ARL-TR-8833 • Oct 2019

Dialogue Structure Annotation Guidelines for Army Research Laboratory (ARL) Human–Robot Dialogue Corpus

by Claire Bonial, David Traum, Cassidy Henry, Stephanie M Lukin, Matthew Marge, Ron Artstein, Kimberly A Pollard, Ashley Foots, Anthony L Baker, and Clare R Voss

NOTICES

Disclaimers

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

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

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

ARL-TR-8833 ● OCT 2019

Dialogue Structure Annotation Guidelines for Army Research Laboratory (ARL) Human–Robot Dialogue Corpus

Claire Bonial, Stephanie M Lukin, Matthew Marge, and Clare R Voss
Computational and Information Science Directorate, CCDC Army Research Laboratory

Kimberly A Pollard, Ashley Foots, and Anthony L Baker
CCDC Data & Analysis Center

David Traum and Ron Artstein
Institute for Creative Technologies, University of Southern California

Cassidy Henry
University of Maryland, College Park, Department of Linguistics

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED (From - To)
October 2019	Technical Report	1 January 2013–30 September 2019

4. TITLE AND SUBTITLE
Dialogue Structure Annotation Guidelines for Army Research Laboratory (ARL) Human–Robot Dialogue Corpus

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Claire Bonial, David Traum, Cassidy Henry, Stephanie M Lukin, Matthew Marge, Ron Artstein, Kimberly A Pollard, Ashley Foots, Anthony L Baker, and Clare R Voss

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
CCDC Army Research Laboratory*
ATTN: FCDD-RLC-IT
2800 Powder Mill Road, Adelphi, MD 20783-1138

8. PERFORMING ORGANIZATION REPORT NUMBER
ARL-TR-8833

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

10. SPONSOR/MONITOR'S ACRONYM(S)

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

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

13. SUPPLEMENTARY NOTES
* The work outlined in this report was performed while the US Army Research Laboratory (ARL) was part of the US Army Research, Development, and Engineering Command (RDECOM). As of 31 January 2019, the organization is now part of the US Army Combat Capabilities Development Command (formerly RDECOM) and is now called CCDC Army Research Laboratory. As of February 2019, the US Army Research Laboratory has been renamed the US Army Combat Capabilities Development Command Army Research Laboratory (CCDC ARL).

ORCID ID(s): Claire Bonial, 0000-0002-3154-2852

14. ABSTRACT
Here we provide detailed guidelines on how to annotate a multifloor human–robot dialogue for structure elements relevant to informing dialogue management in robotic systems. We start with transcribed and time-aligned dialogue data collected from participants and Wizards of Oz across multiple years of an Army Research Laboratory human–robot interaction experiment (the "Bot Language" project). We define structure elements and annotation protocol for marking up these dialogue data, with the aim to inform development of a dialogue management system onboard a robot.

15. SUBJECT TERMS
dialogue, human–robot interaction, annotation, dialogue structure, natural-language processing

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Claire Bonial
Unclassified	Unclassified	Unclassified	UU	63	19b. TELEPHONE NUMBER (Include area code) (301) 394-1431

Contents

List of Figures

List of Tables

Acknowledgments

The authors wish to acknowledge the contributions of the following individuals for their assistance in the guidelines established in this report:

- Carla Gordon (Institute for Creative Technologies [ICT])
- Jill Boberg (ICT)
- Elia Martin (University of Maryland [UMD])
- Mitchell Abrams (Georgetown University)
- Brecken Keller (UMD)
- Sue Hill (US Army Combat Capabilities Development Command Army Research Laboratory)

1. Introduction and Annotation Overview

Dialogue systems for robot or agent interaction can yield multiple benefits to the user. An intuitive natural dialogue interface can reduce the need to train human users on special commands, and a voice-operated interaction frees up a user's hands and eyes for other tasks and improved situational awareness, respectively. An important step in developing automated dialogue management systems is to understand the structural relationships that comprise human–robot and human–agent verbal exchanges. Annotated training data can then be used to inform systems with increasing levels of automation.

This report presents an annotation schema for capturing information structure in dialogue. The schema clusters individual utterances into higher-level transaction structures, which aim to achieve an explicit understanding of dialogue intention and relations between individual utterances that are part of this transaction (Traum et al. 2018). Three kinds of annotations are performed for each utterance (further details and definitions follow in the remainder of the document):

1) indicating the transaction unit (TU) it is a part of
2) indicating the direct relation type to the most immediate antecedent (Ant)
3) indicating the antecedent of that relation (Rel)

This annotation schema is applicable to any dialogue meeting the following criteria (further discussed in Section 2): having multiple conversational interlocutors and more than one nonmutual "conversational floor". A conversational floor is an interactional structure that can be thought of as the time and metaphorical space to speak (Edelsky 1981); here, we refer to collaboratively created floors, as opposed to singly created (e.g., a lecture), involving two interlocutors taking turns holding the floor. This schema was developed for cases of multiple "nonmutual" floors in the sense that one interlocutor participates in two conversational floors with distinct conversational partners in each floor, and those conversational partners are not privy to the other conversational floor. The annotation schema serves two main purposes:

1) Allows a formal characterization of dialogue flow, looking at how each task is broken down into different intentional units and how intentions are established, including translations across different conversational floors, clarifications, and acknowledgement of different steps in the process.
2) Serves as training and evaluation data for automated language understanding and dialogue management policies, indicating how the

human "dialogue manager" participant engaged in response and translation activities across floors.

The remainder of this report describes the annotation schema in the context of the US Army Combat Capabilities Development Command Army Research Laboratory (ARL) Bot Language Project. Section 2 begins with an overview of SCOUT, which fulfills the criteria described previously.

2. Understanding the Domain and Preparing the Corpus

We aim to support natural language understanding within the broader context of ongoing research to develop a spoken dialogue system (Marge et al. 2016) that will run onboard a remotely located, autonomous robot collaborating with humans in search and navigation tasks (e.g., disaster relief). In developing this dialogue system, we are making use of and providing annotations over the Situated Corpus of Understanding Transactions (SCOUT), a corpus of human–robot dialogue (Lukin et al. 2018). This corpus was collected via a phased "Wizard-of-Oz" (WoZ) methodology, in which human experimenters, or "wizards", perform the planned dialogue and navigation capabilities of the robot during experimental trials, unbeknownst to participants interacting with the "robot" (Marge et al. 2017).

The WoZ method is bottom-up in the sense that we do not assume that we can know a priori how humans communicate with a robot in a shared task. Instead, our WoZ methodology facilitates a data-driven understanding of how people talk to robots in our collaborative domain. Similar to DeVault et al. (2014), we use the WoZ methodology only in the early stages of a multistage development process to refine and evaluate the domain, and provide training data for automated dialogue system components. In all stages of this process, participants communicating with the "robot" speak freely, even as increasing levels of automation are introduced in each subsequent stage or "experiment". The iterative automation process utilizes previous experiments' data. Currently, we are in the fourth experiment of the ongoing series, and the corpus utilized in the present annotation guidelines includes data and annotations from the first three experiments.

In those experiments, a naïve participant (in that they are unaware of the utilization of wizards for the experiment) is tasked with instructing a robot to navigate through a remote, unfamiliar, house-like environment and asked to find and count objects such as shoes and shovels. In reality, the participant (given the role of Commander [CMD] in these experiments) is not speaking directly to the robot, but rather to an unseen Dialogue Manager (DM) wizard who listens to the participant's spoken instructions, and in turn passes simplified instructions to a Robot Navigator (RN) wizard, who joysticks the robot to complete the instructions.

The configuration and exchange of information in Bot Language Experiments 1–3 is depicted in Fig. 1 and consists of three conversational interlocutors, four communication streams (Fig. 1 includes their method of communication; speech or text messages), and two distinct, nonmutual floors (Fig 1. depicts with whom they can communicate).

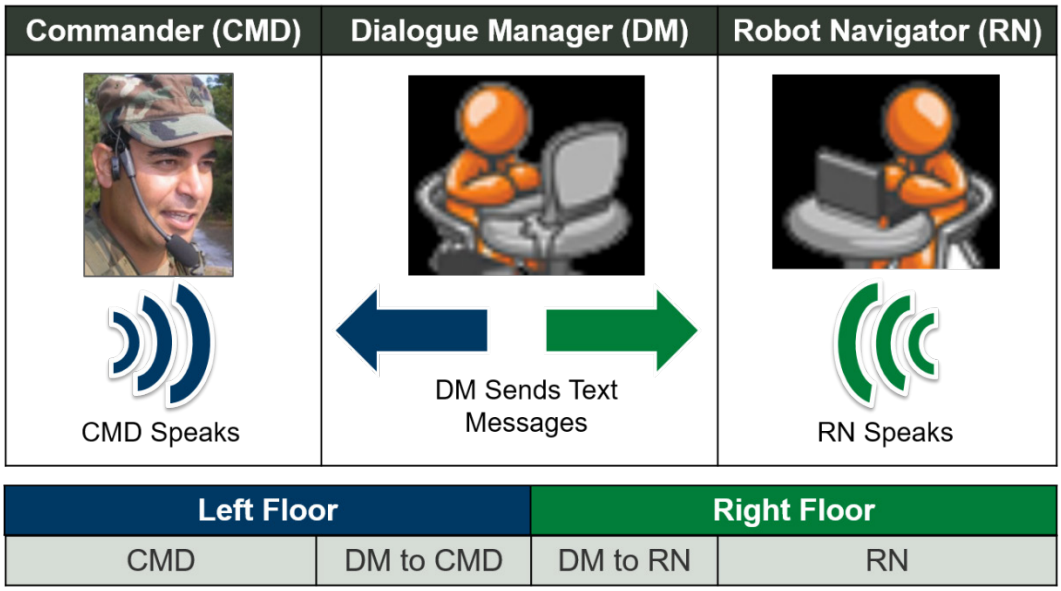


Fig. 1 Configuration and exchange of information between the CMD, DM, and RN in the Bot Language Experiments 1–3

The conversational interlocutors are the CMD, an experimental participant interacting with "a robot"; the DM, a human wizard who plays the part of the front end of the robot, interacting with both the CMD and the operational/navigational component of the robot; and the RN, a human wizard standing in for the navigation component of the robot, taking commands from DM, teleoperating the robot to complete those commands, and communicating robot and task state to DM.

The four communication streams and their medium of communication are "CMD" (spoken), "DM to CMD" (text messages), "DM to RN" (text messages), and "RN" (spoken).

Given that the DM acts as an intermediary passing communications between the CMD and the RN, the dialogue takes place across two nonmutual conversational floors: the left (L) and right (R) floors. The L is communication between CMD and "the robot", with the DM acting as the front end and sending messages, and contains streams "CMD" and "DM to CMD". The R is between the DM and RN, and contains streams "DM to RN" and "RN". The RN cannot directly access the L floor, and CMD cannot directly access the R floor. In Experiments 1–3, the "CMD" and "RN" streams were spoken (captured in audio files and transcribed). The DM is the

only interlocutor who has access to both conversational floors and serves as the mediator of information exchange between the two floors.

The flow of dialogue from CMD to DM, DM to RN, and subsequent feedback to the CMD, as annotated by the guidelines presented in this document, can be seen in Table 1.

Table 1 Example of a minimal TU in a SCOUT dialogue annotation, which contains an instruction initiated by the CMD, its translation to a simplified form that is passed to the robot navigator ("DM to RN"), the acknowledgement of the task execution, and the passing of that acknowledgement back to the CMD. TU, Ant, and Rel types are indicated in the right columns (Traum et al. 2018).

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

In Experiment 1, the communication from "DM to RN" and "DM to CMD" involved text messages typed by the DM and captured in time-stamped logs. In Experiments 2 and 3, the view of the "DM to RN" and "DM to CMD" messages were seen as text messages by the RN and CMD, respectively; however, the messages were not typed but were selected using a graphical user interface (GUI) that would send prewritten text when a button was pushed. In some cases, the GUI button press would provide prewritten text with an open field for typing limited content into a form (e.g*., "Move forward ___ feet."*). In Experiments 2 and 3, there was also an audio signal accompanying a "DM to CMD" message to alert the CMD for some kinds of messages.

In Experiment 1, there was some audio from the RN that was not captured, but in Experiments 2 and 3, an additional recording device was used to capture these messages. Times were synchronized across the text message channels ("DM to CMD" and "DM to RN"); however, the timing of audio messages was according to the audio channel and only semi-automatically synchronized with the other streams. In some cases, this yielded message times and ordering of messages that were inaccurate. Both of these points can impact the data in ways that may be noted in the guidelines to follow: annotators may suspect either misalignment or extraneous communications not captured in the transcript.

These annotation guidelines assume that the message streams have been compiled into a transcription file with the following column headers (see the annotated example illustrated in Fig. 2)*:

A. *ID#*: Each utterance is given a distinct ID (here, a positive integer)

B. *Timestamp*: what time this utterance was completed (for audio messages in CMD and RN streams, this is not necessarily accurate with respect to other streams)

C. *Commander*: Speech transcription from the CMD, part of the L floor

D. *DM->CMD*: Text messages from "DM to the CMD", part of the L floor

E. *DM->RN*: Text messages from "DM to the RN", part of the R floor

F. *RN*: Speech transcription from the RN, part of the R floor

G. *Transaction*: The TU (if any) that this utterance is a part of (see transaction unit annotation in Section 3)

H. *Antecedent*: The annotated antecedent (indicated by the ID) generally of the most immediate direct relation between this utterance and a prior utterance (the antecedent; see antecedent coding in Section 4.1).

