Making Sense of Stop

Claire Bonial[1], Taylor Hudson[2], Anthony L. Baker[2], Stephanie M. Lukin[1], and David Traum[3]
[1]U.S. Army Research Laboratory, Adelphi, MD 20783
[2]Oak Ridge Associated Universities, Oak Ridge, TN 37831
[3]USC Institute for Creative Technologies, Playa Vista, CA 90094
claire.n.bonial.civ@army.mil

Abstract

The instruction to “stop” in human-robot interactions is packed with multiple interpretations. “Stop” can convey the operator’s intent to indicate where the robot should halt motion, or it can convey the operator’s realization that the robot is not executing an instruction satisfactorily and begin the process of repair. We analyze cases of “stop” in a corpus of human-robot dialogue, characterizing them along the dimensions of repair status and timing within the interaction, in order to discover patterns and develop design recommendations for how robots should make sense of “stop.”

1 Introduction

In instructional dialogue in which one participant instructs the other about what to do, including moving to a goal, following a path, constructing an item, or otherwise manipulating an environment, the instructor sometimes tells the other to “stop.” But what does such an instruction mean, and how does it relate to ongoing and planned instructions and executions? In this paper, we attempt to make sense and categorize the usage of “stop” in robot-directed dialogue. The challenges surrounding understanding “stop” arose out of our broader efforts to develop a dialogue system onboard a robot for search and navigation tasks, such disaster-relief efforts, where a human operator uses natural language to instruct a robot on what to look for and where to go in a remote environment (Lukin et al., 2018; Gervits et al., 2021).

We examine a human-robot dialogue corpus (§2) and how instructions of “stop” are used in the interaction. We identify two dimensions of how “stop:” is used: (1) whether it is part of a repair sequence of a problematic aspect of a previously issued instruction (and if so, what kind of repair), and (2) the time during the instructional sequence at which the operator issues the “stop” (§3). We provide examples of each of type, and note that instructions can slide between interpretations and are not mutually exclusive (§4). We then consider how these cases of “stop” might be treated in related work on dialogue systems and human-robot interaction (§5), and conclude with design recommendations to be explored in future work (§6).

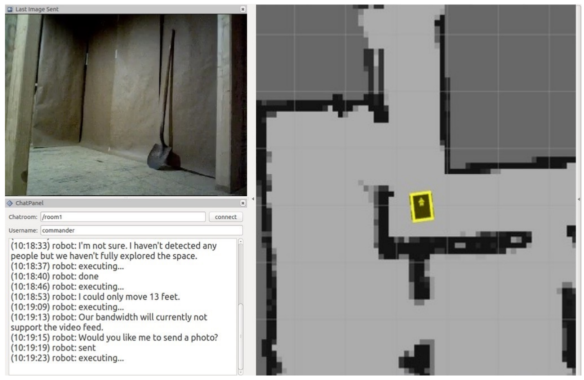

Figure 1: Operator workstation in SCOUT: including image of robot view, chat from robot, and LIDAR map showing terrain features (Marge et al., 2016b).

2 Background: SCOUT Corpus

We leverage the Situated Corpus of Understanding Transactions (SCOUT), a collection of 278 human-robot dialogues acquired through a Wizard-of-Oz paradigm. Here, a human operator instructed what they believed to be an autonomous robot in a remote location through a series of search and navigation tasks, such as finding doorways in an abandoned house, and detecting evidence that a location has been recently occupied (Marge et al., 2016a,b). In the data collection scenario, the human operator speaks to the robot in natural language while sitting at a workstation with three sources of information (shown in Figure 1): a chat stream of text replies from the robot; a 2D terrain (LIDAR) map of the robot’s location that dynamically updates as the robot moves to reveal structural features such as

	Floor 1		*Floor 2*	
#	**Operator**	**DM > Operator**	**DM >RN**	**RN**
69	move forward ten feet			
70			move forward 10 feet	
71		execut- ing...		
72				done
73	stop			
74		done		
75	take a picture			
76			image	
77				image sent
78		sent		

Table 1: Operator instruction to "stop" in 73 is left unexecuted as the robot is already done with the motion instruction and stopped, as evidenced by the Robot Navigator (RN) wizard message "done" to the Dialogue Manager (DM) wizard in 72.

walls and doorways; and images taken at the operator's request from a static, front-facing camera on the robot. Additionally, the operator is shown a picture of the robot (a Clearpath Jackal ground robot that looks like a little truck), but given no other instruction as to what the robot can or cannot do or how to communicate with the robot.

