APPROVED FOR PUBLIC RELEASE

USC Institute for Creative Technologies

U.S. ARMY COMBAT CAPABILITIES DEVELOPMENT COMMAND –
ARMY RESEARCH LABORATORY

Context Is Key:
Annotating Situated Dialogue Relations in Multi-floor Dialogue

Claire Bonial[1],
Mitchell Abrams[2], Anthony L. Baker[3], Taylor Hudson[3], Stephanie M. Lukin[1],
David Traum[4] & Clare R. Voss[1]

[1] U.S. Army DEVCOM ARL
[2] Institute for Human and Machine Cognition (IHMC)
[3] Oak Ridge Associated Universities (ORAU)
[4] USC Institute for Creative Technologies

APPROVED FOR PUBLIC RELEASE

TALK OUTLINE

Goal: Extend our *multi-floor* dialogue annotation schema to account for features of *situated* dialogue—interpretation draws upon info from physical environment, conversational history, robot's physical form, etc.

1. **Multi-floor dialogue**
   - Introduction
   - 2018 unmodified annotation schema
2. ***Situated* dialogue**
   - Situated human-robot dialogue corpus
   - Strengths & weaknesses of original schema annotations as training data
3. **Schema Extensions (Paper Focus)**
   - uniquely mark language that must be grounded to situational context
4. **Annotated Corpus**:
   - Analysis of annotation type frequencies
   - Inter-Annotator Agreement
5. **Conclusions & Future Work**

TALK OUTLINE

Goal: Extend our *multi-floor* dialogue annotation schema to account for features of *situated* dialogue—interpretation draws upon info from physical environment, conversational history, robot's physical form, etc.

1. **Multi-floor dialogue**
   - Introduction
   - 2018 unmodified annotation schema
2. ***Situated* dialogue**
   - Situated human-robot dialogue corpus
   - Strengths & weaknesses of original schema annotations as training data
3. **Schema Extensions (Paper Focus)**
   - uniquely mark language that must be grounded to situational context
4. **Annotated Corpus**:
   - Analysis of type frequencies
   - Inter-Annotator Agreement
5. **Conclusions & Future Work**

USC Institute for Creative Technologies

ihmc ORAU

APPROVED FOR PUBLIC RELEASE

1. MULTI-FLOOR DIALOGUE: INTRODUCTION

Conversational floor: shares common set of speakers and observers

Multi-floor Dialogue: high-level dialogue purposes are the same, and some content is shared, but other aspects (participant structure, turn-taking expectations) are distinct

Time

	Conversational Floor 1		Conversational Floor 2		
ID	Woman	Server	Server	Cook	Drink Server
1	I'll have a cheeseburger and a small coke				
2		Ah no ah, no coke, pepsi			
3	pepsi				
4			One cheeseburger one pepsi		
5				cheeseburger	
6					pepsi

Traum et al. 2018, LREC

1. MULTI-FLOOR DIALOGUE: INTRODUCTION

Conversational floor: shares common set of speakers and observers

Multi-floor Dialogue: high-level dialogue purposes are the same, and some content is shared, but other aspects (participant structure, turn-taking expectations) are distinct

Floor 1 | Floor 2 | Floor 3 | Floor 4

Time

ID	Officer 1	Officer 2	Officer 2	Trans lator	Trans lator	Ricky	Ricky	Lucy
1	Ha!							
2			Ha!					
3					Ha!			
4							Ha!	
5								
6								

End (final exchange)

Traum et al. 2018, LREC

1. MULTI-FLOOR DIALOGUE: 2018 UNMODIFIED ANNOTATION SCHEMA

Transaction Unit (TU): a group containing the initiation and (potential) fulfillment of an intent.

TU 1

1. **Customer:** I'd like a cheeseburger
2. **Waiter:** one cheeseburger.
3. **Waiter:** (placing burger in bag) here you go.
4. **Customer:** thanks!

TU 2

5. **Waiter:** would you like fries with that?
6. **Customer:** Sure, a large one please!
7. **Waiter**: (placing fries box in bag): one large fries.

1. MULTI-FLOOR DIALOGUE: 2018 UNMODIFIED ANNOTATION SCHEMA

Relations: describe the structure between pairs of utterances within a TU

- **Antecedent**: the utterance that a subsequent utterance is addressing (e.g., 2 → 1)
- **Relation-Type:** relationship between utterance and antecedent (e.g., Acknowledgment)

1. **Customer:** I'd like a cheeseburger
2. **Waiter:** one cheeseburger.
3. **Waiter:** (placing burger in bag) here you go.
4. **Customer:** thanks!
5. **Waiter:** would you like fries with that?
6. **Customer:** Sure, large please!
7. **Waiter**: (placing fries in bag): one large fries.

Acknowledgement

Acknowledgement

3rd turn feedback

Answer

Acknowledgement

1. MULTI-FLOOR DIALOGUE: 2018 UNMODIFIED ANNOTATION SCHEMA

Relation Super-Types

Expansions - relate utterances produced by the *same* participant within the *same* floor (4 Subtypes)

Example Subtypes:

1. Customer→Waiter: I'll have a cheeseburger
2. Customer→ Waiter: and a small coke

Continue

Responses - relate utterances by *different* participants in the *same* floor (24 Subtypes)

1. Customer→ Waiter: a small coke, please
2. Waiter→Customer: here you go

Acknowledgement-done

Translations - relate utterances in *different* floors (4 Subtypes)