I. *Relation*: The annotated relation between this utterance and the antecedent (see relation coding in Section 4.2).

J. *Notes*: Any notes about the annotation (e.g., if unsure how to annotate, note suspicions of misalignment, cases where the annotation schema does not seem clear, or the best way to capture the observed relationships). If there are multiple interpretations of how to describe a segment or annotate relations and antecedents, one interpretation should be selected and used consistently; the Notes column can be used to describe the interpretation.

A	B	C	D	E	F	G	H	I
ID#	Timestamp	Commander	DM->CMD	DM->RN	RN	Transaction	Antecedent	Relation
199	14:43:58.60	face east				35		
200	14:44:00.22	take picture				35	199	continue
201	14:44:02.54		ok			35	200*	ack
202	14:44:05.02			turn to face East		35	199	translate-r
203	14:44:06.31			then. . .		35	202	link-next
204	14:44:07.54			send image		35	200	translate-r
205	14:44:21.41				done and sent	35	204*	ack-done
206	14:44:23.46		done, sent			35	205	translate-l
207	14:44:30.63	face south				36		

Fig. 2 Aligned transcript with annotations

* Columns may vary across the final annotation spreadsheets for Experiments 1 and 2; the ones described in this document are the columns for the Experiment 3 annotations):

3. Transaction Unit (TU) Annotation

Each utterance is placed into a group (a TU) defined by the initiation and fulfillment of an intent. A TU contains an initial message (typically a command or a question) by one speaker and all subsequent messages by the same and other speakers across channels to complete the initial intent (i.e., a set of commands from the CMD and responses/communications from the DM and RN that complete a single or grouped set of intentions).

3.1 TU Annotation

For each utterance, annotate the TU in column G such that every member of the TU has the same number. The first TU that is started in the dialogue should be annotated with 1, with each new TU that is started receiving the next highest integer (e.g., 2, 3, 4….). Use the same TU label for all utterances related to extending, clarifying, completing, cancelling, and/or acknowledging this task transaction (Fig. 3).

A	B	C	D	E	F	G	H	I
ID	Timestamp	Commander	DM->CMD	DM->RN	RN	Transaction	Antecedent	Relation
12	136.47	move forward three feet				3		
13	140.49	send picture				3	12	continue
14	146.82			move forward 3 feet, image		3	13*	translation-r
15	150.09		executing...			3	13*	ack-doing
16	155.75				image sent	3	14	ack-done
17	158.83		sent			3	16	translation-l

Fig. 3 Simple transaction, fully annotated

One difficult point is that an intention is often broken into a sequence of related utterances by the same speaker rather than just being one utterance. For example, the CMD often issues a sequence of commands. The rule of thumb to use in deciding whether subsequent commands are part of the same TU or the initiation of a new TU is whether the DM (and possibly the RN) has started to react to the sequence. Thus, a sequence of commands from the CMD that are not interrupted by a DM acknowledgement would all be seen as part of the same unit (see Figs. 2 and 3 for examples).

If, following a DM request for clarification, the CMD clarifies and continues with further commands before the DM acknowledges or translates to the RN, keep the new commands as part of the first TU. For example, lines 132–139 are all part of the same TU in Fig. 4. It may be difficult to tell whether an instruction that follows a clarification is trying to amend the previous instruction or abandoning it and starting a new instruction; use your best judgment and follow conventions established in past challenges cases, a variety of which are given in the Appendix.

.

ID	Timestamp	Commander	DM->CMD	DM->RN	RN	Transaction
132	1007.38	turn left two feet				24
133	1048.81		I don't know what you mean by turn left two feet. Do you want me to move to face something?			24
134	1064.78	turn left fifty degrees				24
135	1070.79	send photo				24
136	1078.83			turn left 50, image		24
137	1082.47		executing...			24
138	1094.41				image sent	24
139	1096.77		sent			24

Fig. 4 TU with clarification: TU 24 begins at ID 132 and continues until ID 139. This is all one TU. The clarification request and answer in lines 133–134 are part of this TU because they are in service of completing this same single intent.

In contrast, new commands that occur after the DM's response would be seen as starting a new TU (unless they are clarifications or repairs of the previous command, such as restating or changing what was previously specified). For example, if a suggestion to do an action is given by the DM in response to a question from the CMD, and then the CMD issues an instruction to do a different action, the instruction should start a new TU that is separate from the question and response. Fig. 5 contains an example of this situation.

ID#	Timestamp	Commander	DM->CMD	DM->RN	RN	Transaction
173	15:07:40.54	what is behind the cone				31
174	15:07:50.12		I'm not sure.			31
175	15:07:51.53		I can move to take a good picture of an object that you are interested in. <beep>			31
176	15:08:00.22	take a good picture of the cone				32
177	15:08:04.21		processing. . .			32
178	15:08:15.20			send image of the cone		32
179	15:08:19.26		executing. . .			32
180	15:08:22.82				uh done and sent	32
181	15:08:25.89		done, sent			32

Fig. 5 Example of a new TU: The DM suggests something in ID 175, but the CMD ignores it and suggests a new course of action (new intent) in ID 176, so this would begin a new TU

Figure 6 shows an extended example of how TUs are annotated. In the figure, several clarifications and installments are needed before the full intention is realized.

ID	Timestamp	Commander	DM->CMD	DM->RN	RN	Transaction
17	113.15	move forward a little bit again				4
18	113.15	and then turn left				4
19	119.33	then take a photo				4
20	128.4		Does a little bit mean two feet?			4
21	137.16			move forward two feet		4
22	143.43				done	4
23	156.6	yes it does				4
24	172.66		How far should I turn left? Until facing the door to my left?			4
25	177.87	perpendicular to the doorway				4
26	202.24			move until perpendicular to the first doorway to your left		4
27	207.04		executing...			4
28	209.9				done	4
29	220.91			turn to face the doorway to the left		4
30	230.18				done	4
31	231.27		done			4
32	235.18	take a photo				5
33	241.3			photo		5
34	242.13				image sent	5
35	244.01		sent			5
36	244.05	move around the cone into the next room				6
37	249.01	turn right ninety degrees				6
38	252.38	and take a photo				6

Fig. 6 Example of an extended transaction: TU 4 begins on ID 17 and continues until ID 32, where TU 5 begins. TU 4 is an example of an extended transaction without overlap.

Overlap between TUs is common, where one transaction might begin before another is finished (Fig. 7).

ID	Timestamp	Commander	DM->CMD	DM->RN	RN	Transaction
87	646.79	go back to the other doorway				15
88	659.39			return to the other doorway		15
89	662.92		executing...			15
90	668.77	face the doorway				16
91	673.31	and back up until you hit the wall				16
92	675.8		done			15
93	678.57	or just about hit the wall				16
94	689			turn to face the doorway, back up to the wall		16
95	692.48		executing...			16
96	697.84	stop				17
97	700.56	take a photo				17
98	710.01		done			17
99	712.46			photo		17
100	712.84				image sent	17
101	714.96		sent			17

Fig. 7 Example of overlap: TU 16 is introduced before TU 15 ends. This is overlap.

3.2 Supporting Project-Specific TU Annotation

In any dialogue data collection, there may be project-specific nuances that appear in the aligned transcripts. This subsection presents three particular cases for TU annotation, with strategies for generalizing to other projects.

1) Calibration of recording equipment at the start of an experiment

Experiments collecting speech data may require calibration of multiple recording devices. The verbal instruction to "calibrate" may occur before the recording actually begins. However, an assumption may be made that the "calibrate" instruction was indeed issued if the recording has begun.

In this project, most transcripts begin with the DM sending the RN the "calibrate" command. This command is the DM and RN's signal to begin recording and move the robot a bit so that the 2-D map populates some for the CMD to see. This command is actually initiated by the CMD, who requests that the robot calibrate, but because recording does not begin until that command is given, it is never included in the recording. Nonetheless, to preserve the dialogue structure relationships of the portion of calibration that is captured, we insert a single row for "(calibrate)" from the CMD with ID 0 and a "0" timestamp at the very beginning of the trial (as in Fig. 8).

A	B	C	D	E	F	G
ID#	Timestamp	Commander	DM->CMD	DM->RN	RN	Transaction
0	0	(calibrate)				1
1	)7 14:17:56.65		calibrating. . .			1
2	)7 14:18:00.14			calibrate		1
3	)7 14:18:45.76	<X: audio sync in three two one>				X-CMD
4	)7 14:18:47.58				<X: in three two one>	X-CMD
5	)7 14:18:49.35	<loud noise>				X-CMD
6	)7 14:18:49.35				<loud noise>	X-CMD
7	)7 14:18:57.59	<X: xxx>				X-CMD
8	)7 14:18:59.33		calibration complete			1
9	)7 14:19:11.05	i'm ready				2
10	)7 14:19:16.40			participant is ready		2
ID#	Timestamp	Commander	DM->CMD	DM->RN	RN	Transaction
301	)7 14:39:09.90	turn fifteen degrees to your left				50
302	)7 14:39:18.88		I will turn left 15			50
303	)7 14:39:21.07			turn left 15 degrees		50
304	)7 14:39:25.86				done	50
305	)7 14:39:28.14		done			50
306	)7 14:39:45.28	<X: so how many doorways did you find>				X-CMD
307	)7 14:39:47.69	five doorways				X-CMD
308	)7 14:39:48.50			please wait		X-CMD
309	)7 14:39:49.57	<X: okay>				X-CMD
310	)7 14:39:50.54	<X: and how many cones>				X-CMD
311	)7 14:39:52.16	two				X-CMD
312	)7 14:39:54.51	<X: and do you think there's any type of container in which a soldier might be able to move things>				X-CMD
313	)7 14:39:59.42	two				X-CMD
314	)7 14:40:02.60	<X: okay>				X-CMD
315	)7 14:40:05.86		Good job!			X-CMD
316	)7 14:40:10.27			end task		X-CMD

Fig. 8 Calibrate command and experimenter-CMD floor: The row with ID 0 and timestamp 0 is manually inserted with "(calibrate)." Note that all communications involving the experimenter floor should have their TU marked as "X-CMD" and their antecedents and relations left blank.

2) Dialogue between the CMD participant and experimenter

Verbal communication may take place between a CMD participant and the experimenter conducting the study that is not relevant to the dialogue structure annotation presented in this document, and may be coded as extraneous. These dialogues may occur when the CMD is about to begin the experiment, if the CMD has a question, or at the conclusion. While this project does not separate out the CMD and experimenter from the same floor, other projects may wish to add an additional floor to preserve this communication.

In this project, there are recorded communications from the experimenter generally at both the very beginning of a trial and at the end of the trial. This includes audio sync communications at the beginning and the experimenter's task questions at the end of the trial. These communications stemming from the experimenter and/or between the experimenter and the CMD (which take place on a distinct conversational floor from the floors captured in our annotation) all receive the marker X-CMD (i.e., experimenter-CMD) in the Transaction column as their TU (Fig. 8).

Communications marked as X-CMD also will not receive any antecedent or relation markings as described in Section 4.

3) Handling of human-in-the-loop or WoZ errors

If a dialogue system has a human-in-the-loop or a Wizard-of-Oz supporting the communication, mistakes may be made as a result of human error. Depending on the nature of the mistake and how the participant or other parties react to it, these cases may be coded in different ways.

In our project, it is possible for the DM to press the wrong button on the communications interface and therefore send a message that is inappropriate for the current context. These should be handled differently depending upon which of two basic cases occur:

a) The DM sends a corrected, appropriate message before either the CMD or the RN reacts to the mistaken message—the TU for the mistaken message should be "ERR" and no antecedent/relation marked (i.e., the mistaken utterance can effectively be ignored/removed and the remaining dialogue structure is sensible and unaffected) (Fig. 9)

ID#	Timestamp	Commander	DM->CMD	DM->RN	RN	Transaction
211	1 15:10:39.43	move forward until i say stop				37
212	1 15:10:45.86		There's too much lag in our communications for instantaneous instructions. <beep>			37
213	1 15:11:04.36	proceed forward				38
214	1 15:11:05.46		You can tell me to turn a number of degrees or to face something. <beep>			ERR
215	1 15:11:09.19		How far forward should I go? <beep>			38
216	1 15:11:14.74	ten inches				38
217	1 15:11:22.28		Hmm. . .			38
218	1 15:11:31.44		I will move forward 1 foot			38
219	1 15:11:33.91			move forward 1 foot		38
220	1 15:11:37.21				done	38
221	1 15:11:39.75		done			38

Fig. 9 ERR*:* Neither the CMD nor the RN react to ID 214, which appears to be a mistaken button press by the DM given that a response discussing "turning" does not fit the context. The corrected response is given in ID 215, and the remaining dialogue structure is sensible and unaffected taking into account only ID 215 while effectively ignoring ID 214 by treating it as ERR.

b) Either the CMD or the RN (or both) react to the mistake—the TU, antecedent, and relation annotations must be marked to the best of the annotator's ability as part of the ongoing dialogue structure (i.e., the mistaken utterance *cannot* be ignored/removed without the remaining dialogue structure being affected as it is motivated by and/or a response to the mistaken utterance). For example, the CMD says "move forward 2 feet" but the DM passes "move back 2 feet", which is acted upon by the RN.

4. Utterance Antecedent and Relations Annotation

All utterances that are part of the same TU will have one or more relations between utterances in that TU. The relations between utterances are partly defined by which streams the antecedent (prior utterance) and "follow-up" (utterance that is related to the antecedent) utterances are part of, and partly by the relationship of the semantic and pragmatic contents.

Relations (enumerated in Table 2 and described in Section 4.2) are annotated by marking the relation type (in column I) and the antecedent for this relation (in column H). Figure 10 shows an example of this markup and the TUs (column G).