The technical abilities of the robot are provided by two "wizard" experimenters acting out a Wizard-of-Oz experimental paradigm. The Dialogue Manager (DM) wizard stood in for the understanding and dialogue management components by interpreting the operator's instructions, selecting responses, and passing the operator's intent along to another wizard, the Robot Navigator (RN) wizard, who stood in for the planning and motor execution components by joysticking the robot to complete the instruction. The DM interacts directly with the operator while the RN only receives information that is conveyed by the DM. This results in two conversational floors (see Table 1, showing the time aligned messages that the operator saw from the DM Wizard, as well as the messages between the DM and the RN wizard, which the operator does not see or hear).

To support the use of SCOUT to serve as training data for dialogue systems, we have annotated the corpus with a dialogue structure schema (Traum et al., 2018; Bonial et al., 2021). This annotation includes identifying and demarcating Transactional Units (TUs) that include related utterances specifying and fulfilling a speaker's intent, where intent is fairly fine-grained (e.g., a movement to a specified location or a request for the robot to send a picture). In addition to TUs, each utterance within a TU is annotated with its dialogue relation to its antecedent, or the previous utterance it addresses (e.g., "ok" has an *acknowledgement* relation to the preceding utterance). Two questions that arose in the course of this annotation that are explored in this paper are: What kinds of strategies of repair (potentially annotated as a *correction* relation under the schema) exist in the corpus? When is the repair incomplete, such that information must be carried over from the repaired utterance and therefore be part of the same, ongoing TU; and when does the repair express a complete intent that should be a part of a new, distinct TU from the utterance being repaired?

3 Approach to Understanding "Stop"

We analyze all 278 twenty-minute trials of SCOUT for usages of "stop" issued by the operator, and characterize usages according to two primary dimensions: (1) the status of the term with respect to whether or not we observe evidence that it is creating an opportunity for *repair*, or possibly serving as part of the repair, of some problematic aspect of a previously issued instruction (Clark and Brennan, 1991), and (2) the timing of the operator's issuing of "stop." These dimensions are summarized in Table 2 and described in §3.1 and §3.2 to follow. In §3.3, we consider how certain senses of "stop" align with the dimensions of repair and timing.

3.1 Marking Repair

We assume that coordinated activity between two individuals, even human and robot, requires establishing *common ground* in the form of shared mutual knowledge and assumptions (Clark and Marshall, 1981). In conversation, this requires a process of *grounding*, or trying to establish both what has been said and understood between conversational partners (Clark and Wilkes-Gibbs, 1986; Clark and Schaefer, 1989). This includes both positive and negative feedback (Allwood et al., 1992). Traum (1994) presented a computational model of grounding in which different *grounding acts* updated the state of common ground for content under consideration. These acts included initiating a new common ground unit, continuing an existing one by adding more material, acknowledging, repairing, and canceling. We find that ”stop” can play each of these roles.

There have been several efforts to provide a finer-grained taxonomy of repair. Schegloff et al. (1977) characterize repairs as to which participant (self or other) initiates and performs the repair. Levelt (1983) examines a corpus of self-repair to draw distinctions between different types of repairs, and correlates them to the timing of the repair and how much of the original structural material is used in the repair itself. This analysis relies on the assumption that speakers continuously monitor their own production of an utterance, as well as how it is received, for evidence of whether or not the produced utterance achieves the desired effect. In this monitoring process, the speaker may realize that the production is ambiguous in comparison to their intention (requiring an *appropriateness* repair), or that there has been a lexical or syntactic error made during production (requiring an *error* repair). Levelt (1983) finds that appropriateness repairs are much more likely to leverage a *fresh start* strategy, where the repair itself doesn't re-use any structure from the original utterance being repaired. Error repairs, in contrast, tend to retrace and replace the trouble word. The two kinds of repairs are also distinct in their timing: while error repairs tend to be immediate repairs, correcting the mistake in the same utterance, appropriateness repairs are more likely to occur later in the conversation, as the speaker perceives interactional evidence of unsuccessful grounding.

