their optional comments:

- "That was probably the most fun thing I have ever done on mTurk."
- "Wow this was one mind bending games"

We measure system performance by lowest *number of scrolls*, i.e., we want the human to have to sift through as few computer interpretations as possible. Through inspection, we found that 22 players were unable to teach the computer an actual language, and instead finished the game primarily by scrolling. On average, these 22 players scrolled 21.6 scrolls per command versus only 7.4 for the others. In Figure 4, we show examples of the types of languages that humans naturally chose to pilot the computer system.

Overall, we find that many players adapt to the task by becoming more consistent, less verbose, and more precise. For example, some players became more consistent over time (e.g. from using both REMOVE and DISCARD to only using REMOVE). We were also encouraged to see that our system was flexible enough to be taught commands in several different languages. Different users were able to successfully pilot the system using English, Arabic, Polish, and a custom programming language. This flexibility is a rarity in natural language interfaces, and a key feature of our system

We believe these observations are central to the goals of the CwC program. In particular, we observed that our system allowed humans to both complete a complex task while *simultaneously* teaching the system new language and concepts. To be capable of adapting to new challenges, collaborative computer systems must learn new concepts through interaction with users and improve over time, and SHRDLURN accomplishes this difficult goal.

2.2 VOXELURN

In SHRDLURN, our language learning system started from scratch for each user engagement. However, while this system allowed humans to teach diverse sets of

Most successful players (1st–20th)		
rem cy pos 1, stack or blk pos 4, rem blk pos 2 thru 5, rem blk pos 2 thru 4, stack bn blk pos 1 thru 2, fill bn blk, stack or blk pos 2 thru 6, rem cy blk pos 2 fill rd blk **(3.01)**	remove the brown block, remove all orange blocks, put brown block on orange blocks, put orange blocks on all blocks, put blue block on leftmost blue block in top row **(2.78)**	Remove the center block, Remove the red block, Remove all red blocks, Remove the first orange block, Put a brown block on the first brown block, Add blue block on first blue block **(2.72)**
Average players (21th–50th)		
reinsert pink, take brown, put in pink, remove two pink from second layer, Add two red to second layer in odd intervals, Add five pink to second layer, Remove one blue and one brown from bottom layer **(9.17)**	remove red, remove 1 red, remove 2 4 orange, add 2 red, add 1 2 3 4 blue, emove 1 3 5 orange, add 2 4 orange, add 2 orange, remove 2 3 brown, add 1 2 3 4 5 red, remove 2 3 4 5 6, remove 2, add 1 2 3 4 6 red **(8.37)**	move second cube, double red with blue, double first red with red, triple second and fourth with orange, add red, remove orange on row two, add blue to column two, add brown on first and third **(7.18)**
Least successful players (51th–)		
holdleftmost, holdbrown, holdleftmost, blueonblue, brownonblue1, blueonorange, holdblue, holdorange2, blueonred2 , holdends1, holdrightend, hold2, orangeonorangerightmost **(14.15)**	'add red cubes on center left, center right, far left and far right', 'remove blue blocks on row two column two, row two column four', remove red blocks in center left and center right on second row **(12.6)**	laugh with me, red blocks with one aqua, aqua red alternate, brown red red orange aqua orange, red brown red brown red brown, space red orange red, second level red space red space red space **(14.32)**
Spam players ($\sim$ 85th–100)		
next, hello happy, how are you, move, gold, build goal blocks, 23,house, gabboli, x, run,,xav, d, j, xcv, dulicate goal (21.7)		
Most interesting		
usuń brązowe klocki, postaw pomarańczowy klocek na pierwszym klocku, postaw czerwone klocki na pomarańczowych, usuń pomarańczowe klocki w górnym rzędzie	rm scat + 1 c, + 1 c, rm sh, + 1 2 4 sh, + 1 c, - 4 o, rm 1 r, + 1 3 o, full fill c, rm o, full fill sh, - 1 3, full fill sh, rm sh, rm r, + 2 3 r, rm o, + 3 sh, + 2 3 sh, rm b, - 1 o, + 2 c,	mBROWN,mBLUE,mORANGE RED+ORANGEˆORANGE, BROWN+BROWNm1+BROWNm3, ORANGE +BROWN +ORANGEˆm1+ ORANGEˆm3 + BROWNˆˆ2 + BROWNˆˆ4

Figure 4: Example commands, along with the average number of scrolls for that player in parentheses. Success is measured by the number of scrolls, where the more successful players need less scrolls. 1) The 20 most successful players tend to use consistent and concise language whose semantics is similar to our logical language. 2) Average players tend to be slightly more verbose and inconsistent (left and right), or significantly different from our logical langauge (middle). 3) Reasons for being unsuccessful vary. Left: no tokenization, middle: used a coordinate system and many conjunctions; right: confused in the beginning, and used a language very different from our logical language.