A	B	C	D	E	F	G	H	I
ID	Timestamp	Commander	DM->CMD	DM->RN	RN	Transaction	Antecedent	Relation
0	0	(calibrate)				1		
1	3.18				calibration complete	1	0	ack-done
2	9.69		calibration complete			1	1	translation-l
3	13.43		Please be aware that there may be lag times in receiving and processing your requests. I'll say DONE when I've completed your request, or SENT after sending you a photo, or I may ask for more information or let			1		
4	21.35	take a photo				2		
5	28.95			photo		2	4	translation-r
6	29.97				image sent	2	5	ack-done
7	32.1		sent			2	6	translation-l
8	52.62	I'm ready				3		

Fig. 10 Simple follow-up example: In ID 5 in the DM-> RN column, the DM is following up the instruction in ID 4, so the antecedent marking would be "4". This sequence is uncomplicated by multiple commands or overlap, so each one follows the other in succession (5 is preceded by 4, 6 by 5, etc.).

4.1 Antecedent Annotation

Generally, an utterance is a follow-up to an antecedent line if it is the most recent direct follow-up to the antecedent. In this case, there will also be a specific type of follow-up relation between the antecedent and follow-up, as described next. In column H, enter the utterance ID (column A) that the utterance under annotation is a follow-up to (the "antecedent" of, according to the relation in column I).

An example is shown in Fig. 10. In ID 4, the CMD says "take a photo," a follow-up is the DM to RN "photo" request, so the antecedent for the "photo" request in ID 5 is ID 4. The RN confirming "image sent" in ID 6 is a follow-up to ID 5, so its antecedent is 5. This is then consequently followed up by the "DM to CMD" "sent" confirmation in ID 7, so the antecedent for this is ID 6—the most recent relation. Antecedents must always "match" the relation as the relation decided upon stems from determining how the follow-up under consideration addresses the antecedent (described in Section 4.2).

For multiple commands in succession by the same speaker and part of the same group, each line has the preceding line as its antecedent. For an utterance that is directly related to a whole sequence of utterances from the same speaker, use the last line of that sequence along with an asterisk (e.g., 39*), which would be referring to the set of commands ended by line 39 (Fig. 11).

ID#	Timestamp	Commander	DM->CMD	DM->RN	RN	Transaction	Antecedent
38	15:56:38.68	rotate forty five degrees right				6	
39	15:56:41.19	and take a picture				6	38
40	15:56:45.73		processing. . .			6	39*

Fig. 11 Example of * antecedent: a straightforward example of a line (ID 40) with an antecedent of the previous sequence of utterances ending with 39

We do not have a way to indicate a subset of antecedents; rather this interpretation will be applied post-annotation. Therefore, we also interpret the * to refer to "the set of commands ending with the annotated line and starting with the last line not already encapsulated by a different relation". In other words, we interpret the * to mean "everything above this point that that relation has not already been applied to". Figure 12 shows an example of a chain of antecedents where a follow-up is associated only with a subset in that chain (relations are described in Table 2 in Section 4.2).

ID#	Timestamp	Commander	DM->CMD	DM->RN	RN	Transaction	Antecedent	Relation
35	2018-08-07 10:49:22.58	move forward two feet				6		
36	2018-08-07 10:49:24.68	<no speech>						
37	2018-08-07 10:49:25.29	face				6	35	continue
38	2018-08-07 10:49:26.31	east				6	37	continue
39	2018-08-07 10:49:26.84	and take a picture				6	38	continue
40	2018-08-07 10:49:30.00		processing. . .			6	39*	processing
41	2018-08-07 10:49:33.94			move forward 2 feet		6	35	translation-r
42	2018-08-07 10:49:35.51			then. . .		6	41	link-next
43	2018-08-07 10:49:36.77		moving. . .			6	35	ack-doing
44	2018-08-07 10:49:38.34			turn to face East		6	38*	translation-r
45	2018-08-07 10:49:39.61			then. . .		6	44	link-next
46	2018-08-07 10:49:41.87		turning. . .			6	38*	ack-doing
47	2018-08-07 10:49:43.80			send image		6	39	translation-r

Fig. 12 Example of a chain of antecedents: The antecedent of ID 44 is 38*, indicating "everything including and above 38 that has not been included in that same translation-r relation" (see Table 2, Section 4.2). Note that ID 35 has the translation-r relation applied to it by ID 41, therefore we interpret the antecedent of 44 to mean inclusively 36–38.

Another instance we have noticed is partial antecedents, where the DM responds to only part of a sequence. There may be times where a CMD is giving multiple commands and then the DM sends them incrementally to the RN as opposed to clustering the commands into a single response to the RN. If the entire content of a translation is contained within a single utterance, use that utterance as the antecedent (as in ID 59 of Fig. 13—only ID 57 of the CMD sequence is antecedent to the clarification question of distance). If on the other hand, the content comes from multiple utterances, use the * notation to indicate the last utterance in the sequence with translated content (as in ID 58, where the original instructions in lines 55–57 are translated in one line). Figure 13 shows an example of multiline commands in which the DM responds to individual increments or lines of that command.

ID	Timestamp	Commander	DM->CMD	DM->RN	RN	Transaction	Antecedent
54	387.1	okay				8	53
55	387.1	then turn around one hundred and eighty degrees				9	
56	391.75	move forward about two feet				9	55
57	394.28	then turn left and continue down the hallway you're in				9	56
58	410.64			turn around 180, move forward two feet, turn left to face back down the hallway		9	57*
59	422.27		How far should I continue down the hallway?			9	57
60	444.26				done	9	58

Fig. 13 Example of multiline commands: In the example from the May 17 alley, lines 55–57 compose one sequence by the CMD. ID 55 is the antecedent to ID 56, 56 is antecedent for row 57, and so forth. The antecedent of the "DM to RN" communication in ID 58 is the entire sequence, indicated by 57*. The antecedent of ID 59 is only ID 57—the portion containing the command being clarified.

Usually, there will not be a direct antecedent relation between utterances in columns D and E (i.e., the two DM-originating streams). This would mean that the DM communicates to one party because of their own communication to another (e.g., commenting to the RN what they have told to CMD). More commonly, either the CMD or RN's relevant preceding utterance would be an antecedent for both DM follow-ups. An example can be seen in Fig. 13 in that ID 57 is antecedent to both IDs 58 and 59.

In some cases, this "incremental processing" results in part of the command being ignored or lost by the DM and clarification or repetition is needed. Figure 14 shows partial follow-ups with a repetition.

ID	Timestamp	Commander	DM->CMD	DM->RN	RN	Transaction	Antecedent
39	299.02	move forward till you reach the wall				7	
40	304.91	then turn ninety degrees right				7	40
41	308.1	send picture				7	41
42	335.75		I see a few walls. Which wall should I move to?			7	39
43	335.76	the wall straight ahead				7	42
44	360.35			Move to the wall directly ahead of you		7	43*
45	370.96				done	7	44
46	372.92		I'm moving to the wall ahead of me. Then what?			7	43
47	374.19	turn right ninety degrees				7	46
48	378.94	send picture				7	47
49	384.47			turn right 90, image		7	48*
50	387.78		executing...			7	48*
51	393.54				image sent	7	49
52	396.96		sent			7	51

Fig. 14 Partial follow-ups: The DM is "incrementally processing" the CMD's instructions, as evidenced by IDs 44 and 49 containing separate pieces of the same original instruction. After clarification of the first portion of the instruction, the CMD is asked to repeat the later portions of the instruction.

4.2 Relation Annotation

Utterance relations are used to describe the mechanics and structure of a conversation. There are many possible relations between different pairs of utterances. In general, the relations fall under three broad relation types:

1) **Expansion**: by same speaker within the same stream
2) **Translation**: content on one floor being communicated (by the DM) to the other floor
3) **Response**: by other speaker in the same floor

Table 2 summarizes all relations; details and examples of each relation type follow.

Table 2 Relations summarized by type

General relation type	Relation	Annotation label
Expansion *Relation between utterances of the same speaker*	Continue	continue
	Correction	correction
	Link-next	link-next
	Summarization	summarization
Translation *Relation between utterances of different speakers*	Translation-left	translation-l
	Translation-right	translation-r
	Partial translation	translation-l-partial, translation-r-partial
	Quotation	quotation
	Comment	comment
Response *Relation between utterances of different speakers*	Processing	processing
	Acknowledge (general, underspecified)	ack
	Acknowledge understand	ack-understand
	Acknowledge unsure	ack-unsure
	Acknowledge try	ack-try
	Acknowledge will comply	ack-wilco
	Acknowledge doing	ack-doing
	Acknowledge done	ack-done
	Acknowledge can't	ack-cant
	Partial acknowledgment	ack-understand-partial, ack-unsure-partial, etc.
	Negative acknowledgment	nack
	Missing information	missing-info
	Request clarification	req-clar
	Clarification repair	clar-repair

Table 2 Relations summarized by type (continued)

General relation type	Relation	Annotation label
Response *Relation between utterances of different speakers*	Request repeat	req-repeat
	Clarification repeat	clar-repeat
	Request done status	req-done
	Clarification done status	clar-done
	Answer	answer
	Non-answer response	nar
	Make offer	offer
	Offer accept	offer-accept
	Offer reject	offer-reject
	Reciprocal response	reciprocal
	Third-turn feedback	3feedback
	Other response	other

The antecedent for an utterance (annotated in column H and previously discussed in Section 4.1) determines the relation type that will be used to annotate that utterance. Determining linguistic antecedents and relations involves some judgment; however, there are several strict guidelines to follow when assigning relations. If it has been determined that the antecedent of an utterance is the same speaker, the relation for that utterance *must* be an Expansion type. Expansions cannot have an antecedent is a different speaker from that of the follow-up utterance. Both Translation and Response types *must* have an antecedent is a different speaker, while Translation types *must* have an antecedent that is from a conversational floor distinct from the conversational floor of the follow-up utterance under consideration.

As mentioned, the antecedent for an utterance determines the relation type, and usually the focus is the utterances' *most immediate linguistic antecedent*, even though it may be involved in several previous antecedents. Regarding the general rule of annotating the most immediate linguistic antecedent, as shown in Fig. 15; utterance ID 7 is both an acknowledgement (ack) that the command in utterance ID 4 has been done (ack-done) as well as a translation-left of the utterance in line ID 6, where the same information is conveyed by the RN. However, we only annotate the *most recent direct relation*, therefore we relate ID 7 to ID 6 as the antecedent instead of to ID 4, and assign the "translation-l" relation to ID 7.

A	B	C	D	E	F	G	H	I
ID	Timestamp	Commander	DM->CMD	DM->RN	RN	Transaction	Antecedent	Relation
0	0	(calibrate)				1		
1	3.18				calibration complete	1	0	ack-done
2	9.69		calibration complete			1	1	translation-l
3	13.43		Please be aware that there may be lag times in receiving and processing your requests. I'll say DONE when I've completed your request, or SENT after sending you a photo, or I may ask for more information or let			1		
4	21.35	take a photo				2		
5	28.95			photo		2	4	translation-r
6	29.97				image sent	2	5	ack-done
7	32.1		sent			2	6	translation-l
8	52.62	I'm ready				3		

Fig. 15 Simple follow-up example: In ID 5 in the DM-> RN column, the DM is following up the instruction in ID 4, so the antecedent marking would be "4". This sequence is uncomplicated by multiple commands or overlap, so each one follows the other in succession (5 is preceded by 4, 6 by 5, etc.). Note that relation types are determined by the most immediate antecedent; thus, ID 6 is an ack-done of ID 5, but not ID 4.

An exception to the rule of considering the most direct/recent antecedent is the case where an utterance is both a continuation of the same stream but also a direct relation to a single utterance in another stream. In this case, mark the other relations (e.g., translate-r or ack-doing) rather than the expansion. If the relation is to multiple utterances in a sequence, then mark the continuation relation.

4.2.1 Expansion Relations

Expansion relation types are used for utterances between the same speaker within the same stream. The specific expansion relations are as follows:

1) **continue**: add more content (could include more specific discourse or rhetorical relations) (Fig. 16). This includes the case where the previous utterance by this speaker is another relation such as response or translate, particularly if partial and continued by this utterance.

ID	Commander	DM->CMD	DM->RN	RN	Transaction	Antecedent	Relation
1				calibration complete	1	0	ack-done
2		calibration complete			1	1	translation-l
3	move forward five feet				2		
4	turn south thirty degrees				2	3	continue

Fig. 16 continue: ID 4 is a "continue" relation following up the antecedent in ID 3

2) **correction**: replace some content or change one or more prior-expressed values. This includes utterances such as "cancel", "stop", and "nevermind", which may cancel instructions underway (see Fig. 17), as well as expressions of the correction/replacement itself (e.g., "turn right" after having requested "turn left" previously).

ID	Timestamp	Commander	DM->CMD	DM->RN	RN	Transaction	Antecedent	Relation
164	745.14	go through the door to the north west				22		
165	754.43		processing. . .			22	164	processing
166	757.93			move into Cleaning room		22	164	translate-r
167	766.63		moving. . .			22	164	ack-doing
168	777.77	stop				22	164	correction
169	779.8			stop		22	168	translate-r

Fig. 17 correction

Note that an utterance should be marked as correction only if there is no intervening clarification request by another speaker about the antecedent on that floor. If there is, then the utterance should be marked as "clar-repair" (described later) rather than correction. Correction, on the other hand, should be marked when the same speaker corrects their previous instructions (hence it is a relation that applies to same speaker within same stream as marked in Fig. 17). In other words, correction does not cross speakers or floors, meaning that a CMD message cannot be a correction of the DM's message.