Operating under this theoretical framework, we assume that "stop" may be issued upon the operator's realization that an instruction has not been successfully grounded, thus calling for some kind of repair. Following terminology from Levelt (1983) and Heeman and Allen (1994), adapted in Bohus and Rudnicky (2008), the kind of repair that follows may be a *change* to the original instruction (for example, swapping a word), or it may be a *fresh start*. For our purposes, where we are also annotating this data for TUs to support use in training dialogue systems, we draw a distinction between the two kinds of repair that is related to the TU structure. Repairs that supplement only the information being updated with respect to the original utterance are annotated as *change* repairs. For example, "*Go to the fire hydrant*...**or not fire hydrant, fire extinguisher**." The bold-faced repair changes the destination of "go," but the repair is an incomplete expression of the operator's intent. As a result, the repair must be linked to the italicized *reparandum* within the same TU in order to extract a full expression of the desired behavior: "go to the fire extinguisher." In contrast, repairs that provide a complete expression of the operator's intent are classified as *fresh starts.* Because *fresh start* repairs do not require information from the reparandum to be completely understood, the *fresh start* can be treated as a new intention and new TU, canceling the prior intention.

In summary, repair status annotations of the word "stop" are one of three categories:

1. No evidence of repair in what follows "stop;" the operator wants the robot to halt motion
2. Evidence in what follows "stop" that the operator is repairing the instruction preceding "stop," and that repair is a *change* repair
3. Evidence in what follows "stop" that the operator is repairing the instruction preceding "stop," and that repair is a *fresh start*

3.2 Timing During Instructional Sequence

Past work on repair outlined here indicates that there is an interplay of the interpretation of an utterance as some type of repair and its positioning within the conversational structure. Thus, we also explore the timing of when the operator issues "stop." This can occur during a number of phases within the instructional sequence:

1. As part of the initial instruction, prior to execution
2. During the grounding of the instruction (e.g., if the addressee is clarifying, questioning, or negotiating some aspect)
3. After the instruction has been given and accepted, but before execution
4. During execution, when part has been performed and part remains unperformed
5. After execution, but before grounding the fact that execution has (successfully or unsuccessfully) terminated
6. After it has been agreed that the action has terminated

We annotate each corpus occurrence of "stop" with its timing, where each of the above numbered timings is an annotation category. Table 2 summarizes the timings, maintaining the numbering above.

		Repair Status		
		No evidence of repair - halt motion	Evidence of repair - change strategy	Evidence of repair - fresh start strategy
Timing	1. Original instruction	39	0	0
	2. During grounding	2	5	0
	3. After grounding, before execution	2	0	1
	4. During execution	95	0	18
	5. After execution, before grounding termination	1	0	0
	6. After grounding termination	8	0	0

Table 2: Corpus counts of "stop," as characterized along the dimensions of the timing of issuance and the status as to whether it is creating an opportunity for repair to come, either change or fresh start repair strategies.

3.3 Senses of Stop

The use of "stop" is polysemous in human-robot explorative dialogue. One common sense is a navigation domain action—the opposite of "go," where "go" means to accelerate from zero, and "stop" means to decelerate to zero.[1] Sometimes "stop" means pause rather than terminate, where the expectation is that motion will be resumed after an appropriate interval, which might involve waiting for something else to happen or some change to the future instructions. A stop sign has this meaning: one installed on the street means to wait until the path is clear of other traffic or pedestrians. We roughly group these two senses as relating to halting motion, potentially temporarily.