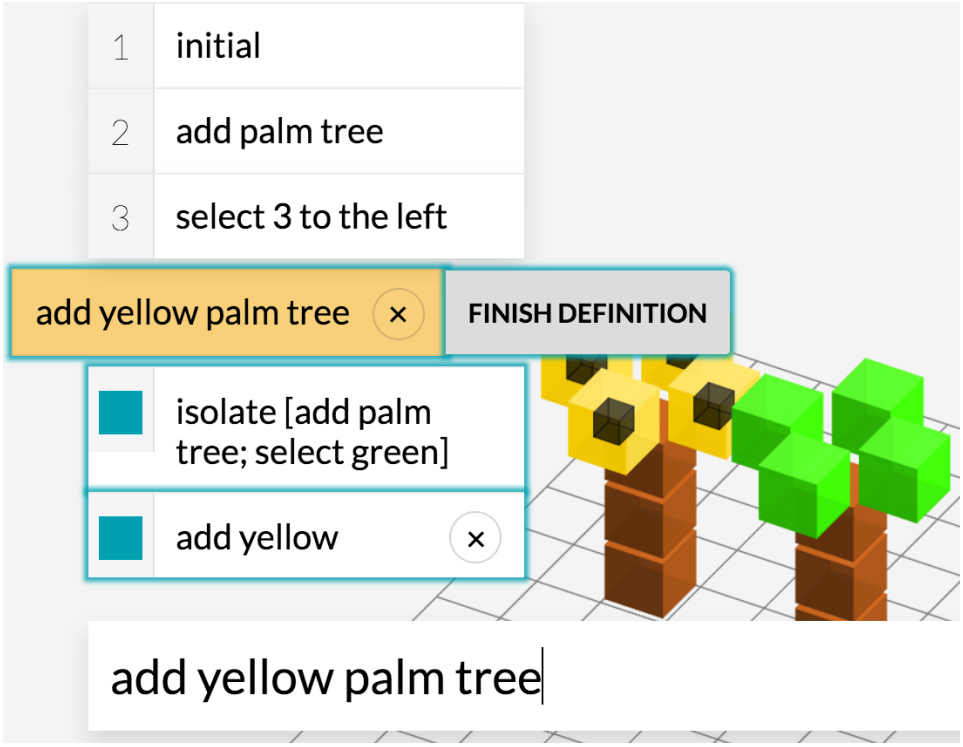

Figure 5: VOXELURN web interface, allowing users to build structures via natural language commands.

commands to the computer, we observed in practice that many users entered similar commands. Inspired by this, we aspired to build a similar system where users could build upon the teaching efforts of users before them, hopefully allowing humans to perform more complex tasks in less time.

In VOXELURN [Wang et al., 2017], users are presented with a similar interface to that of SHRDLURN: a text box which accepts language commands (Figure 5).[5] However, instead of attempting to parse natural language commands right off the bat and have users select the appropriate parse among several, we initially allow users access to a precise *programming language*. This programming language gives users precise and powerful control over their environment, but is more difficult to use than natural language commands (bottom of Figure 5). To get the best of both worlds, users can *naturalize* this programming language over time: attaching *definitions* to sequences of commands in the original programming lan-

[5] `http://local.voxelurn.com/#/build`

Figure 6: Examples of structures that participants in our Mechanical Turk user study created using VOXELURN.

guage. This allows users to connect interpretable natural language concepts to precise functionality.

Definitions equate a novel command to a sequence of commands that the system already understands. For example, GO LEFT 6 AND GO FRONT might be defined as REPEAT 6 [GO LEFT]; GO FRONT, which eventually can be traced back to the expression REPEAT 6 [SELECT LEFT OF THIS]; SELECT FRONT OF THIS in the core language. Unlike function definitions in programming languages, the user writes concrete values rather than explicitly declaring arguments. The system automatically extracts arguments and learns to produce the correct generalizations. Compared to standard machine learning, say from demonstrations, definitions provide a much more powerful learning signal: the system is told directly that A 3 BY 4 RED SQUARE is 3 RED COLUMNS OF HEIGHT 4, and does not have to infer how to generalize from observing many structures of different sizes.