1. Customer→ Waiter: I'll have a cheeseburger
2. Waiter→Cook: Cheeseburger!!

Translation-right

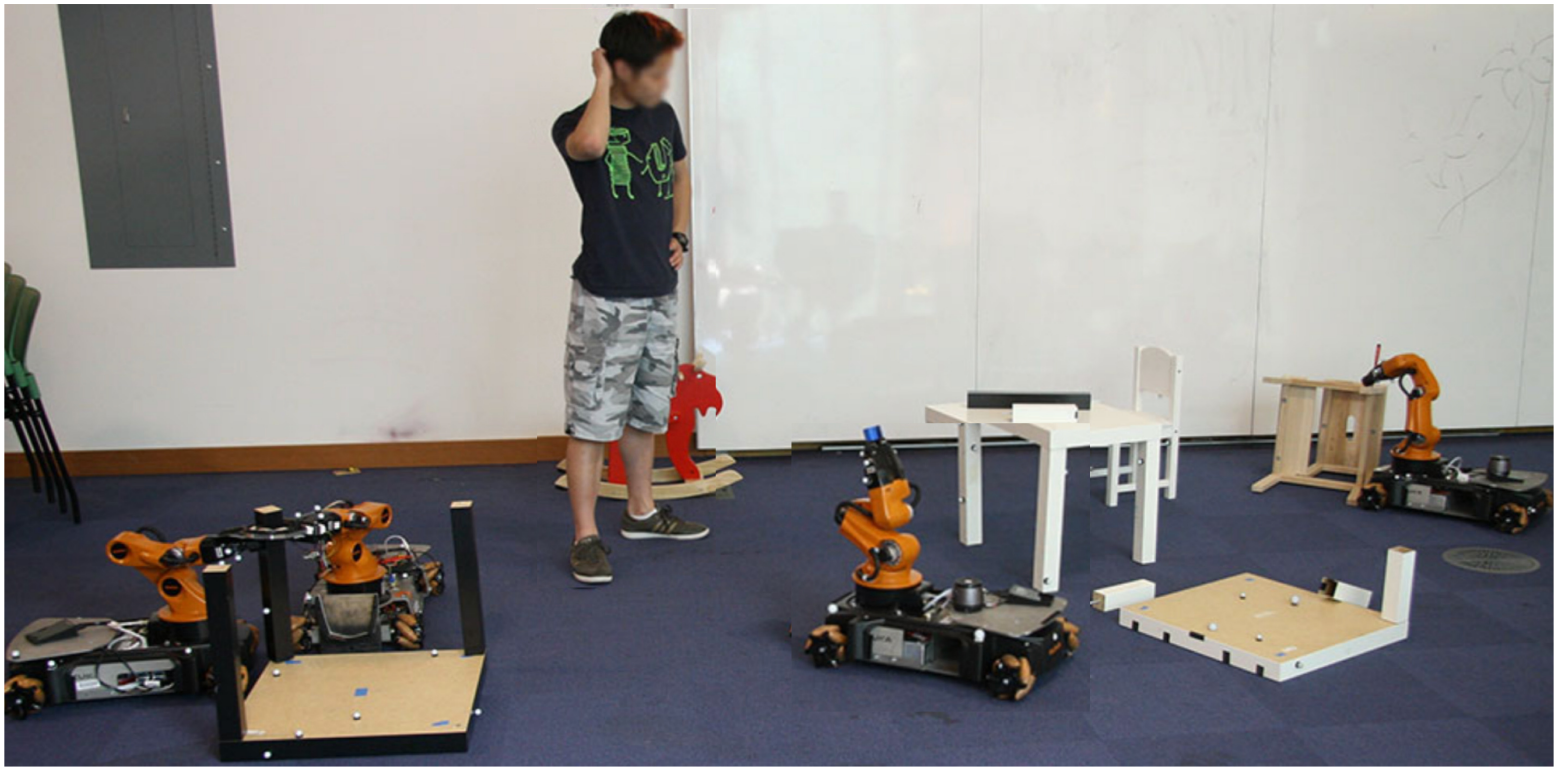
TALK OUTLINE

Goal: Extend our *multi-floor* dialogue annotation schema to account for features of *situated* dialogue—interpretation draws upon info from physical environment, conversational history, robot's physical form, etc.

1. **Multi-floor dialogue**
   - Introduction
   - 2018 unmodified annotation schema
2. ***Situated* dialogue**
   - Human-robot dialogue corpus
   - Strengths & weaknesses of original schema annotations as training data
3. **Schema Extensions (Paper Focus)**
   - uniquely mark language that must be grounded to situational context
4. **Annotated Corpus**:
   - Analysis of type frequencies
   - Inter-Annotator Agreement
5. **Conclusions & Future Work**

2. SITUATED DIALOGUE: HUMAN-ROBOT DIALOGUE CORPUS

Goal: Extend our *multi-floor* dialogue annotation schema to account for features of *situated* dialogue—interpretation draws upon info from physical environment, conversational history, robot's physical form, etc.



Picture Credit: Knepper et al. 2015

2. SITUATED DIALOGUE: HUMAN-ROBOT DIALOGUE CORPUS

How might people talk to a robot in a collaborative task?

- **Wizard-of-Oz** methodology: human “wizard” stands in for automated components
- **Phased** WoZ: use data collected in last phase to train additional automated components in next phase (*DeVault et al. 2014, Artstein et al. 2015*)

2. SITUATED DIALOGUE: HUMAN-ROBOT DIALOGUE CORPUS

“Commander” Participant

What the Participant says

speaks verbal instructions

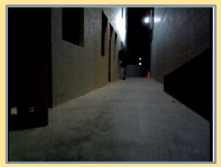
Travel forward down the hall

What the Participant sees

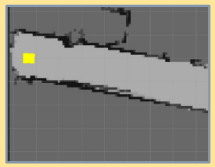
messages from the robot

`How far forward should I move?`

static image

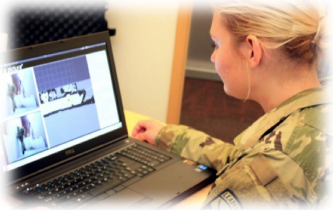


dynamic map



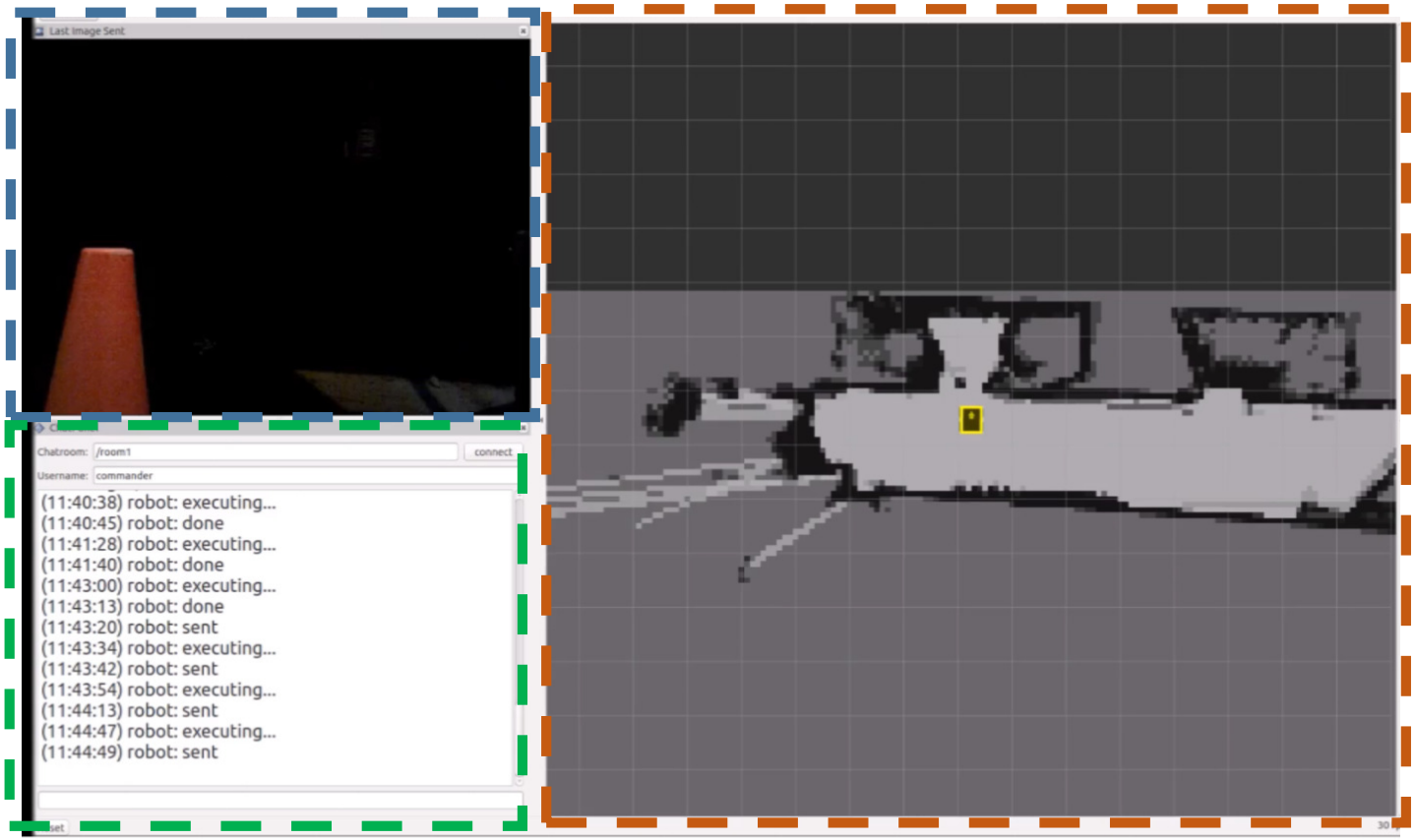
Situated Corpus of Understanding Transactions: Marge et al., RO-MAN 2016