3) **link-next**: an explicit discourse connective marker (e.g., "and", "then", or "but") that indicates that the antecedent will have a relation with the following utterance (Fig. 18).

ID	Timestamp	Commander	DM->CMD	DM->RN	RN	Transaction	Antecedent	Relation
147	509.68	rotate right ninety degrees				28		
148	512.03	and take a picture				28	147	continue
149	516.25		processing. . .			28	148*	processing
150	519.73			turn right 90 degrees		28	147	translation-r
151	520.62		I will turn right 90 degrees			28	147	ack-wilco
152	521.91		and. . .			28	151	link-next
153	522.51			then. . .		28	150	link-next
154	523.68			send image		28	148	translation-r

Fig. 18 link-next

Note that when link-next connects two or more utterances that, combined, form a complex response addressing a single command utterance ID (i.e., there is one CMD utterance being addressed by two or more DM utterances including a link-next), the antecedent of the first line of the complex response (a translation-r-partial) is the CMD utterance, while the link-next and continuations of the response take the previous portion of the response as their antecedent. See IDs 155–157 and 171–173 in Fig. 19 for examples. Contrast these with the link-next usage in Fig. 18, in which the link-next connects utterances that each address a single command utterance (i.e., there is one CMD utterance being addressed by one DM utterance).

153	667.27	<pause> go west about five feet				1		
154	674.73		processing. . .			1	153	processing
155	681.65			turn to face West		1	153	translation-r-partial
156	682.8			then. . .		1	155	link-next
157	684.12			move forward 5 feet		1	156	continue
158	699.47		moving. . .			1	153	ack-doing
159	701.97				heh <pause> uh done	1	157*	ack-done
160	707.51		done			1	159	translation-l
161	710.38	<notification sound>						
162	711.03		I'm facing a wall now. <beep>			2		
163	711.68		Did I successfully do what you asked? <beep>			2	162	continue
164	717.38	<inhale; notification sound>						
165	718.47		If you're having trouble figuring out how far I should move or turn to get to something, you can always try telling me to go directly to whatever you are interested in. <beep>			2	163	continue
166	727.84	go through the doorway				3		
167	735.01		I see more than one doorway. <beep>			3	166	missing-info
168	737.63		Which doorway? <beep>			3	166	req-clar
169	740.57	go through the doorway to your <long pause> left				3	168	clar-repair
170	750.64		processing. . .			3	169	processing
171	757.41			move into Shoe Start room		3	169	translation-r-partial
172	762.12			via. . .		3	171	link-next
173	771.21			TV hallway		3	172	continue
174	776.14		moving. . .			3	169	ack-doing
175	783.01	stop				3	169	correction
176	784.14				done	3	173*	ack-done
177	784.99		done			3	176	translation-l

Fig. 19 link-next connecting multiple utterances in a complex response to a single CMD utterance

4) **summarization**: an utterance that does not add to (continue) or remove from (correction or clar-repair) a prior set of utterances, but just restates all or part of it, perhaps, but not necessarily, in other words or another order (Fig. 20). A heuristic for determining if an utterance is a summarization is if the utterance can be removed and the intent still be understood in full. If the command that is being summarized can already be fully executed as specified, then the follow-up command that restates it can be a summarization. This helps distinguish from the continue relation, which adds information.

   Summarization does not cross speakers or floors, and thus it can only be in response to one's own messages. In other words, a message from the CMD cannot be a summarization of the DM's message.

ID	Timestamp	Commander	DM->CMD	DM->RN	RN	Transaction	Antecedent	Relation
35	274.61	robot go to that white sign that's in the picture				5		
36	278.8	and take a picture of it				5	35	continue
37	281.06	also do you speak any other languages				5	36	continue
38	305.46		I do not speak. I can recognize some English words.			5	37	answer
39	317.79	go to the end of the hallway that is to your right				5	36*	summarization
40	322.74	and take a picture of the white sign that's on the doorway or on the uh wall				5	36*	summarization

Fig. 20 summarization

4.2.1 Translation Relations

Translation relation types are used by the DM following an utterance by a speaker in another floor. The specific translation relations are as follows:

1) **translation-l**: from the right floor to the left, providing the same content to the CMD that the RN provided to DM (Fig. 21).

ID	Timestamp	Commander	DM->CMD	DM->RN	RN	Transaction	Antecedent	Relation
7	92.51	yes send picture				2	6	clar-repair
8	105.07			turn right 30 degrees, image		2	7*	translation-r
9	108.69		executing...			2	7*	ack-doing
10	110.82				image sent	2	8	ack-done
11	122.47		sent			2	10	translation-l

Fig. 21 translation-l: "sent", following "image sent"

2) **translation-r**: from the left floor to the right, providing the same content to the RN that has been provided by the CMD to DM (Fig. 22).

ID	Timestamp	Commander	DM->CMD	DM->RN	RN	Transaction	Antecedent	Relation
6	84.99		I'm not sure what you mean by turn south 30 degrees. Should I turn further to the south by 30 degrees?			2	4	req-clar
7	92.51	yes send picture				2	6	clar-repair
8	105.07			turn right 30 degrees, image		2	7*	translation-r

Fig. 22 translation-r: "turn right 30 degrees, image", following "turn south 30 degrees", clarification-request, clarification, and then "yes send picture"

3) **–partial**: either of the previous two relations can be "partial" if it only translates part of the command of an utterance or sequence (with part of it being translated in a later utterance). The example in Fig. 23, "move east 10 feet", requires first a turn then a movement to complete the instruction. The first translation-r is annotated with a "partial" and any subsequent translations in the sequence with a "continue".

ID#	Timestamp	Commander	DM->CMD	DM->RN	RN	Transaction	Antecedent	Relation
47	2018-08-07 11:22:01.47	move <pause .34> east ten feet				8		
48	2018-08-07 11:22:06.01		processing...			8	47	processing
49	2018-08-07 11:22:08.97			turn to face East		8	47	translation-r-partial
50	2018-08-07 11:22:10.02			then...		8	49	link-next
51	2018-08-07 11:22:11.35		turning...			8	47	ack-doing
52	2018-08-07 11:22:12.48			move forward 10 feet		8	50	continue

Fig. 23 partial translation: partial translations right are used to achieve execution of instructions

4) **quotation**: telling the speaker in one floor what was said by the speaker in the other floor, but without the same illocutionary force as the original. Have not seen an example yet, but an example might be "asked us to read the Arabic writing".

5) **comment**: talking about a speaker/utterance in one floor to the speaker in the other floor, without relaying a command or translation (Fig. 24).

ID	Timestamp	Commander	DM->CMD	DM->RN	RN	Transaction	Antecedent	Relation
119	817.56			Move forward until you reach the wall closest to you, turn left 90, image		16	116*	translation-r
120	824.39				direct daily ahead	16	188	comment
121	829.63				This is the wall directly ahead	16	119	comment

Fig. 24 comment

4.2.3 Response Relations

Response relation types are used by one speaker following an utterance by the other speaker in the same floor. The specific **response** relations are as follows:

1) **processing**: indicates a message was received and is being worked on (similar to ellipses shown in text messaging) (Fig. 25). Does not explicitly indicate understanding, as the next utterance might be a clarification rather than acknowledgement or something that implies understanding. Can be realized in experiment with "processing…" or "Hmm".

ID	Timestamp	Commander	DM->CMD	DM->RN	RN	Transaction	Antecedent	Relation
198	704.17	rotate right twenty degrees				35		
199	708.36	and move three feet forward				35	198	continue
200	712.08		processing. . .			35	199*	processing
201	719.81			turn right 20 degrees		35	198	translation-r
202	721.32			then. . .		35	201	link-next
203	722.87			move forward 3 feet		35	199	translation-r
204	725.31		turning. . .			35	198	ack-doing
205	728.61		moving. . .			35	199	ack-doing
206	733.18				done	35	203*	ack-done
207	734.9		done			35	206	translation-l

Fig. 25 processing

Within the **response** relation types, we have the **acknowledgement** relations, which show understanding of the previous utterance:

2) **ack**: demonstrates very general receipt of previous utterance, underspecified, and/or ambiguous as to whether the previous utterance was fully understood or will be acted upon or agreed to (Fig. 26). Cases where there are several, plausible interpretations, such as ack-understand or ack-wilco, should be annotated as a generic ack.

ID#	Timestamp	Commander	DM->CMD	DM->RN	RN	Transaction	Antecedent	Relation
26	018-08-06 14:48:58.34	move to the object <pause> on your left				6		
27	018-08-06 14:49:05.00		processing. . .			6	26	processing
28	018-08-06 14:49:12.66		If you describe an object, you can help me locate it. <beep>			6	26	missing-info
29	018-08-06 14:49:15.19	it's a yellow cone				6	28	clar-repair
30	018-08-06 14:49:19.32		ok			6	29	ack
31	018-08-06 14:49:38.93		processing. . .			6	29*	processing
32	018-08-06 14:49:50.85			move forward to yellow cone		6	29*	translation-r

Fig. 26 ack

3) **ack-understand**: expresses or shows understanding without commitment to action or agreement (Fig. 27). Includes repetitions of what was said, affirmative cue words like “yes” or “uh-huh”.

ID	Timestamp	Commander	DM->CMD	DM->RN	RN	Transaction	Antecedent	Relation
94	500.44	<pause> uh rotate ninety le <left>				17		
95	506.82		ok, I think I got it.			17	94	ack-understand
96	510.25			turn left 90 degrees		17	94	translation-r
97	511.19		I will turn left 90 degrees			17	94	ack-wilco

Fig. 27 ack-understand

4) **ack-unsure**: acknowledgment of the understanding of a command, expressing uncertainty about whether it can/will be done (Fig. 28). Not clearly an ack-cant or ack-try, but also distinct from an ack-understand because of some explicit statement of doubt about possibility or future action.

ID	Timestamp	Commander	DM->CMD	DM->RN	RN	Transaction	Antecedent	Relation
90	625.63	okay can you move at least fifty feet forward				13		clar-repair
91	630.85		processing. . .			13		processing
92	658.19			move forward 50 feet		13		translation-r
93	667.45		I'm not sure if I can move that far forward. <beep>			13		ack-unsure
94	669.44		I will move forward as far as I can, ok? <beep>			13		ack-try

Fig. 28 ack-unsure

5) **ack-try:** acknowledgment of a command and promise to try to do it (explicitly falling short of guaranteeing success) (Fig. 29).

ID	Timestamp	Commander	DM->CMD	DM->RN	RN	Transaction	Antecedent	Relation
90	625.63	okay can you move at least fifty feet forward				13		clar-repair
91	630.85		processing. . .			13		processing
92	658.19			move forward 50 feet		13		translation-r
93	667.45		I'm not sure if I can move that far forward. <beep>			13		ack-unsure
94	669.44		I will move forward as far as I can, ok? <beep>			13		ack-try

Fig. 29 ack-try

6) **ack-wilco**: acknowledgment of a command and promise to do it in the future (Fig. 30). A simple “ok” may be interpreted as an ack-wilco if there is no reason to doubt that the DM is going to perform the action and indeed the DM does subsequently translate the command (Fig. 31).

ID#	Timestamp	Commander	DM->CMD	DM->RN	RN	Transaction	Antecedent	Relation
143	018-08-06 15:01:08.39	move forward two feet				26		
144	018-08-06 15:01:14.45			move forward 2 feet		26	143	translation-r
145	018-08-06 15:01:15.08		I will move forward 2 feet			26	143	ack-wilco
146	018-08-06 15:01:17.79				done	26	144	ack-done
147	018-08-06 15:01:19.40		done			26	146	translation-l

Fig. 30 ack-wilco

ID#	Timestamp	Commander	DM->CMD	DM->RN	RN	Transaction	Antecedent	Relation
39	018-08-06 14:50:39.83	move to the brown object <pause> behind the cone				8		
40	018-08-06 14:50:44.58		ok			8	39	ack-wilco
41	018-08-06 14:51:03.47			move forward to stairs side crate		8	39	translation-r
42	018-08-06 14:51:12.73				done	8	41	ack-done
43	018-08-06 14:51:14.30		done			8	42	translation-l

Fig. 31 ack-wilco showing "ok"

7) **ack-doing**: acknowledgment that the speaker understands the command and it is underway (Fig. 32).

8) **ack-done**: acknowledgment that a command or prior planned act has been completed successfully (Fig. 32).

ID	Timestamp	Commander	DM->CMD	DM->RN	RN	Transaction	Antecedent	Relation
12	136.47	move forward three feet				3		
13	140.49	send picture				3	12	continue
14	146.82			move forward 3 feet, image		3	13*	translation-r
15	150.09		executing...			3	13*	ack-doing
16	155.75				image sent	3	14	ack-done
17	158.83		sent			3	16	translation-l

Fig. 32 ack-doing (ID 15 "executing") and ack-done (ID 16 "image sent")

9) **ack-cant**: an expression that the previous command was understood but cannot be executed (Fig. 33).

ID	Timestamp	Commander	DM->CMD	DM->RN	RN	Transaction	Antecedent	Relation
111	558.96	<pause> can you <small pause> go through <small pause> that opening				20		
112	558.99		done			19	110	translation-l
113	566.63		processing. . .			20	111	processing
114	576.27		no			20	111	ack-cant
115	577.92		There's an obstruction preventing me from doing that. <beep>			20	114	continue

Fig. 33 ack-cant

10) –**partial**: any ack commands can be qualified with "partial" if only part of the antecedent is acknowledged explicitly to that degree (generally there will be an implicit acknowledgement to a different degree) (Fig. 34).