To evaluate the strength of this setup, we conduct a user study on Mechanical Turk, where dozens of users simultaneously taught language to the model and could build on top of each other's abstractions. First, we had *qualifier* tasks, in which an AMT worker was instructed to replicate a fixed target exactly, ensuring that the initial users are familiar with at least some of the core language, which is the starting point of the naturalization process.

Next, we allowed the workers who qualified to enter a second open-ended task, in which they were asked to build any structure they wanted in 30 minutes. This process was designed to give users freedom while ensuring quality. The analogy of this scheme in a real system is that early users (or a small portion of expert users) have to make some learning investment, so the system can learn and become easier

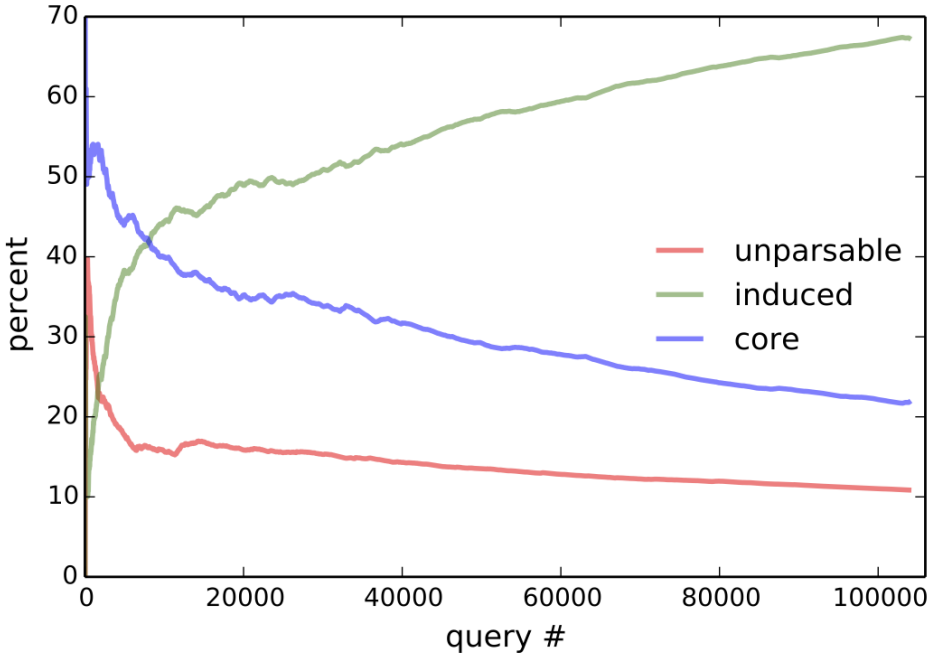

Figure 7: Percentage of all user-specified commands in our VOXELURN user study that are part of the *core* programming language, the *induced* language (more natural), or *unparsable* by the system.

for other users. 70 workers passed the qualifier task, and 42 workers participated in the final free-building experiment. Together, they built 230 structures, some of which were surprisingly ornate (Figure 6). In total, over 100k commands were entered consisting of over 5k distinct tokens.

Over the course of the experiment, we observed that the percentage of commands expressed by users in the core language decreased, while the percentage of commands expressed by users in the induced language increased (Figure 7). We interpret this as evidence that this naturalization process is both useful and effective, in that users willingly chose to write more commands in the induced language as time progressed.

With VOXELURN, we demonstrated a viable strategy for combining the flexibility and precision of programming languages with the ease-of-use of natural language. Using a mixture of these strategies, users were able to build complex structures, building upon the added natural abstractions of previous users. We believe that these principles behind VOXELURN are broadly applicable to computer

systems which might help humans accomplish many different kinds of complex tasks more effectively. Moreover, our observations during both SHRDLURN and VOXELURN reinforce the notion that computer systems must have the ability to adapt and learn from users in order to be effective in scenarios that were not envisioned beforehand.

3 Other

In addition to our primary efforts described above, our CwC team made numerous research contributions in the fields of conversational agents [Li et al., 2016b,a,c, Li and Jurafsky, 2017], dialogue systems [He et al., 2017, Li et al., 2017, He et al., 2018], and other topics in natural language processing [See et al., 2016, Hudson and Maning, 2019, Khandelwal et al., 2019].

References[6]

Jonathan Berant and Percy Liang. Imitation learning of agenda-based semantic parsers. *TACL*, 2015.

* Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. *NAACL*, 2018.

Chris Donahue, Mina Lee, and Percy Liang. Enabling language models to fill in the blanks. In *ACL*, 2020.

Kelvin Guu, Tatsunori Hashimoto, Yonatan Oren, and Percy Liang. Generating sentences by editing prototypes. In *ACL*, 2018.