2. SITUATED DIALOGUE: HUMAN-ROBOT DIALOGUE CORPUS



WHAT THE PARTICIPANT SEES

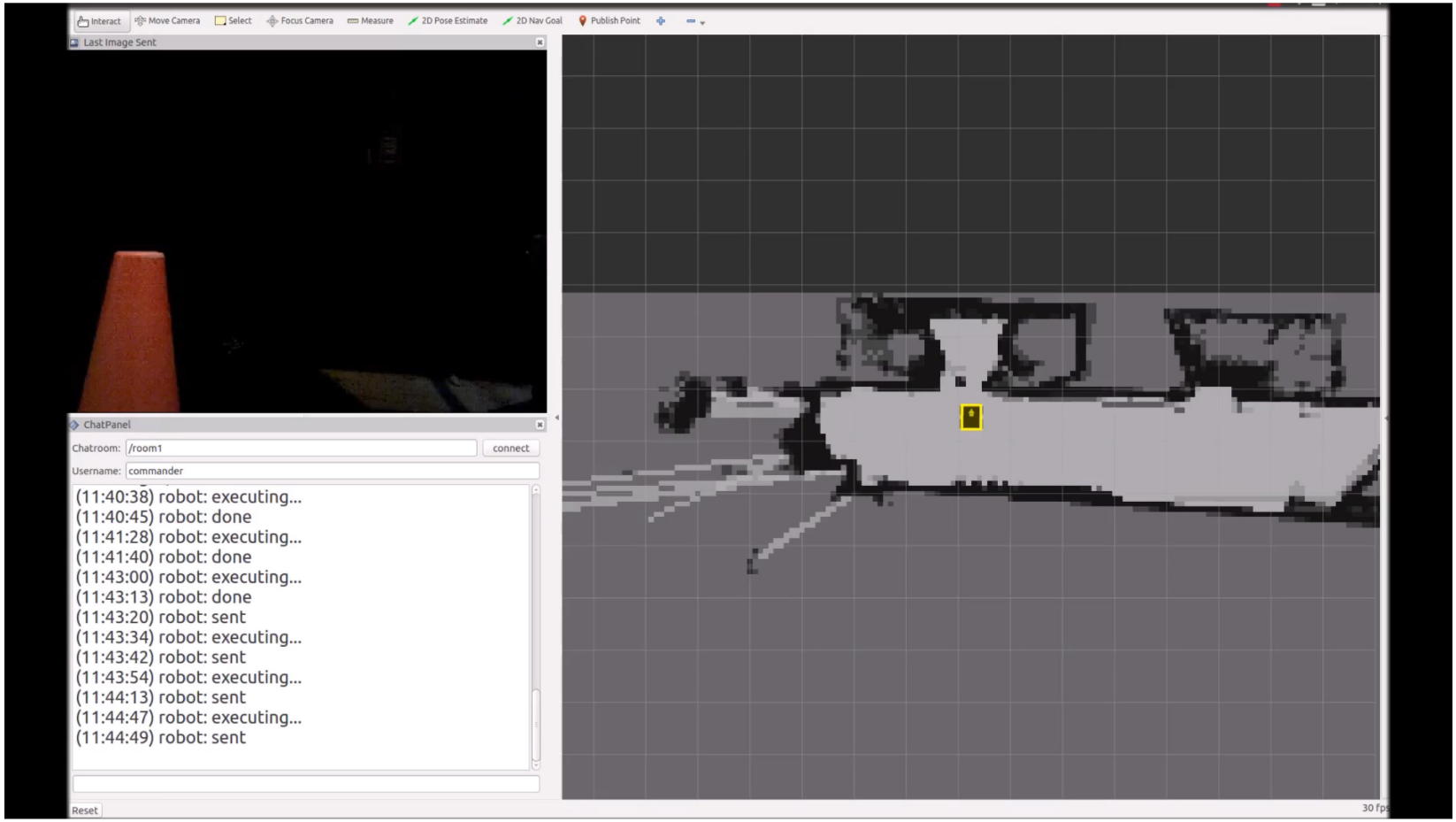
last photo sent by Robot



Robot's replies

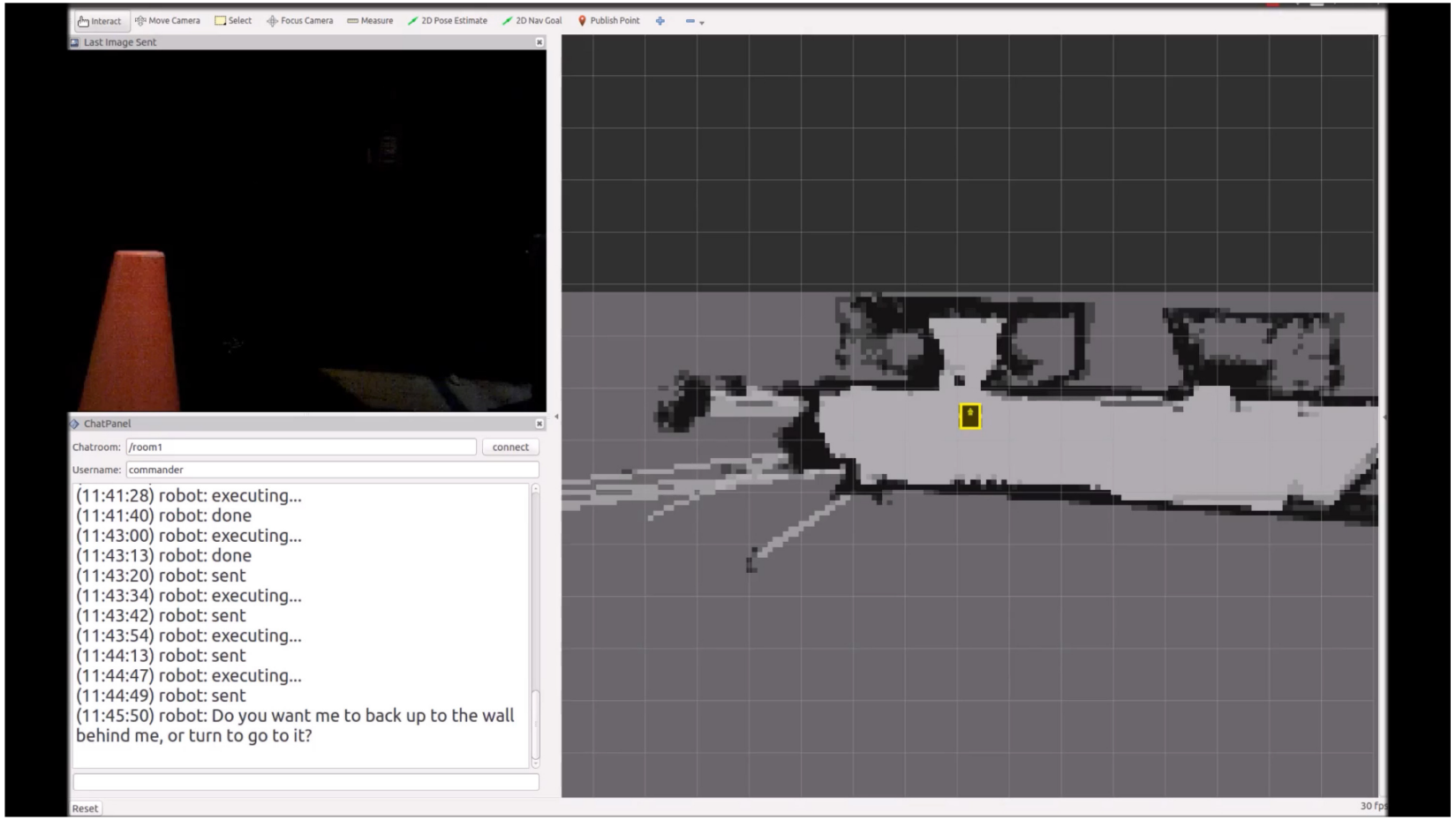