ID	Timestamp	Commander	DM->CMD	DM->RN	RN	Transaction	Antecedent	Relation
138	1199.97			move back three feet, turn 90 degrees right, then another 30 degrees right, photo		19	137*	translation-r
139	1203.85		executing...			19	137*	ack-doing
140	1236.4				done	19	138	ack-done-partial
141	1241.39				image sent	19	138	ack-done-partial

Fig. 34 -partial

Within the **response** relations, we also have **clarification** relations. These relations indicate, resolve, or attempt to resolve problems in interpreting a prior utterance:

1) **nack**: indicate that the antecedent could not be understood well enough to act on, but not explicitly requesting action (e.g., "no copy" or "I don't understand") (Fig. 35).

ID	Timestamp	Commander	DM->CMD	DM->RN	RN	Transaction	Antecedent	Relation
29	134.03	can you go that way				7		
30	137.97		processing. . .			7	29	processing
31	156.71		Hmm, I'm not sure what you would like me to do. <beep>			7	29	nack

Fig. 35 nack

2) **missing-info**: indicate a specific part of the antecedent was not interpretable well enough to act on, but not requesting further action (e.g., "I don't know which object you are referring to") (Fig. 36). The other party has the option to clarify-repair or move on and do something else.

ID	Timestamp	Commander	DM->CMD	DM->RN	RN	Transaction	Antecedent	Relation
176	856.64	<pause> robot proceed to the doorway ahead of you				23		
177	869.51		processing. . .			23	176	processing
178	886.96		I see more than one doorway. <beep>			23	176	missing-info
179	888.54	<notification noise> the doorway closest to you				23	178	clar-repair

Fig. 36 missing-info

3) **req-clar**: request for clarification; indicate that something in the prior utterance was not clear and ask the other speaker to do something about it, such as answer a question or confirm a possibility (Fig. 37).

ID	Timestamp	Commander	DM->CMD	DM->RN	RN	Transaction	Antecedent	Relation
3	23.25	move forward five feet				2		
4	28.87	turn south thirty degrees				2	3	continue
5	43.26			move forward 5 feet, turn south 30 degrees		2	4*	translation-r
6	84.99		I'm not sure what you mean by turn south 30 degrees. Should I turn further to the south by 30 degrees?			2	4	req-clar
7	92.51	yes send picture				2	6	clar-repair

Fig. 37 req-clar

4) **clar-repair**: provide other-initiated self-repair to a prior utterance, after prompting by another (if unprompted this would be a correction) (Fig. 38).

ID	Timestamp	Commander	DM->CMD	DM->RN	RN	Transaction	Antecedent	Relation
30	252.39	turn ninety				6		
31	254.33		done			5	29	translation-I
32	256.61	send picture				6	30	continue
33	266.1		Which way should I turn 90 degrees?			6	32*	req-clar
34	267.93	ninety degrees right				6	33	clar-repair

Fig. 38 clar-repair

5) **req-repeat**: request to repeat a prior utterance (Fig. 39).

6) **clar-repeat**: provide other-initiated repetition, after prompting to repeat with a req-repeat (Fig. 39).

ID	Timestamp	Commander	DM->CMD	DM->RN	RN	Transaction	Antecedent	Relation
58	265.11	robot turn right eighteen degrees				12		
59	277.67		Can you repeat that? <beep>			12	58	req-repeat
60	279.68	robot turn right <pause> thirteen degrees				12	59	clar-repeat

Fig. 39 req-repeat and clar-repeat

4) **req-done**: request for more information about whether the task has been completed successfully. Something has been done in response to a prior command, but the speaker is not confident enough that the action is correct, and therefore do not report an ack-done (Fig. 40).

5) **clar-done**: provide (other-initiated) clarification about whether the action completed was to the speaker's satisfaction, generally in response to a req-done utterance (Fig. 40).

ID#	Timestamp	Commander	DM->CMD	DM->RN	RN	Transaction	Antecedent	Relation
284	14:55:54.46	can you turn maybe thirty degrees to your right				43	282	continue
285	14:55:56.76	and then go back through that door				43	284	continue
286	14:56:01.98		processing. . .			43	285*	processing
287	14:56:10.23			turn right 30 degrees		43	284	translation-r
288	14:56:12.40			then. . .		43	287	link-next
289	14:56:16.00			move into Stairwell hallway		43	285	translation-r
290	14:56:18.58		executing. . .			43	285*	ack-doing
291	14:56:26.16				done	43	289*	ack-done
292	14:56:28.48		Did I successfully do what you asked? <beep>			43	285*	req-done
293	14:56:29.58	yes				43	292	clar-done

Fig. 40 req-done and clar-done

Within the **response** relations, we also have **question-response** relations. For these, the antecedent is a question (information request), and the response indicates understanding, not a clarification, and some attempt to address the question:

1) **answer**: answer a question (Fig. 41). If something is a response to a request for clarification or repair, then even if the response is an answer to a kind of question, it should be annotated as clar-repair or clar-repeat, respectively, not "answer".

ID	Timestamp	Commander	DM->CMD	DM->RN	RN	Transaction	Antecedent	Relation
37	281.06	also do you speak any other languages				5	36	continue
38	305.46		I do not speak. I can recognize some English words.			5	37	answer

Fig. 41 answer

2) **nar**: a non-answer-response that addresses a question without providing an answer (e.g., explains why an answer will not be given, the question is not

relevant, or a helpful suggestion of how the requested information might be arrived at) (Fig. 42).

ID	Timestamp	Commander	DM->CMD	DM->RN	RN	Transaction	Antecedent	Relation
73	480.62	do you think there's uh some type of container that a soldier might be able to use to move things				12		
74	490.49		Hmm. . .			12		processing
75	506.82		I see objects all around me; I need your help to decide which are important. <beep>			12		nar
76	509.24		What do you think? <beep>			12		req-clar

Fig. 42 nar

Within the **response** relations, we also have **offer-accept/reject** relations. For these, the antecedent is a request by the speaker to do an action, and the response indicates acceptance or rejection of the proposed action:

1) **offer**: an offer by the speaker to perform an action (Fig. 43).

2) **offer-accept**: a response to an offer accepting the proposed action (Fig. 43).

ID#	Timestamp	Commander	DM->CMD	DM->RN	RN	Transaction	Antecedent	Relation
129	10:34:14.93	can you tell me the approximate size of the objects that you're looking at				17		
130	10:34:22.05		Hmm. . .			17	129	processing
131	10:34:31.20		I'm not sure.			17	129	answer
132	10:34:52.75		I think you are more familiar with the objects than I am.			17	131	continue
133	10:35:02.66		I can move to take a good picture of an object that you are interested in. <beep>			17	132	offer
134	10:35:08.84	yeah that'd be great				17	133	offer-accept

Fig. 43 offer and offer-accept

3) **offer-reject**: a response to an offer rejecting the proposed action, for example “no thank you”.

4) **reciprocal**: a response that indicates the responder is expressing the same or same or similar content as the prior speaker in the antecedent (e.g., “hello” in response to “hello” or an explicit “me too”) (Fig. 44).

ID	Timestamp	Commander	DM->CMD	DM->RN	RN	Transaction	Antecedent	Relation
5	19.37		Hello! <beep>			2		
6	20.6	hello				2	5	reciprocal

Fig. 44 reciprocal

5) **3feedback**: a third turn feedback response to a response (e.g., "thanks" after acknowledgement of an action or an evaluation after the answer to a question) (Fig. 45).

ID	Timestamp	Commander	DM->CMD	DM->RN	RN	Transaction	Antecedent	Relation
19	143.01	move forward to the second doorway				4		
20	149.62		processing. . .			4	19	processing
21	157.67			move to alley Door 2		4	19	translation-r
22	163.7		moving. . .			4	19	ack-doing
23	181.46				alright done	4	21	ack-done
24	183.39		done			4	23	translation-l
25	186.19	okay				4	24	3feedback

Fig. 45 3feedback

6) **other**: a response that does not fit into one of the other categories (Fig. 46). All responses other than clarifications will indicate acknowledgement of understanding of the antecedent, but answer, reciprocal response, third turn feedback, and other response will also indicate some other function as well.

ID	Timestamp	Commander	DM->CMD	DM->RN	RN	Transaction	Antecedent	Relation
231	813.66	r rotate				40		
232	813.91		I'm facing a wall now. <beep>			40	232	other
233	815.23		Did I successfully do what you asked? <beep>			40	231	req-clar
234	818.38	yes				40	233	clar-repair

Fig. 46 other

5. Project-Specific Annotation Steps

5.1 Improving Annotation Speed in Excel via Relation Shortcuts

For this project, TU, antecedent, and relation annotation is conducted in an Excel file, and the process can be tedious. For Experiment 3 annotation, the two methods for inputting relations are 1) typing the full name of the relation in the appropriate cell or 2) clicking on the cell to select a relation from a drop-down list. The latter method is possible if Data Validation settings are used to ensure that the data in the relations column match a source list of the possible relations. For Experiment 3, we placed the source list of relations on the second sheet of a workbook.

In practice, the annotation of TUs can be done quickly. In comparison, the annotation of relations has been limited by the speed at which the annotator can input data using one of the two methods discussed in the preceding paragraph. To overcome that limitation and improve the speed of annotation, a method was developed to annotate relations using shortcuts.

In this method, the annotator simply types shortcuts in the Shortcut column of an annotation sheet. The Relation column then populates with the relation that is associated with the shortcut. Figure 47 depicts an example. In ongoing annotation,

these shortcuts will be populated in the aligned spreadsheets prior to distribution for annotation.

ID#	Times	Commander	DM->CMD	DM->RN	RN	Tran	Antec	Relation	Shortcut
123	:49.51	proceed three feet				20			
124	:51.79	pivot fifty degrees right				20		continue	con
125	:56.73		processing. . .			20		processing	pr
126	:59.97			move forward 3 feet		20		translation-r	tr
127	:02.23		executing. . .			20		ack-doing	doing
128	:04.08			then. . .		20		link-next	x
129	:09.37			turn right 30 degrees		20		translation-r	tr
130	:13.17				done	20		ack-done	done
131	:14.22		done			20		translation-l	tl
132	:15.11	send me a picture				21			
133	:17.65			send image		21		translation-r	tr
134	:18.32				sent	21		ack-done	done
135	:19.12		sent			21		translation-l	tl

Fig. 47 Example of annotation using relation shortcuts. The annotator types the shortcuts in the rightmost column, and the appropriate relation is then automatically populated in the Relation column.

Table 3 contains a list of all shortcuts and their associated relations. The shortcuts were selected to maximize brevity, differentiability, and typing speed. Consider that a commonly-seen sequence might involve a **translation-r-partial** followed by a **link-next** followed by a **continue**. Instead of typing out the relations or clicking through drop-downs to select the relations, the user can enter **trp**, **x**, and **con** in successive cells. While there are likely better ways to optimize this process, piloting has suggested that this approach can significantly streamline annotation: ideal, uncomplicated transcripts can be annotated for TUs and relations in about 10 min at the fastest, though transcripts are often less than ideal and annotation of antecedents takes additional time.

Table 3 List of relation shortcuts

Shortcut	Annotation label	Relation
con	continue	continue
cor	correction	correction
x	link-next	link-next
sum	summarization	summarization
tl	translation-l	translation-left
tr	translation-r	translation-right
tlp, trp	translation-l-partial, translation-r-partial	partial translation
qu	quotation	quotation
com	comment	comment
pr	processing	processing
ack	ack	acknowledge (general, underspecified)
und	ack-understand	acknowledge understand
uns	ack-unsure	acknowledge unsure
try	ack-try	acknowledge try
wil	ack-wilco	acknowledge will comply
doing	ack-doing	acknowledge doing
done	ack-done	acknowledge done
cant	ack-cant	acknowledge can't
***p. ex: undp, unsp, tryp, etc.	ack-understand-partial, ack-unsure-partial, etc.	partial acknowledgment
nack	nack	negative acknowledgment
mis	missing-info	missing information
req	req-clar	request clarification
clar	clar-repair	clarification repair
rp	req-repeat	request repeat
clp	clar-repeat	clarification repeat
rd	req-done	request done
cld	clar-done	clarification done
ans	answer	answer
nar	nar	non-answer response
off	offer	offer
ofa	offer-accept	offer accept
ofr	offer-reject	offer reject
rec	reciprocal	reciprocal response
3	3feedback	third-turn feedback
other	other	other response

5.2 After the Files Are Validated

The following are project-specific steps describing the pipeline after the files have been annotated and subsequently validated:

1) For the cases that are easily fixable, the original annotator should make the corrections. If all the changes were made to a file (i.e., there were no borderline cases), then please check it into the SVN (subversion control system for tracking changes to shared files) and mark it as completed on the tracking spreadsheet (e.g., exp3_annotation_tracking.xlsx).

2) If the file contains a borderline case, the original annotator should make all the other easily fixable changes to the file, then check it into the SVN without marking it as completed on the tracking spreadsheet. The annotator should take a screenshot of the problematic annotation, then create a new tab in the issues tracking sheet (dialogue_structure_issues_tracking.xlsx).

3) For the sake of keeping the annotated files themselves clean, remove the easily fixable validation notes from the annotated file after making the corrections, but leave the notes for the borderline cases until resolved.

4) Outstanding issues in the issue tracking sheet are discussed either at the annotation meetings or in the MatterMost online chat system. The validator can provide their comments about why they think it was a borderline instance, and the team will work towards a resolution together.

5) While waiting for these borderline issues to be resolved, annotators and validators should proceed with annotating and validating new files as they have the time. Our goal is to resolve the borderline instances quickly so that there are not a lot of issues floating around that might make it more difficult for the annotators to keep track of all the files they are in charge of.