Tatsunori B Hashimoto, Kelvin Guu, Yonaten Oren, and Percy Liang. A retrieve-and-edit framework for predicting structured outputs. In *NeurIPS*, 2018.

Tatsunori B Hashimoto, Shiori Sagawa, Yonaten Oren, Kelvin Guu, and Percy Liang. Distributionally robust language modeling. In *EMNLP*, 2019.

He He, Anusha Balakrishnan, Mihail Eric, and Percy Liang. Learning symmetric collaborative dialogue agents with dynamic knowledge graph embeddings. In *ACL*, 2017.

[6] All referenced papers were original research at Stanford as part of the DARPA CwC program (under ARO prime contract no. W911NF-15-1-0462), except those marked with *.

He He, Derek Chen, Anusha Balakrishnan, and Percy Liang. Decoupling strategy and generation in negotiation dialogues. In *EMNLP*, 2018.

He He, Nanyun Peng, and Percy Liang. Pun generation with surprise. In *NAACL*, 2019.

Drew A Hudson and Chris Maning. GQA: A new dataset for real-world visual reasoning and compositional question answering. In *CVPR*, 2019.

Dan Iter, Kelvin Guu, Larry Lansing, and Dan Jurafsky. Pretraining with contrastive sentence objectives improves discourse performance of language models. In *ACL*, 2020.

Robin Jia and Percy Liang. Data recombination for neural semantic parsing. In *ACL*, 2016.

Urvashi Khandelwal, He He, Peng Qi, and Dan Jurafsky. Sharp nearby, fuzzy far away: How neural language models use context. In *ACL*, 2018.

Urvashi Khandelwal, Kevin Clark, Dan Jurafsky, and Lukasz Kaiser. Sample efficient text summarization using a single pre-trained transformer. *arXiv*, 2019.

Julia Kruk, Jonah Lubin, Karan Sikka, Xiao Lin, Dan Jurafsky, and Ajay Divakaran. Integrating text and image: Determining multimodal document intent in instagram posts. In *EMNLP*, 2019.

Mina Lee, Chris Donahue, Robin Jia, Alexander Uhunoma, and Percy Liang. SWORD: Improving data coverage and quality for lexical substitution. In *NAACL*, 2021.

Jiwei Li and Dan Jurafsky. Neural net models of open-domain discourse coherence. In *EMNLP*, 2017.

Jiwei Li, Michel Galley, Chris Brockett, Jianfeng Gao, and Bill Dolan. A diversity-promoting objective function for neural conversation models. In *NAACL*, 2016a.

Jiwei Li, Michel Galley, Chris Brockett, Jianfeng Gao, and Bill Dolan. A persona-based neural conversation model. In *ACL*, 2016b.

Jiwei Li, Will Monroe, Alan Ritter, Michel Galley, Jianfeng Gao, and Dan Jurafsky. Deep reinforcement learning for dialogue generation. In *EMNLP*, 2016c.

Jiwei Li, Will Monroe, Tianlin Shi, Sébastien Jean, Alan Ritter, and Dan Jurafsky. Adversarial learning for neural dialogue generation. In *EMNLP*, 2017.

Juncen Li, Robin Jia, He He, and Percy Liang. Delete, retrieve, generate: A simple approach to sentiment and style transfer. In *NAACL*, 2018.

* Percy Liang and Christopher Potts. Bringing machine learning and compositional semantics together. *Annual Reviews Linguist*, 2015.

Reid Pryzant, Richard Martinez, Nathan Dass, Sadao Kurohashi, Dan Jurafsky, and Diyi Yang. Automatically neutralizing subjective bias in text. In *AAAI*, 2020.

Abigail See, Minh-Thang Luong, and Christopher D Manning. Compression of neural machine translation models via pruning. In *SIGNLL*, 2016.

Abigail See, Aneesh Pappu, Rohun Saxena, and Akhila Yerukola. Do massively pretrained language models make better storytellers? In *CoNLL*, 2019.

Sida I Wang, Percy Liang, and Chris Manning. Learning language games through interaction. In *ACL*, 2016.

Sida I Wang, Samuel Ginn, Percy Liang, and Christopher D Manning. Naturalizing a programming language via interactive learning. In *ACL*, 2017.

* Luke S Zettlemoyer and Michael Collins. Learning to map sentences to logical form: Structured classification with probabilistic categorial grammars. In *UAI*, 2005.

Hugh Zhang, Tatsunori B Hashimoto, and Percy Liang. Unifying human and statistical evaluation for text generation. In *NAACL*, 2019.