Robot's LIDAR map of searched area
(LIDAR: Light Detection And Ranging)

2. SITUATED DIALOGUE: HUMAN-ROBOT DIALOGUE CORPUS



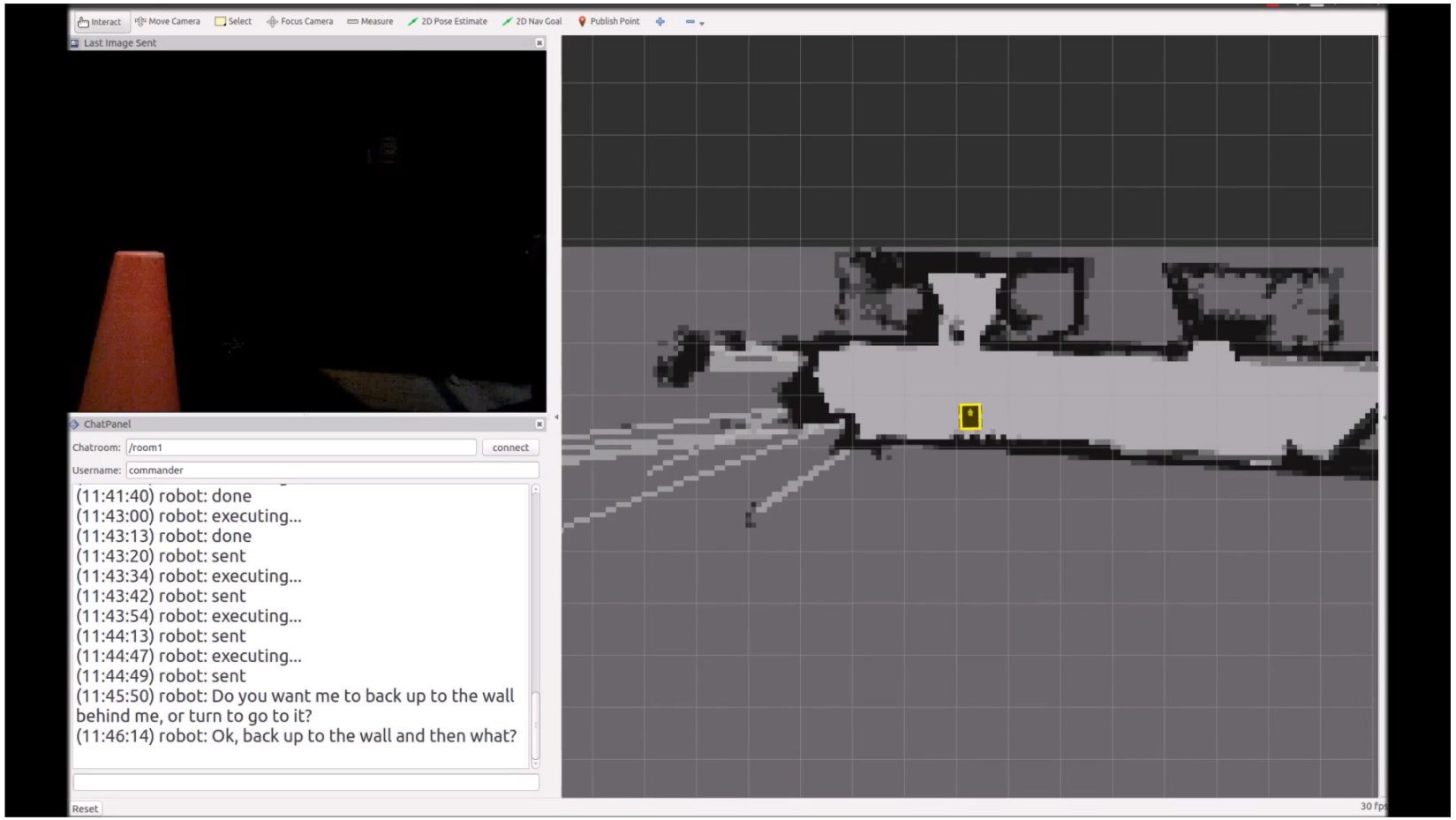
"go to the wall behind you, face north and then take a picture"