5.3 After Borderline Cases Are Resolved

The following are project-specific steps describing the process for resolving annotation borderline cases:

1) Have the original annotator make the agreed-upon corrections and commit them to the SVN.

2) Let annotation manager know when it is done so they can mark it on the verification sheet (e.g., exp3_annotation_tracking.xlsx).

3) Take a screenshot of the corrected annotation on the issues tracking sheet (e.g., dialogue_structure_issues_tracking.xlsx).

4) Add the screenshot to the issues resolved document (e.g., dialogue_structure_issues_resolved.docx) and provide a brief summary of the decision. The format of the previous decisions can be used as examples.

5) Delete the tab for that issue from the issues tracking sheet. *Once a sheet is deleted from an Excel file, one cannot recover it using Undo or Ctrl+Z, so be sure to be ready before deleting the sheet.*

6. Conclusion and Recommendations

In this report, we have outlined procedures for providing dialogue structure annotations over transcribed and time-aligned, human–robot dialogue data. This protocol has been under development and used to annotate over 80 human–robot interactions in the Bot Language Experiments 1–3.

7. References

DeVault D, Artstein R, Benn G, Dey T, Fast Ed, Gainer A, Georgila K, Gratch J, Hartholt A, Lhommet M, et al. SimSensei Kiosk: a virtual human interviewer for healthcare decision support. In: Proceedings of the 2014 International Conference on Autonomous Agents and Multi-agent systems. International Foundation for Autonomous Agents and Multiagent Systems; 2014 May. p. 1061–1068.

Edelsky C. Who's got the floor? Language in Society. 1981;10(3):383–421.

Lukin SM. Gervits F, Hayes CJ, Leuski A, Moolchandani P, Rogers JG 3rd, Traum D. Scoutbot: a dialogue system for collaborative navigation. Ithaca (NY): Cornell University; 2018. http://arXiv:1807.08074.

Marge M, Bonial C, Foots A, Hayes C, Henry C, Pollard KA. Artstein R, Voss C, Traum D. Exploring variation of natural human commands to a robot in a collaborative navigation task. In: Proceedings of the First Workshop on Language Grounding for Robotics; 2017 Aug. p. 58–66.

Marge M, Bonial C, Pollard KA, Artstein R, Byrne B, Hill SG, Voss C, Traum D. Assessing agreement in human-robot dialogue strategies: a tale of two wizards. In: International Conference on Intelligent Virtual Agents; 2016 Sep; Cham, Switzerland. Springer. p. 484–488.

Traum D, Henry C, Lukin S, Artstein R, Gervits F, Pollard K, Bonial C, Lei S, Voss C, Marge M. Dialogue structure annotation for multi-floor interaction. In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC-2018); 2018 May.

Appendix. Annotation Questions and Resolutions
(Author: AL Baker)

A.1 Introduction to this Appendix

During the course of annotation, existing guidelines and practices are sometimes challenged by new situations; indeed, no annotation schema survives first contact with participants. For Experiment 3, we collected information about edge cases, ambiguous situations, and scenarios that appeared to fall outside of current guidelines. After making collective decisions about how to address those situations, we recorded our decisions and justifications. The goals of recording this information were to increase the consistency of our annotation, and to improve the ability of the annotation schema to accommodate the variety of interactions encountered in our data.

The following sections in this Appendix contain our decisions on several situations. Section A.2 contains a few clarifications on the annotation guidelines. Section A.3 consists of potentially ambiguous situations encountered in the data transcripts. Each situation is accompanied by a question that illustrates the issue, an answer that discusses our decision, and one or more screenshots of the situation. This Appendix therefore serves as a complement to the annotation schema.

A.2 Miscellaneous Rulings and Clarifications

1) With **missing-info**, the Dialogue Manager (DM) points out an inability to do the command *without* motivation of any response from Commander (CMD). **Req-clar** points out an inability to do the command and explicitly motivates a response from CMD.

2) A line that only contains "…" is usually tagged **link-next**, but rarely you'll see that it takes the place of the DM saying "processing…", so those would be annotated as **processing**.

3) Instances like Lines 18–20 in the following image are annotated as a sequence of **translation-r-partial – link-next – continue**. The antecedent for the **continue** is the **link-next**. This is a fairly common sequence that is seen when a single command is translated in multiple parts. This sequence can be extended with additional links and continues, as can be seen in Q11 later in this section.

16	1	15:24:21.23	take a picture east				3	
17	1	15:24:27.37		processing. . .			3	16
18	1	15:24:36.06			turn to face East		3	16
19	1	15:24:37.56			then. . .		3	18
20	1	15:24:38.51			send image		3	19
21	1	15:24:40.90				uh done	3	20*
22	1	15:24:43.34		done, sent			3	21

processing
translation-r-partial
link-next
continue
ack-done
translation-l

4) In some cases a situation may be unclear on whether it should be annotated using **offer** and **offer-accept** versus using **req-clar** and **clar-repair.** Compare the following two examples.

   In the following image, Line 181 should be **offer** and Line 182 should be **offer-accept**. All the Lines in the image would be within TU 26.

179	2018-08-01 10:38:04.27	is that a plant in front of you				26		
180	2018-08-01 10:38:12.84		I think so.			26	179	answer
181	2018-08-01 10:38:16.07		I can move to take a good picture of an object that you are interested in. <beep>			26	180	continue
182	2018-08-01 10:38:21.35	do that				26	181	3feedback
183	2018-08-01 10:38:25.49		processing. . .			26	182	processing
184	2018-08-01 10:38:32.58			move to Alley plant		26	182*	translation-r
185	2018-08-01 10:38:39.27		moving. . .			26	182	ack-doing
186	2018-08-01 10:38:45.16				uh done	26	184	ack-done
187	2018-08-01 10:38:47.48		done			26	186	translation-l

   Contrast this with Lines 122 and 123 in the following image. For Line 122, **req-clar** fits better because the CMD is asking for a clarification of the command specified in 121. In contrast, Line 181 in the previous example involves a brand new offer rather than a request for clarification of the command.

118	2018-08-01 10:31:32.83	what do you see				17		
119	2018-08-01 10:31:33.24		done			16	117	translation-l
120	2018-08-01 10:31:44.33		I'm not sure.			17	118	ack-unsure
121	2018-08-01 10:31:47.35	send me a picture				17	120	clar-repair
122	2018-08-01 10:31:49.55		Would you like me to send a picture? <beep>			17	118	req-clar
123	2018-08-01 10:31:53.78	yes				17	122	clar-repair
124	2018-08-01 10:32:00.38			send image		17	21*	translation-r
125	2018-08-01 10:32:00.81				sent	17	124	ack-done
126	2018-08-01 10:32:02.51		sent			17	125	translation-l

A.3 Annotation Resolutions

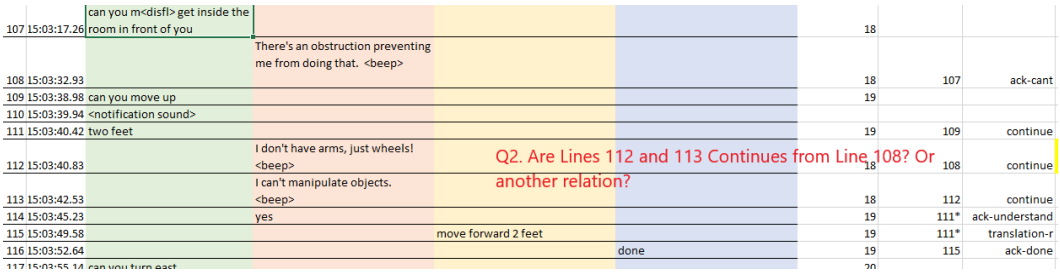
Q1. New transaction unit (TU) in response to an answer

91	15:02:09.52	can you move to the co<disfl> to the cone		Q1. Is Line 93 part of the same TU as 91 and 92?		15		
92	15:02:17.67		yes			15	91	answer
93	15:02:21.90	move to the cone				15	92	3feedback
94	15:02:29.65			move to Alley cone		15	93	translation-r
95	15:02:31.36		executing. . .			15	93	ack-doing
96	15:02:46.62				done	15	94	ack-done
97	15:02:48.70		done			15	96	translation-l

Is Line 93 part of the same TU as Lines 91 and 92?

Lines 91 and 92 should be treated as a question-**answer** in isolation, and Line 93 starts a new TU, which is the execution of a command (which happens to be what the person was asking about prior). If the DM had instead said, "yes, should I move to it" that would not necessarily be annotated as an "answer". Instead, it would be annotated as **offer**, and if the CMD responded "yes". then that response would be annotated as **offer-accept**, and all would be part of the same TU.

Q2. Disjointed continues

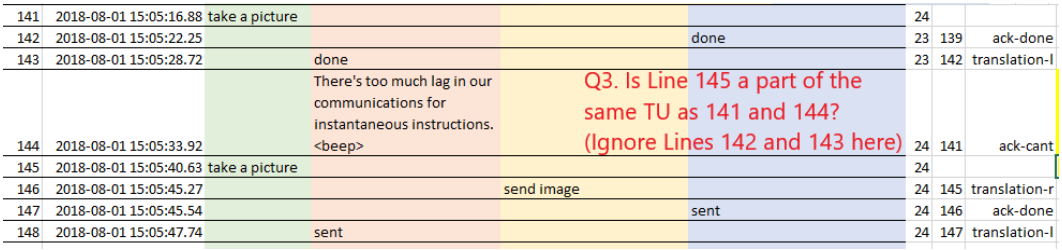


107	15:03:17.26	can you m<disfl> get inside the room in front of you				18			
108	15:03:32.93		There's an obstruction preventing me from doing that. <beep>			18	107	ack-cant	
109	15:03:38.98	can you move up				19			
110	15:03:39.94	<notification sound>							
111	15:03:40.42	two feet				19	109	continue	
112	15:03:40.83		I don't have arms, just wheels! <beep>			18	108	continue	
113	15:03:42.53		I can't manipulate objects. <beep>			18	112	continue	
114	15:03:45.23		yes			19	111*	ack-understand	
115	15:03:49.58			move forward 2 feet		19	111*	translation-r	
116	15:03:52.64				done	19	115	ack-done	
117	15:03:55.14	can you turn east				20			

Are Lines 112 and 113 ***continues*** *from Line 108, or would they be another relation?*

Line 108 should be **ack-cant** and Lines 112–113 should be **continue**. **Continue** "adds more content" (definition from the guide) to the **ack-cant**. The delay that is seen in the transcript is because the DM was probably looking for the right buttons and the CMD got the information they needed from the first DM response on Line 108.

Q3. Atypical summarization and DM mistaken response



141	2018-08-01 15:05:16.88	take a picture				24			
142	2018-08-01 15:05:22.25				done	23	139	ack-done	
143	2018-08-01 15:05:28.72		done			23	142	translation-l	
144	2018-08-01 15:05:33.92		There's too much lag in our communications for instantaneous instructions. <beep>			24	141	ack-cant	
145	2018-08-01 15:05:40.63	take a picture				24			
146	2018-08-01 15:05:45.27			send image		24	145	translation-r	
147	2018-08-01 15:05:45.54				sent	24	146	ack-done	
148	2018-08-01 15:05:47.74		sent			24	147	translation-l	

Is Line 145 part of the same TU as Lines 141 and 144?

Since Line 141 isn't acted upon, there's a seemingly random response from the DM, and the command from Line 141 is repeated, so these should be treated as one TU with Line 145 as **summarization** of Line 141. In absence of being able to read the DM's mind at the time, we have to assume that Line 144 was not a mistake, and in that case how this scenario was treated in the image seems to be correct—with Line 144 as **ack-cant** with respect to the first request to take a picture.

Q4. Missing-info versus req-clar

160	2018-08-01 15:06:49.33	take a picture northeast of you			27		
161	2018-08-01 15:06:59.87		Hmm. . .		27	160	processing
162	2018-08-01 15:07:05.44		You can tell me to turn a number of degrees or to face something. <beep>	Q4. Is this line a missing-info, or a req-clar? I think missing-info but not sure..	27	160	missing-info

*Is Line 162 a **missing-info** or a **req-clar**?*

In the past, this DM utterance has been annotated as **req-clar**. **Missing-info** has been instances like "I see more than one doorway" or "I'm not sure where or when to stop turning". **Missing-info** is not an actual request for information from the CMD (unlike the **req-clar**) and more of a comment that the DM cannot complete the action given the provided information.

Missing info conveys the missing parameter but does not necessarily request that the CMD do anything or respond in any particular way, while **req-clar** DOES motivate the CMD to clarify something. These "You can tell me..." type responses are borderline because, the way they are phrased, they don't necessarily directly motivate any particular type of response (the way, for example, "How far should I turn to the left?" would) but in context they do tend to be understood and interpreted as a request for the CMD to rephrase the past instructions with a number of degrees or landmark destination. Given that we are always trying to interpret utterances in context according to how they seem to be understood in the larger dialogue, it does seem appropriate to treat these as **req-clar**.

Q5. Misalignment: rows look out of order

45	14:59:34.45	take a picture				5			Q5. out of order?
46	14:59:37.14				sent	5	45	ack-done	
47	14:59:37.15			send image		5	45	translation-r	
48	14:59:38.92		sent			5	46	translation-l	

This example depicts lines that appear to be out of order. How should these situations be handled?