2. SITUATED DIALOGUE: HUMAN-ROBOT DIALOGUE CORPUS



"back up"

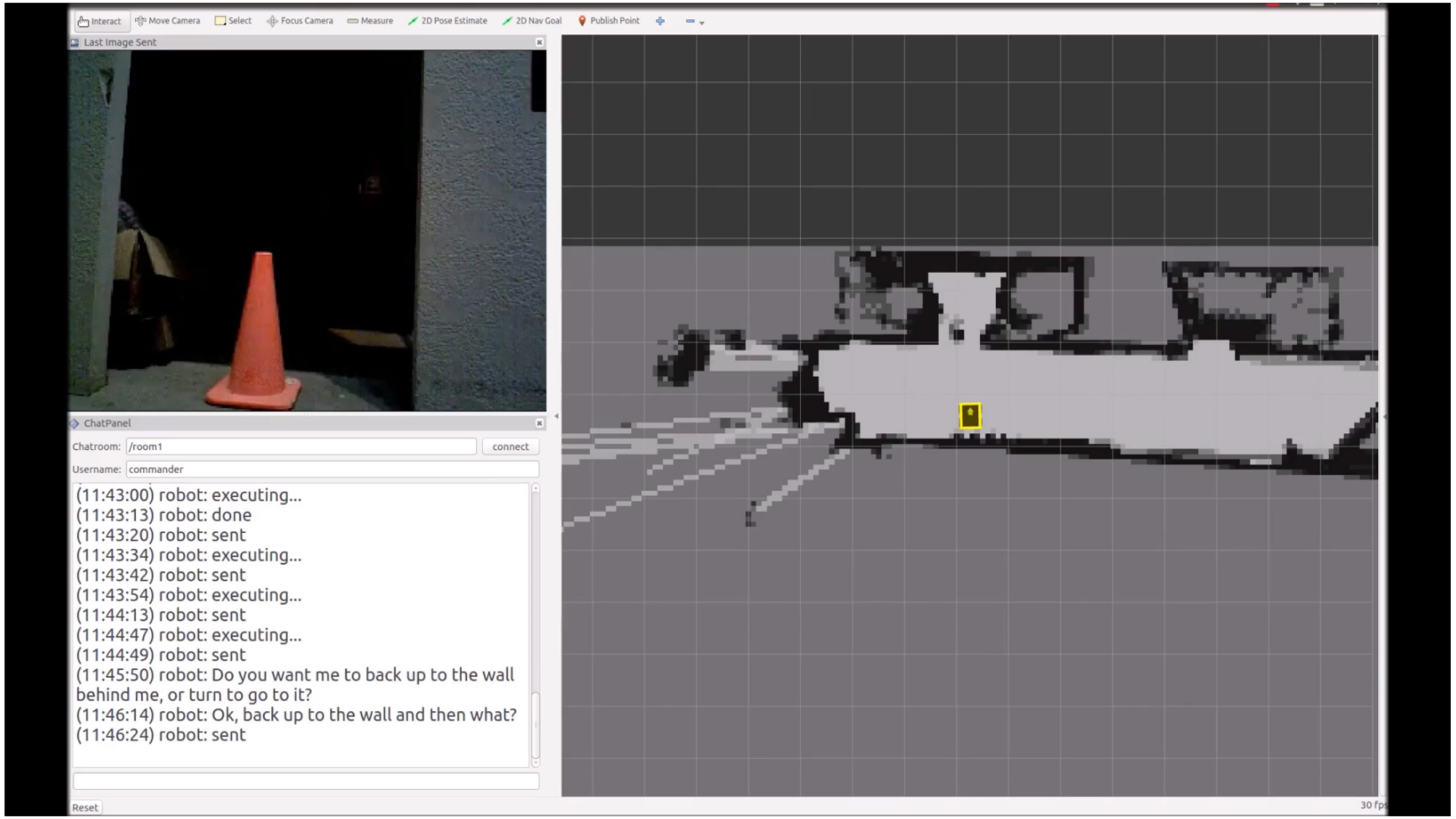
2. SITUATED DIALOGUE: HUMAN-ROBOT DIALOGUE CORPUS



"take a picture"

APPROVED FOR PUBLIC RELEASE

2. SITUATED DIALOGUE: HUMAN-ROBOT DIALOGUE CORPUS



"go to the other door"

2. SITUATED DIALOGUE: HUMAN-ROBOT DIALOGUE CORPUS

What the participant says

speaks verbal instructions

Travel forward down the hall

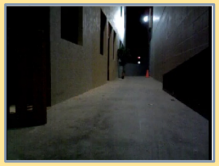
"Commander" Participant

What the participant sees

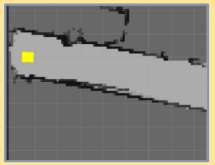
messages from the robot

`How far forward should I move?`

static image



dynamic map



"Behind the scenes"

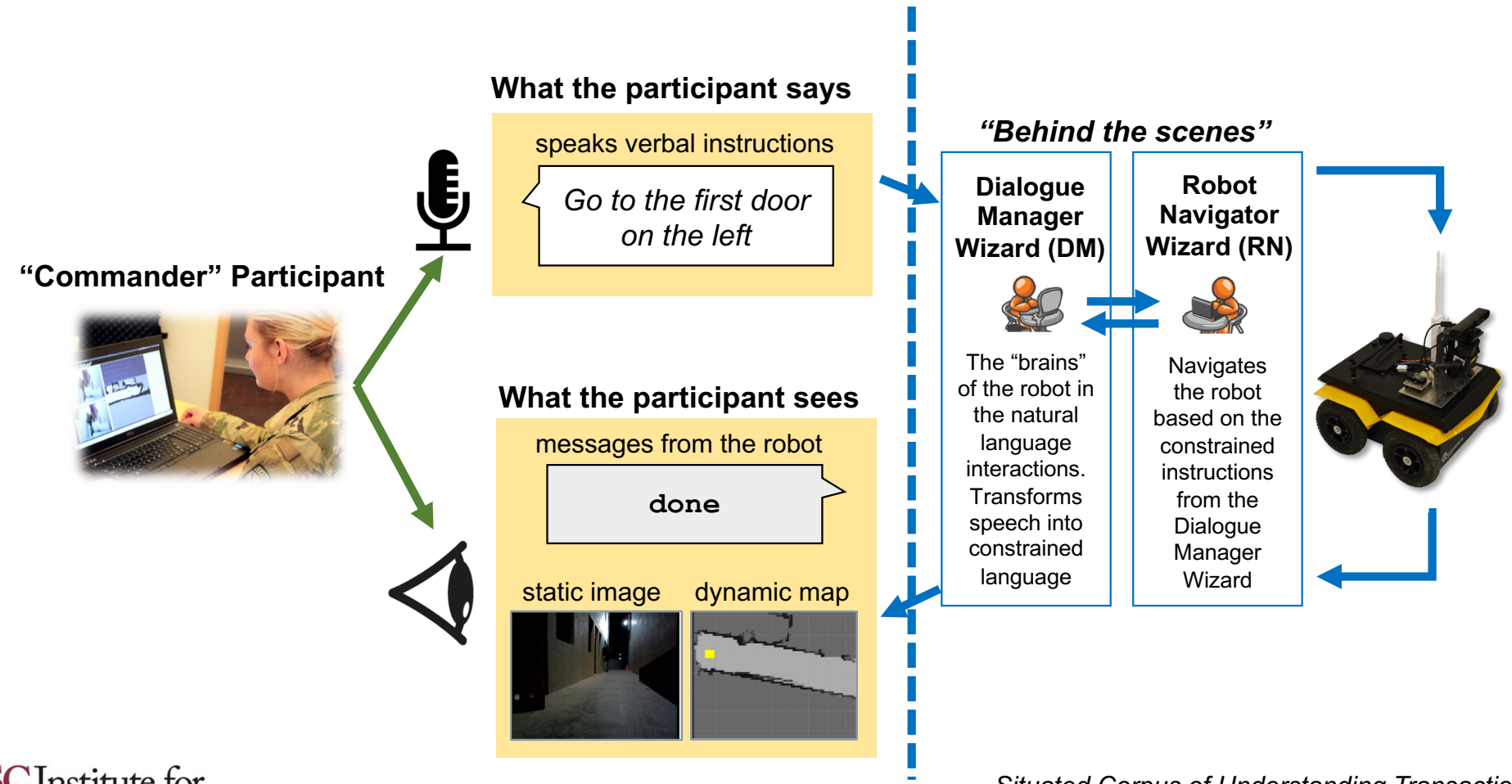
Dialogue Manager Wizard (DM)

The "brains" of the robot in the natural language interactions. May need to request clarification of participant.

Situated Corpus of Understanding Transactions: Marge et al., RO-MAN 2016

APPROVED FOR PUBLIC RELEASE

2. SITUATED DIALOGUE: HUMAN-ROBOT DIALOGUE CORPUS



Situated Corpus of Understanding Transactions: Marge et al., RO-MAN 2016

2. SITUATED DIALOGUE: HUMAN-ROBOT DIALOGUE CORPUS

left floor

Participant

DM Wizard

right floor

RN Wizard

move forward

processing. . .

You can tell me to move a certain distance or to move to an object

go forward 3 feet

processing. . .

move forward 3 feet

moving. . .

done

done

behind the scenes…

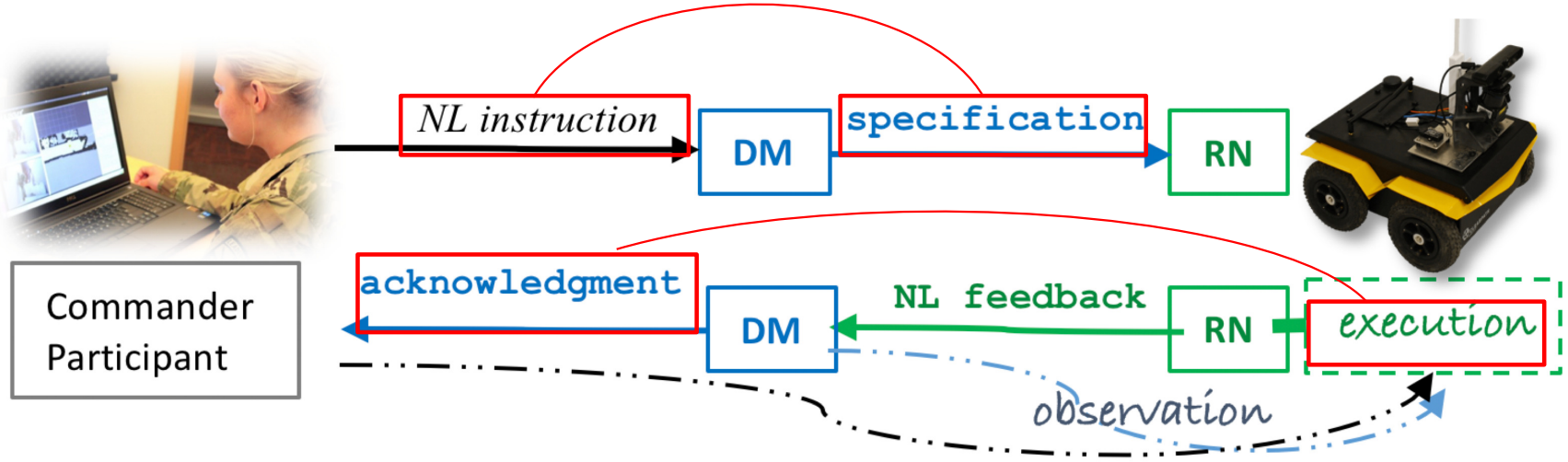
2. SITUATED DIALOGUE: HUMAN-ROBOT DIALOGUE CORPUS

left floor | right floor

ID	Participant (Audio Stream 1)	DM -> Participant (Chat Room 1)	DM-> RN (Chat Room 2)	RN (Audio Stream 2)
1	move forward			
2		processing. . .		
3		You can tell me to move a certain distance or to move to an object		
4	go forward 3 feet			
5		processing. . .		
6			move forward 3 feet	
7		moving. . .		
8				done
9		done		
10	what do you see			
11			send image	
12				sent
13		sent		