The audio for the screen recorder for this particular CMD was unfortunately corrupt, so we can't confirm that the RN spoke "sent" before or after the DM requested "send image." However, the screen recorder did show us that the RN sent the image after the DM requested "send image", so we can perhaps infer that the intention of the RN was to say "sent" when they actually sent the picture. Therefore, we can *treat this as a case of misalignment and swap the two rows in the spreadsheet and adjust their line IDs appropriately.*

Q6. Disjointed continues 2

182.	15:08:27.34	what color is the cone				31				
183.	15:08:38.10		Hmm. . .			31	182	processing		
184.	15:08:45.35		I'm not sure.			31	182	answer		
185.	15:08:46.11		. . .			31	184	link-next		
186.	15:08:51.88	can you turn <pause>				32				
187.	15:08:53.27	west				32	186	continue	Q6.	
188.	15:08:53.37		I think you are more familiar with colors than than I am.			31	185	continue	are these all continues?	
189.	15:08:53.73	and take a picture				32	187	continue		
190.	15:08:57.00		processing. . .			32	189	processing		
191.	15:09:00.88			turn to face West		32	187*	translation-r		
192.	15:09:02.00			then. . .		32	191	link-next		

*Are Lines 187–189 all annotated as **continue**?*

The annotation in this image is *correct*; those are indeed all **Continue**. The key here is to make sure that the TUs are annotated correctly, as they are important for distinguishing conversation threads.

Q7. Summarization, correction, and clar-repair

211.	15:10:39.43	move forward until i say stop				34				
212.	15:10:45.86		There's too much lag in our communications for instantaneous instructions. <beep>			34	211	ack-cant	Q7.	
213.	15:11:04.36	proceed forward				34	211	summarization	Is this correction or summarization?	
214.	15:11:05.46		You can tell me to turn a number of degrees or to face something. <beep>			34	212	continue		
215.	15:11:09.19		How far forward should I go? <beep>			34	213	req-clar		
216.	15:11:14.74	ten inches				34	215	clar-repair		
217.	15:11:22.28		Hmm. . .			34	216	processing		
218.	15:11:31.44		I will move forward 1 foot			34	216	ack-wilco		
219.	15:11:33.91			move forward 1 foot		34	216	translation-r		

Is Line 213 a correction or a summarization?

This is a case that could be handled with the ERR tag—so Line 213 "proceed forward" would be a new TU, Line 214 is ERR with no antecedent/relation, and Line 215 is **req-clar** of Line 213.

Line 213 is a new instruction that may be trying to get at the same higher level intention [as Line 212], but with a new command, not trying to restate (**summarize**) or fix (**correction** or more properly **clar-repair** of Line 212). Line 214 looks like an error on the DM part—probably pressing the wrong button. It's unclear if it is meant as a continuation of Line 212, or a response to Line 213, but it is inappropriate in either case, since it is talking about turning rather than moving forward. Line 215 […] is a proper **req-clar** response to Line 213.

Corrections are explicitly done to fix a previous command. The commands may look similar, but without explicit evidence of restatement or attempt to fix the previous command, they initiate new TUs.

Q8. Sequential occurrences of translate-r

128	1	15:31:32.43	turn north				17		
129	1	15:31:35.05	er				17	128	continue
130	1	15:31:35.83	and move up two feet				17	129	continue
131	1	15:31:50.33		I will turn to face North			17	128	ack-wilco
132	1	15:31:53.01			turn to face North		17	128	translation-r
133	1	15:32:17.18			move to front wall ahead		17	130	Q8. translation-r
134	1	15:32:20.56				uh done	17	133*	ack-done
135	1	15:32:21.91		I moved forward as far as I could.			17	134	translation-l

*It is unusual to see two translations in a row without a link-next in between, so are Lines 132 and 133 both **translate-r**?*

There is a one-to-one relationship between the first line of the instructions and the first **translate-r**, then there is another one-to-one relationship between the second line of the instructions and the second **translate-r**, so yes, Lines 132 and 133 are both annotated as **translate-r** with distinct antecedents.

Q9. Handling overlapping TUs

239	1	15:39:01.35	take a good picture of the calendar		Q9. How to handle these TUs?		32		
240	1	15:39:11.92		ok			32	239	ack
241	1	15:39:27.16	turn left twenty degrees				33		
242	1	15:39:29.00	and take a picture				33	241	continue
243	1	15:39:32.67			send image of calendar		32	239	translation-r
244	1	15:39:35.34		There's too much lag in our communications for instantaneous instructions. <beep>			33	242*	ack-cant
245	1	15:39:38.32				uh done and sent	32	243	ack-done
246	1	15:39:44.09		done, sent			32	245	translation-l
247	1	15:39:51.29	turn left thirty degrees				34		
248	1	15:39:52.60	and take a picture				34	247	continue

Lines 239–248 appear to depict overlapping communication. How should TUs be annotated when commands and messages overlap?

The TUs in this figure are correct because Line 244 is a shut-down of Lines 241–242 while the instructions in Line 239 are carried out.

Q10. Missing-info followed by clar-repair, and summarization

57	14:55:06.33	turn around				9		
58	14:55:07.61	and take a picture				9	57	continue
59	14:55:19.86		I'm unsure where or when I should stop turning. <beep>			9	57	missing-info
60	14:55:24.37	turn a hundred and eighty degrees				9	59	clar-repair
61	14:55:26.19	and take a picture				9	58	summarization
62	14:55:28.30		processing. . .			9	61*	processing
63	14:55:32.48			turn 180		9	60	translation-r
64	14:55:33.56			then. . .		9	63	link-next

124	14:59:47.72	turn around				17		
125	14:59:48.75	and drive ten feet back down the alleyway				17	124	continue
126	14:59:56.46		I'm not sure where or when to stop turning. <beep>			17	124	missing-info
127	15:00:03.53	turn seventy degrees to the left				17	126	clar-repair
128	15:00:05.32	and move forward ten feet				17	125	summarization
129	15:00:10.64		processing. . .			17	128*	processing
130	15:00:16.56			turn left 70 degrees		17	127	translation-r

*Can **clar-repair** have **missing-info** as an antecedent? And should Lines 61 and 128 be **continue** or **summarization**?*

Clar-repair can follow an antecedent **missing-info**. If another part of the command is later repeated after the **clar-repair** (e.g., Lines 61 and 128 in the two images above), those would be **summarizations** since they are repeating parts of the original instruction that were fully specified, not missing some gaps.

Q11. Long partial-link-continue blocks

88	14:57:12.98	turn around to the left				12		
89	14:57:14.32	and take a picture every ninety degrees				12	88	continue
90	14:57:18.93		processing. . .			12	89*	processing
91	14:57:37.27			turn left 90 degrees		12	89	nslation-r-partial
92	14:57:39.30			then. . .		12	91	link-next
93	14:57:40.53			send image		12	92	continue
94	14:57:43.39			turn left 90 degrees		12	93	continue
95	14:57:44.54			then. . .		12	94	link-next
96	14:57:45.74			send image		12	95	continue
97	14:57:47.06		turning. . .			12	89*	ack-doing
98	14:57:49.20			turn left 90 degrees		12	96	continue
99	14:57:50.76			then. . .		12	98	link-next
100	14:57:53.90			send image		12	99	continue
101	14:57:55.30			then. . .		12	100	link-next
102	14:57:57.18			turn left 90 degrees		12	101	continue
103	14:57:58.80			then. . .		12	102	link-next
104	14:57:59.69			send image		12	103	continue
105	14:58:15.54				done and sent	12	104*	

*The figure above indicates that a **translation-r-partial** was followed by what appears to be many **continues**. How should this situation be handled?*

This annotation is correct except Line 91's antecedent should be 89* like Lines 90 and 97, rather than just 89 (or 88). This figure demonstrates that partial-link-continue chains can be extended.

Q12. Ack-try in the form of a question

52	15:51:40.52	move west three feet				7		
53	15:51:51.03		Hmm. . .			7	52	processing
54	15:51:53.13		I will move forward as far as I can, ok? <beep>			7	52	ack-try
55	15:51:57.51	okay				7	54	answer uns
56	15:52:02.60			move to front wall ahead		7	54	translation-r

*If an **ack-try** by the DM is used as a question, how should a response from the CMD be annotated?*

As this figure demonstrates, Line 54 should be **ack-try** and Line 55 would be an **answer**. The antecedent for Line 56 should be 55*. While the CMD is not usually the one to answer questions, this schema fits the situation.

Q13. Vocal fillers and disfluencies/unintelligible messages

31	25:26.98	move back				5		
32	25:28.04	and take a picture				5	31	continue
33	25:43.30	mo<disfl>						
34	25:44.55		How far should I move back? <beep>			5	31	req-clar
35	25:45.44	five feet				5	34	clar-repair
128	31:32.43	turn north				18		
129	31:35.05	er						
130	31:35.83	and move up two feet				18	129	continue
131	31:50.33		I will turn to face North			18	128	ack-wilco
132	31:53.01			turn to face North		18	128	translation-r
133	32:17.18			move to front wall ahead		18	130	translation-r

[from exp3-P4-House1]

How should disfluencies and vocal fillers be annotated? Should they be annotated?

We do not have enough information to determine what Line 33 is. With more context, it might be a **summarize** or **correction**; however, we have no concrete way of knowing. Additionally, it does not add any new information. Therefore, we will remove Line 33 from the TU by not assigning it a TU and not giving it a relation or antecedent.

Line 129 does not provide any new information. Similarly, we suggest removing it from the TU by not assigning it a TU and not giving it a relation or antecedent.

If you judge that the purpose of an utterance is only to hold a turn, and not to convey new information, it can also be removed from the TU. For example, if the CMD says “okay” followed by “could you move…”. If the “okay” is not judged to be a response to anything, or a third turn feedback, then it could be that the CMD just said it to indicate they are starting their conversational turn. If that is the case, the “okay” line would be removed from the TU.

Q14. Questions from the DM as new TUs

156	29:58.24	how many orange cones <pause .31> do you see				24		
157	30:06.92		Hmm. . .			24	156	processing
158	30:13.00		I'm not sure.			24	156	answer
159	30:14.15		. . .			24	158	link-next
160	30:16.73		Two.			24	159	continue
161	30:20.07		What do you think? <beep>			25		
162	30:26.04	i think you are correct				25	161	answer
163	30:34.38		What should we do next? <beep>			26		
164	30:39.18	can you find any more doorways				27		
165	30:55.74		I need your help to find doorways.			27	164	answer
166	30:56.86		. . .			27	165	link-next
167	31:00.45	go forward twenty feet				28		

[from exp3-P7-Alley]

How should questions from the DM be handled? These cases are relatively uncommon.

Questions from the DM like the ones in Lines 161 and 163 would begin new TUs. Line 164 would also begin a new TU. Line 164 could be interpreted as a response to the question posed by the DM in Line 163, however 164 does not directly address what to do next. Therefore, the annotations in the figure are correct as-is.

Q15. Corrections from Commander

"Stop" is seen to be used in two ways by CMDs.

In the first way, it is used to revise or cancel the previously issued command. In the second way, it is used when the CMD has a new plan in mind. Usually, this is seen when the CMD is watching the robot move around the map, and decides that the robot should be instructed to do something different.

You would keep the stop command in the same TU if there's evidence that it is intended to revise the previous command. For example, you may see this in the timing of the message, where the CMD might issue a stop command before the DM has had a chance to translate anything to the RN. Or there may be linguistic cues in addition to "stop", such as some cues linking it back to the former command like "Oh, never mind, stop" or "oops, stop" or "that's not what I meant, stop".

In contrast, you would judge that the stop command begins a new TU when the commander appears to be reacting to something they see on the map. For example, if we don't have evidence that the commander is trying to revise the previous command, and it seems as though they see that the robot is adjacent to an area of interest on the map and want the robot to stop, then this is separate from the past instruction and is a new TU.

244	2018-08-03 14:47:05.18	go back to prior position				43		
245	2018-08-03 14:47:11.78		Can you rephrase that? <beep>			43	244	req-clar
246	2018-08-03 14:47:15.48	go back to <extended pause> prior point				43	245	clar-repair
247	2018-08-03 14:47:24.52		ok, I think I got it.			43	246	ack-understand
248	2018-08-03 14:47:33.58	stop				43	246	correction
249	2018-08-03 14:47:35.48	face uh southwest				44		
250	2018-08-03 14:47:39.22		ok			44	249	ack
251	2018-08-03 14:47:54.23			turn right 90 degrees		44	249	translate-r
252	2018-08-03 14:47:55.66		turning. . .			44	249	ack-doing

[from exp3-P96-house1]

This example would have Line 248 annotated as being within the same TU, as it appears to be an attempt to correct the previous command.

We have the commander trying twice to convey something, and it seems plausible that Line 248 is another attempt to convey their true intention in a clearer way. Part of the contextual evidence here is that the DM doesn't pass anything to the RN, so it's not as if the CMD is seeing some action take place and then deciding suddenly that they want to do something novel based upon what they're seeing (e.g., the robot comes close to a doorway opening on the map)—so it's definitely not an "inspired by the map" case. They've issued one command, but before it's even begun (which would be indicated by the DM passing information to the RN), they've changed their mind and issued another command, so the "stop" here makes reference to cancelling/correcting the previous command, not to stopping ongoing motion to do something new.

In short, you would annotate "stop" as a new TU if it seems that the CMD is telling the robot to stop based on its movement to a location. You would keep it in the same TU if it happens before the robot starts moving or if there is additional communication that suggests that it was meant to revise the previous command.

Q16. Interpreting ack-wilco-partial

258	2018-08-08 11:54:02.93	take a photo of looking south				45			
259	2018-08-08 11:54:10.06		I will turn to face South			45	258	ack-wilco	
260	2018-08-08 11:54:10.74			turn to face South		45	258	e-r-partial	
261	2018-08-08 11:54:12.53			then. . .		45	260	link-next	
262	2018-08-08 11:54:13.59			send image		45	261	continue	
263	2018-08-08 11:54:29.87				done and sent	45	262*	ack-done	
264	2018-08-08 11:54:30.96		sent			45	263	translate-l	

[from exp3-P91-house2]

*Is Line 259 an **ack-wilco-partial**, or an **ack-wilco?***

We decided that this will be an **ack-wilco** despite the fact that at face-value it seems that only part of the CMD instruction is acknowledged here. This decision was based on multiple observations from more of the transcript outside of this TU, in particular, how the DM handled other actions. The DM did not provide **ack-wilco** for other requests for photos. Additionally, the DM did translate the instruction.

- We would treat this as an **ack-wilco-partial** under the following circumstances:
- The CMD later clarified or reminded the DM to take the picture.
- The DM forgot to translate the take a picture.

Q17. Ambiguous translation-r at start of experiment

0	0	(calibrate)				1		
1	:56:11.50			participant is ready		1	0	translatio
2	:56:15.84		calibrating. . .			1	0	ack-doing
1	14:57:50.33	<X: alright>				X-CMD		

[from exp3-P4-Alley]

Line 1 appears to be an attempt to translate something but there's no visible antecedent. How should this be annotated?

The "participant is ready" seems like a mispress from the DM. The expected button press is "calibrate", but because the RN later responds to the instruction, we will treat it here as a **translation-r**. We suggest adding a note saying that this was interpreted by the RN as "calibrate".

Q18. Indicating multiple antecedents; one TU vs two TUs

[from exp3-P7-House1]

170	47:19.50	turn west				27		
171	47:21.71	and drive <pause .81> forward three feet				27	170	continue
172	47:28.87		processing. . .			27	171*	processing
173	47:35.49		I will turn to face West			27	170	ack-wilco
174	47:37.14		and. . .			27	173	link-next
175	47:40.50		I will move forward 3 feet			27	171	ack-wilco
176	47:44.30			turn to face West		27	170	translation-r
177	47:46.70			then. . .		27	176	link-next
178	47:52.19	take a picture				28		
179	47:53.99		I will move forward 3 feet			ERR		
180	47:56.77		executing. . .			27	171*	ack-doing
181	47:59.65		There's too much lag in our communications for instantaneous instructions. <beep>			28	178	ack-cant
182	48:04.14			send image		28	178	translation-r
183	48:05.18				done and sent	28	182*	ack-done
184	48:07.28		done, sent			28	183	translation-l

[from exp3-P8-House2]

Line 183 appears to refer to commands from two TUs. How should the TUs and antecedents be annotated in this scenario?

Note that this example includes *two* TUs that overlap. This situation, in which Line 183 responds to commands from both TUs, is extremely unusual but we need a way to link them correctly.

The TU for Line 183 should therefore be listed as [27,28].

The antecedent for Line 183 should be listed as [176,182].

Note: Excel will remove the comma from the cell if it is input as 176,182 (this would leave the cell value as 176182). The brackets prevent that removal.

Compare the situation above to one that contains only a single TU:

35	49:22.58	move forward two feet				6		
36	49:24.68	<no speech>						
37	49:25.29	face				6	35	continue
38	49:26.31	east				6	37	continue
39	49:26.84	and take a picture				6	38	continue
40	49:30.00		processing. . .			6	39*	processing
41	49:33.94			move forward 2 feet		6	35	translation-r
42	49:35.51			then. . .		6	41	link-next
43	49:36.77		moving. . .			6	35	ack-doing
44	49:38.34			turn to face East		6	38*	translation-r
45	49:39.61			then. . .		6	44	link-next
46	49:41.87		turning. . .			6	38*	ack-doing
47	49:43.80			send image		6	39	translation-r
48	49:48.48				done and sent	6	47*	ack-done
49	49:49.25		done, sent			6	48	translation-l

The annotation in this image is correct; no additional action is required here.

Q19. Should one question or command have multiple Response relations, or only a single relation with multiple Continues?

155	15:29:49.40	do you see any shovels around you				17		
156	15:29:51.16		done			16	154	translation-I
157	15:29:57.49		I'm not sure.			17	155	answer
158	15:30:05.14		I think you are more familiar with shovels than I am.			17	157	continue
159	15:30:11.36	take a picture at north east south and west				18		
160	15:30:12.28		I see objects all around me; I need your help to decide which are important. <beep>			17	158	continue
161	15:30:18.18		processing. . .			18	159	processing
162	15:30:21.67			turn to face North		18	159	nslation-r-partial
163	15:30:23.11			then. . .		18	162	link-next

304	15:55:57.00	are those the same shoes we saw before				45		
305	15:56:09.89		Hmm. . .			45	304	processing
306	15:56:16.13		I'm not sure.			45	304	answer
307	15:56:17.17	have we been in this room before				46		
308	15:56:17.23		. . .			45	306	link-next
309	15:56:22.80		I think you are more familiar with shoes than I am.			45	306	continue
310	15:56:26.04		no			46	307	answer
311	15:56:27.43		I don't think so.			46	310	continue

73	14:47:51.09	enter the doorway				13		
74	14:47:58.53		There's an obstruction preventing me from doing that. <beep>			13	73	ack-cant
75	14:48:01.35		I can't manipulate objects. <beep>			13	74	continue
76	14:48:04.11		I don't have arms, just wheels! <beep>			13	75	continue
77	14:48:07.24		Would you like me to send a picture? <beep>			13	73	continue
78	14:48:09.54	yes				13	77	answer
79	14:48:11.23			send image		13	78*	translation-r

[from exp3-P06 - P08]

For multi-message responses to commands or questions like the ones in the examples above, consider the DM's intent. Usually, the multiple responses all relate to the same intention (i.e., to give the CMD enough information about their request), so in these cases, the DM's responses will generally have a single Response relation (e.g., **answer**, **nar**, **nack**, and so on) followed by multiple **continue** relations. Therefore, the annotations in these images are correct.

•

Q20. “Ok:” ack vs. ack-wilco

73	14:54:30.23	move to the next doorway				15		
74	14:54:35.79		Which doorway? <beep>			15	73	missing-info
75	14:54:38.29	the doorway to your right				15	74	clar-repair
76	14:54:41.67		ok			15	75*	ack-wilco
77	14:54:47.22			move into Cleaning room		15	75*	translation-r
78	14:54:59.63				done	15	77	ack-done

126	14:59:07.84	move to the next doorway				23	123	summarization
127	14:59:11.11	southwest				23	126	continue
128	14:59:16.88		ok			23	127*	ack-wilco
129	14:59:24.10			move to Cleaning room - hall doorway		23	127*	translation-r
130	14:59:36.67				done	23	129	ack-done

How should “ok” be annotated if there is nothing else in that message?

If it’s clear that the DM’s “ok” means that they will do it, code as **ack-wilco** based on whether or not the DM carries out the instruction. In other words, “Ok” could be an **ack-wilco** if the DM eventually translates the message to the RN.

Instead, if the DM’s “ok” does not result in a clear translation, code as **ack**.

When deciding whether or not to use * for the “ok” line, determine whether the acknowledged command is fully specified, or if it is spread across multiple lines.

Q21. Translating poorly-worded commands: An unusual case

93	15:26:35.07	ahead north				20		
94	15:26:36.67	five feet				20	93	continue
95	15:26:38.33		ok			20	94*	ack
96	15:26:41.56			turn to face North		20	93	translation-r
97	15:26:43.38			then. . .		20	96	link-next
98	15:26:46.22			move forward 5 feet		20	94*	nslation-r-partial

Lines 93–94 are strangely phrased and present an unusual case, so how should the DM’s translations in 96–98 be annotated?

Here, “move forward 5 feet” cannot be understood without both “ahead north” to indicate forward movement, and “five feet” to indicate distance. So in isolation, Line 98 needs both antecedents (hence the *) but the translated content is only part of the antecedent sequence (i.e., does not include the “north”) so in that regard, the relation is only a partial translation.

Therefore, this situation would be handled like in the example image.

Q22. Handling TUs continued

126	10:33:53.16	do you know the um what you're looking at right now <pause .25>				16				
127	10:33:56.14	like can you tell me what you're looking at				16	126	continue		
128	10:34:02.07		I don't know what all the objects are around me. <beep>			16	127*	answer		
129	10:34:14.93	can you tell me the approximate size of the objects that you're looking at				17				
130	10:34:22.05		Hmm. . .			17	129	processing		
131	10:34:31.20		I'm not sure.			17	129	answer		
132	10:34:52.75		I think you are more familiar with the objects than I am.			17	131	continue		
133	10:35:02.66		I can move to take a good picture of an object that you are interested in. <beep>			17	132	continue	offer-accept	
134	10:35:08.84	yeah that'd be great				17	133	other	offer-accept	
135	10:35:10.02	could you take a picture of the object <pause .27> to your left				17	134	continue		
136	10:35:16.57		processing. . .			17	135*	processing		
137	10:35:30.64			send image of black barrel 2		17	135*	translation-r		
138	10:35:35.66		executing. . .			17	135*	ack-doing		
139	10:35:40.74				done and sent	17	137	ack-done		
140	10:35:42.18		done, sent			17	139	translation-l		

Does Line 135 begin a new TU, or is it the same TU as Line 134?

This is a new TU because Line 134 is an acceptance of the robot's capabilities, and Line 135 is a new attempt to use those capabilities. Another argument for this case is because the original intent of "do you know what you're looking at" (Line 126) is slightly different from "take a picture of the object on your left" (Line 135) one is a question, one is a command to take a picture

The generic/indefinite references to objects in 132,133 point in this direction as description of capabilities rather than a specific intention toward a single object in TU 17, so Line 135 should begin a new TU.

Q23. CMD responds to visual cues, rather than DM text

313	2018-08-06 10:59:02.68	turn ninety degrees				55	312
314	2018-08-06 10:59:05.17		ok			55	313
315	2018-08-06 10:59:11.53			move forward 4ft		55	307
316	2018-08-06 10:59:12.85			then. . .		55	315
317	2018-08-06 10:59:14.13			turn left 90 degrees		55	313
318	2018-08-06 10:59:18.52		executing. . .			55	313*
319	2018-08-06 10:59:23.46	perfect				55	map
320	2018-08-06 10:59:23.91				done	55	317*

[from exp3-P95-house1]

In Line 119, the CMD appears to say "perfect" before the DM can say that the action was finished. So what is the antecedent for Line 119?

There may be cases like this where it appears that the CMD and DM lines are out of order. First, verify with the screen recording. It may be the case where the CMD was responding to the visual information, rather than the DM text.

If so, we annotate the antecedent here as "map".

In this case, the antecedent is "map" and the relation is **3feedback**.

Q24. DM responds to visual cues, rather than RN speech

264	2018-08-02 11:06:53.44	take a picture				36	
265	2018-08-02 11:06:53.80		done			35	263
266	2018-08-02 11:06:54.51	then turn north				36	264
267	2018-08-02 11:06:56.72			send image		36	264
268	2018-08-02 11:06:59.31		sent			36	map
269	2018-08-02 11:07:02.64			then. . .		36	267
270	2018-08-02 11:07:04.27			turn to face North		36	266

[from exp3-P05-house1]

In Line 268, the DM is saying that an action was complete, but there was no message from the RN. So how should Line 268 be annotated?

There may be cases like this where it appears that the RN "sent" speech is missing. First, verify with the screen recording. If the audio is missing, but the RN did take action and the DM did respond, we can infer that the DM saw that the image had sent and pressed the button.

- If so, we annotate the antecedent here as "map".
- In this case, the antecedent is "map" and the relation is **ack-done**.

Q25. DM apologizes out of context

229	16:03:38.92	take a picture				26		
230	16:03:40.81			send image		26	229	translation-r
231	16:03:41.94				sent	26	230	ack-done
232	16:03:43.88		sent			26	231	translation-l
233	16:03:46.65		Woops!			26	232	other
234	16:03:48.32		Sorry :(			26	233	continue

[from exp3-P06-house2]

There may be cases where the DM apologizes to the CMD but it is not clear from the transcript why. First, verify with the screen recording in case something else was happening (e.g., the robot performed the wrong action).

If it is unclear why the DM apologized and no one responds, then we annotate this as an “ERR” TU.

28	2018-08-01 15:22:00.85	okay turn west			
29	2018-08-01 15:22:08.35		ok		
30	2018-08-01 15:22:09.60		Sorry :(		
31	2018-08-01 15:22:10.98	alright			

In other cases where it’s unclear why the DM is apologizing and if the CMD or the RN does respond, we must annotate it as an **other** relation with the antecedent as the previous utterance (from the DM in this case).

List of Symbols, Abbreviations, and Acronyms

2-D	two-dimensional
3feedback	third-turn feedback
ack	acknowledgement
Ant	Antecedent
ARL	Army Research Laboratory
CCDC	US Army Combat Capabilities Development Command
clar	clarification
CMD	Commander
DM	Dialogue Manager
ICT	Institute for Creative Technologies
ID	identification
L	left
nack	negative acknowledgement
nar	non-answer-response
R	right
Rel	Relation
req	request
RN	Robot Navigator
SCOUT	Situated Corpus of Understanding Transactions
SVN	subversion
TU	transaction unit
UMD	University of Maryland
wilco	will comply
WoZ	Wizard-of-Oz
X	experimenter

1 (PDF)	DEFENSE TECHNICAL INFORMATION CTR DTIC OCA
1 (PDF)	CCDC ARL FCDD RLD CL TECH LIB
1 (PDF)	GOVT PRINTG OFC A MALHOTRA
4 (PDF)	CCDC ARL FCDD RLC IT C N BONIAL S M LUKIN M MARGE C R VOSS
3 (PDF)	CCDC DAC FCCD RLH FC K POLLARD A FOOTS A L BAKER