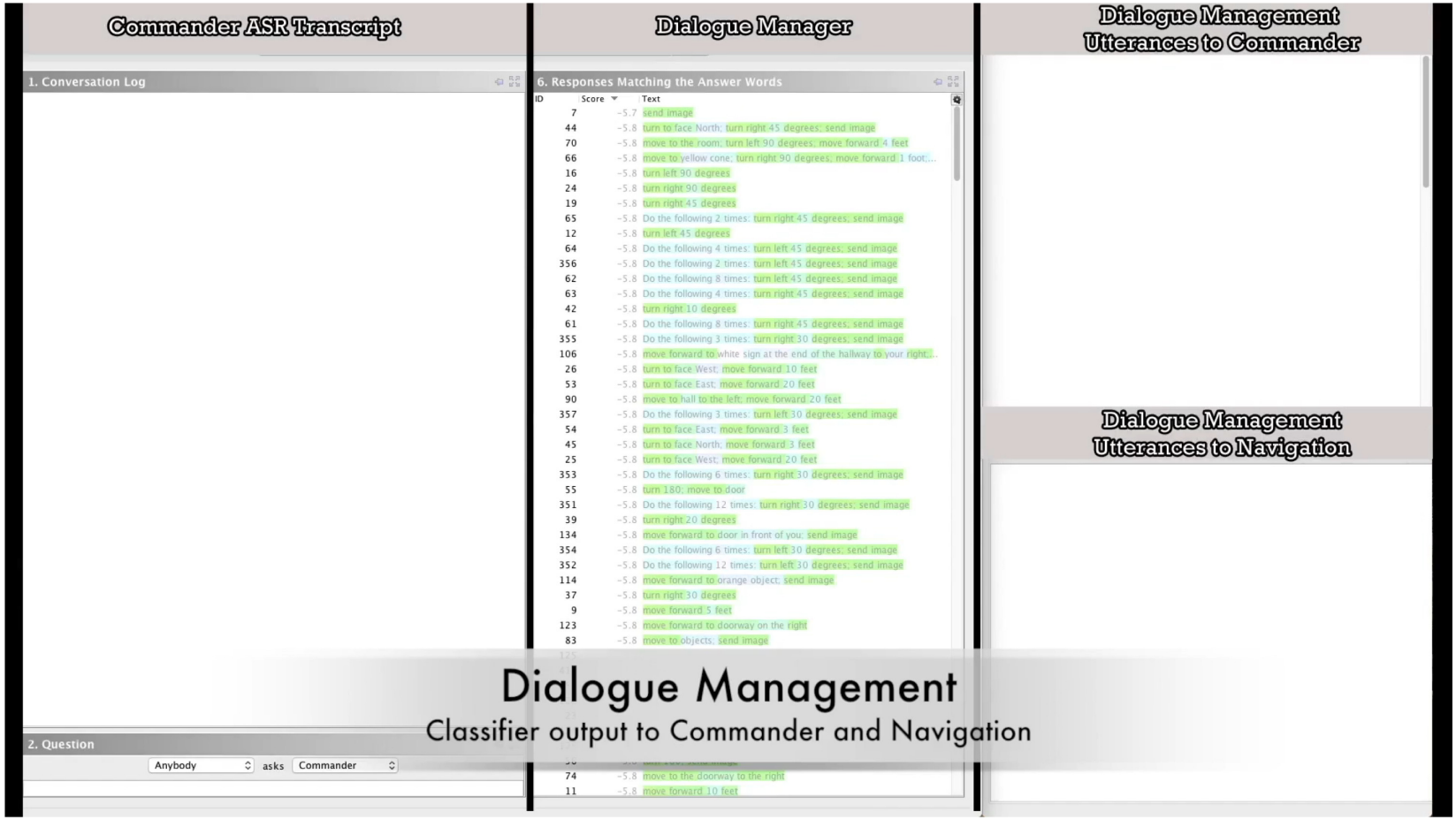
Time

2. SITUATED DIALOGUE: TRAINING DATA STRENGTHS AND WEAKNESSES



- Training data establishes bi-directional associations between NL, execution behavior

2. SITUATED DIALOGUE: TRAINING DATA STRENGTHS AND WEAKNESSES

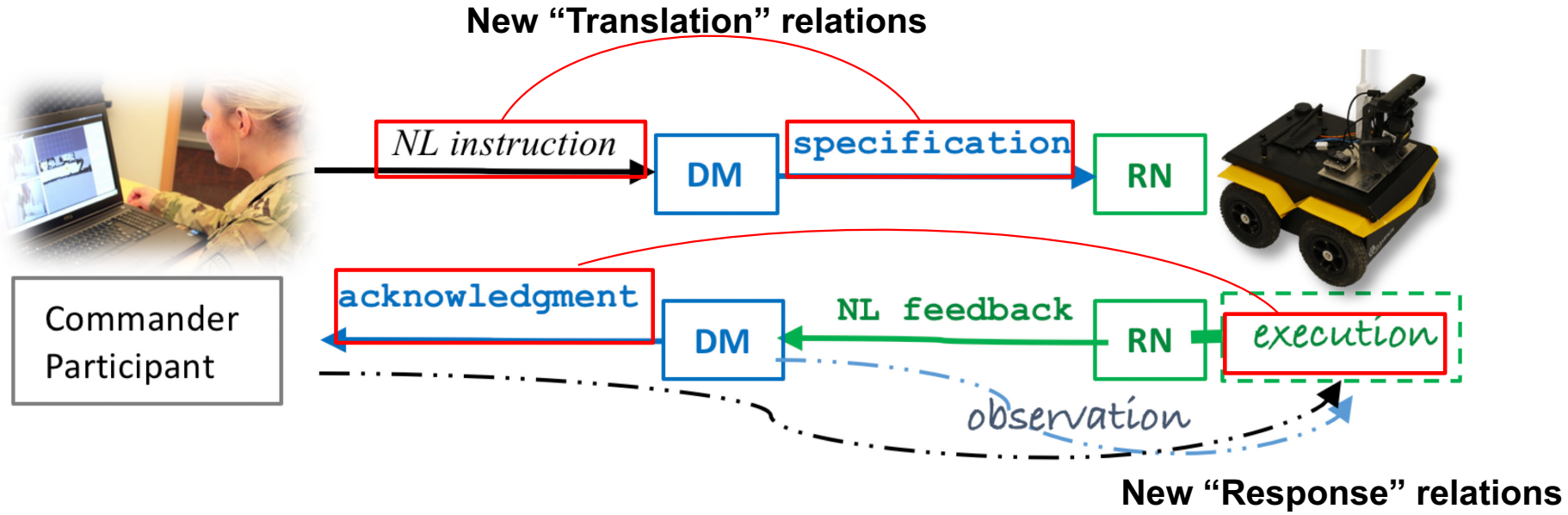


Left Floor: Commander - DM

Right Floor: DM - RN

ScoutBot demo: Lukin et al, ACL 2018

2. SITUATED DIALOGUE: TRAINING DATA STRENGTHS AND WEAKNESSES



- Some **associations** between NL, execution behavior are only valid in particular certain **situated** contexts

TALK OUTLINE

Goal: Extend our *multi-floor* dialogue annotation schema to account for features of *situated* dialogue—interpretation draws upon info from physical environment, conversational history, robot's physical form, etc.

1. **Multi-floor dialogue**
   - Introduction
   - 2018 unmodified annotation schema
2. ***Situated* dialogue**
   - Human-robot dialogue corpus
   - Strengths & weaknesses of original schema annotations as training data
3. **Schema Extensions (Paper Focus)**
   - uniquely mark language that must be grounded to situational context
4. **Annotated Corpus**:
   - Analysis of type frequencies
   - Inter-Annotator Agreement
5. **Conclusions & Future Work**

USC Institute for Creative Technologies

ihmc ORAU

1. MULTI-FLOOR DIALOGUE: 2018 UNMODIFIED ANNOTATION SCHEMA

Relation Super-Types

Expansions - relate utterances produced by the *same* participant within the *same* floor (4 Subtypes)

Responses - relate utterances by *different* participants in the *same* floor ~~(24 Subtypes)~~ (26 Subtypes)

1. Ack-doing-prep
2. Ack-wilco-prep

Translations - relate utterances in *different* floors ~~(4 Subtypes)~~ (10 Subtypes)

1. Translation-r-direct
2. Translation-r-landmark
3. Translation-r-situated
4. Translation-r-default
5. Translation-r-history
6. Translation-r-contextual

Example Subtypes:

1. Customer→Waiter: I'll have a cheeseburger
2. Customer→ Waiter: and a small coke

Continue

1. Customer→ Waiter: a small coke, please
2. Waiter→Customer: here you go

Acknowledgement-done

1. Customer→ Waiter: I'll have a cheeseburger
2. Waiter→Cook: Cheeseburger!!

Translation-right

USC Institute for Creative Technologies

Traum et al. 2018, LREC

3. SCHEMA EXTENSIONS: LANDMARK AND DIRECT TRANSLATION EXTENSIONS

TU	ID	Participant (Audio Stream 1)	DM -> Participant (Chat Room 1)	DM-> RN (Chat Room 2)	RN (Audio Stream 2)	Ante-cedent	Relation-Type
1	1	go through the doorway directly in front of you					
1	2	and take a photo				1	continue
1	3		processing. . .			2*	processing
1	4			move into Conf Room		1	**translation-r-landmark**
1	5			then…		4	link-next
1	6			send image		2	**translation-r-direct**
1	7		moving. . .			1	ack-doing
1	8				uh done and sent	6*	ack-done
1	9		done, sent			8	translation-l

Direct Translations: Uses the same or synonymous words, where the translation is applicable in any physical or conversational context.

Landmark Translations: Refers to a unique landmark name known only to members of the right floor.

3. SCHEMA EXTENSIONS : SITUATED AND DEFAULT TRANSLATION EXTENSIONS

TU	ID	Participant (Audio Stream 1)	DM -> Participant (Chat Room 1)	DM-> RN (Chat Room 2)	RN (Audio Stream 2)	Ante-cedent	Relation-Type
1	1	turn east ninety degrees					
1	2	and travel three feet				1	continue
1	3		processing. . .			2*	processing
1	4			turn left 90 degrees		1	**translation-r-situated**
1	5			then…		4	link-next
1	6			move forward 3 feet		2	**translation-r-default**
1	7		turning...			1	ack-doing
1	8		moving...			2	
1	8				done	6*	ack-done
1	9		done			9	translation-l

Situated Translations: Synonymous with original instruction only in the current physical context, but does not specify a unique landmark.

Default Translations: Supplements information by relying on some default assumption related to a robot behavior or capability

3. SCHEMA EXTENSIONS : HISTORY TRANSLATION EXTENSION

TU	ID	Participant (Audio Stream 1)	DM -> Participant (Chat Room 1)	DM-> RN (Chat Room 2)	RN (Audio Stream 2)	Ante-cedent	Relation-Type
1	1		You often ask for images at the end of movement instructions. Should I send one each time?				
1	2	yes				1	offer-accept
2	3	back up five feet					
2	4			back up 5 feet		3	translation-r-direct
2	5			send image		3	**translation-r-history**
2	7		executing...			3	ack-doing

History Translations: All or part of the translation is only relevant given the dialogue history, in which it was established that a certain instruction should be interpreted in a particular way.

3. SCHEMA EXTENSIONS : CONTEXTUAL TRANSLATION, PREPARATORY ACKNOWLEDGMENTS

TU	ID	Participant (Audio Stream 1)	DM -> Participant (Chat Room 1)	DM-> RN (Chat Room 2)	RN (Audio Stream 2)	Ante-cedent	Relation-Type
1	1	take a picture of the wall on your left					
1	3		processing. . .			1	processing
1	4			move to left wall		1	**translation-r-contextual-partial**
1	6			send image		4	continue
1	7		moving. . .			1	**ack-doing-prep**
1	8				done and sent	6*	ack-done
1	9		done, sent			8	translation-l

Underspecified, Contextual Translations: Draws upon situational or conversational context, but precisely what contextual information is being used is unclear, underspecified, or there are two or more factors.

Acknowledgment – Doing/Will-Comply Preparation: Speaker understands the command and a preparation step required for compliance with the command is underway (doing) or will be done (will-comply).

TALK OUTLINE

Goal: Extend our *multi-floor* dialogue annotation schema to account for features of *situated* dialogue—interpretation draws upon info from physical environment, conversational history, robot's physical form, etc.

1. **Multi-floor dialogue**
   - Introduction
   - 2018 unmodified annotation schema
2. ***Situated* dialogue**
   - Human-robot dialogue corpus
   - Strengths & weaknesses of original schema annotations as training data
3. **Schema Extensions (Paper Focus)**
   - uniquely mark language that must be grounded to situational context
4. **Annotated Corpus**:
   - Analysis of type frequencies
   - Inter-Annotator Agreement
5. **Conclusions & Future Work**

4. ANNOTATED CORPUS: ANALYSIS OF TYPE FREQUENCIES

Corpus Overview:

- 168 human-robot dialogues annotated, validated
- Total of 40,873 relations

Super-Type Relations Frequencies:

- 36.4% Acknowledgment super-type
- 36.5% Translation super-type

New Relation Types Frequencies:

- 70% Direct Translations (no situated language)
- 30% have situated language
- New preparatory acknowledgments have small, but critical impact

Relation
Translation-r
Direct
Direct-partial
Contextual
Contextual-partial
Landmark
Landmark-partial
Situated
Situated-partial
History
History-partial
Default
Default-partial
Updated Ack Types
Will-comply
Doing
Will-comply-prep
Doing-prep

situated (Contextual … Default-partial)

situated (Will-comply-prep, Doing-prep)

Table 5: Frequencies and % of updated relations.

4. ANNOTATED CORPUS: INTER-ANNOTATOR AGREEMENT

Reliability measured through

Inter-Annotator Agreement (IAA)

- Comparable or higher IAA than original, unmodified schema
- New annotation categories are clearly identifiable

Markable Type	Agreement		Distance Metric
	Unmodified Schema	Modified Schema	
Antecedents	0.72–0.82	0.79-**0.94**	Nominal[a]
Relation Types	0.77–0.89	0.83-**0.93**	Nominal[a]
Transaction Units	0.48–**0.93**	0.65-0.85	MASI[b]

[a]Krippendorff (1980) [b]Passonneau (2006)

Table 6: IAA of the original, unmodified schema of Traum et al. (2018) and our modified schema.

5. CONCLUSIONS & FUTURE WORK

Extended multi-floor, multi-party dialogue structure annotations to uniquely mark *situated dialogue*

- Prevents invalid associations of NL - execution behavior in training data
- Critical step in exploration of how to relate and ground language to the context

Ongoing: Bringing together dialogue structure & propositional content with Dialogue-AMR