Another sense of "stop" is a meta-instruction, meaning 'stop doing what you are doing,' which might be equivalent to the first sense, if what you were doing was moving. However, this sense could be applied to any other action, even stopping—one might say *stop stopping* as an instruction to either maintain current speed (above zero) or revert to the previous speed before slowing down with the intention to stop.[2] This sense of "stop" terminates some action.

Although we do not explicitly annotate the sense of "stop" in our corpus analysis, we note that the first sense is often leveraged by operators in contexts unrelated to repair: no misunderstanding or problem has taken place in the collaborative dialogue; the operator simply wants the robot to halt its motion, often to transition to another action such as turning or taking a picture. The second sense can be leveraged by operators to stop an ongoing action that is perceived as problematic or mismatched to their intent; therefore, calling for the robot to stop provides an opportunity for repair to take place in the collaborative interaction, where dialogue is interleaved with different actions. The two senses are not mutually exclusive, and we see overlap in the domain of collaborative exploration in particular, where the action interleaved with dialogue is often motion. Thus, to create the opportunity for repair, the robot must stop its ongoing action, which often is a motion action. When creating the opportunity for repair, "stop" can have some characteristics of an *edit marker*, as it can signal repair to come; yet "stop" also carries its own event semantics, making it much richer than typical edit markers (e.g., "uh").

4 Corpus Analysis of "Stop"

To understand the interplay of timing and the intention behind "stop," and thereby inform our system design, we analyze 171 SCOUT instances of "stop" along the dimensions of repair status and timing. Table 2 provides a matrix of each repair-timing pair, including counts of the number of "stop" instances characterized by that pair. These counts were obtained through double annotation followed by discussion and adjudication. Annotators were two authors trained to do the dialogue structure annotation of the same corpus, therefore also familiar with TU structure. Inter-annotator agreement was high—97% for timing and 85% for repair status, as measured by Gwet's AC-2, a simplified variant of Krippendorff's α (Gwet, 2001). The sections that follow analyze "stop" first according to timing and second by repair status within each timing.

[1]Corresponding to FrameNet's *Halt* frame for words that "denote a Theme ceasing motion" (Baker et al., 1998)

[2]This sense is represented in VerbNet's *Stop* class, in which an Affector ceases to engage in an eventuality (Schuler, 2005)

4.1 Original instruction

No Evidence of Repair, Halt: Operators frequently tell the robot to "stop" in the course of issuing a new instruction. In this timing position, it is generally clear that the operator intends to pinpoint where the robot should halt or pause its motion (39 instances in Table 2). Specifically, some operators include instructions to "stop" at certain landmarks, which apply vacuously in terms of execution, as the robot in this interaction would default to a stop after having achieved the desired end position.[3] For example, *Move forward and stop in front of orange object*. Similarly, some operators include instructions to "stop" in between actions in sequences of complex instructions. For example, *Move forward up to yellow cone, then stop, and turn left ninety degrees*. Interestingly, while some operators do tend to either include or not include explicit instructions to "stop" at the end of a motion or between actions, we did not observe any operators who did so in an entirely consistent fashion for all instructed stopping points, as might be required for a system that did not stop by default.

Evidence of Repair: SCOUT contains clear examples of the *change* strategy of self repairs in the original instruction—clear corrections of a particular instruction word or parameter—but no clear cases of "stop" used in this timing to create an opportunity for repair. What we do see could be characterized as the typically preferred style of *self-initiated, self-repair* (Schegloff et al., 1977). Most involve traditional edit markers, such as "uh," and echo some portion of the original instruction. For example, *go* (pause) *west...no, uh, go east*. Thus, while we do see this kind of repair strategy in the corpus, "stop" is not a felicitous way to mark or create the opportunity for repair of the operator's own production of an utterance.

4.2 During Grounding of Instruction

No Evidence of Repair, Halt: We observed two instances of "stop" in its halt-motion sense as a kind of transition marker during grounding—when responding to a clarification request, the operator repeated the same instruction as the original, but inserted "stop" between actions in the sequence. For example, an operator instructs the robot to *Turn ninety north*, to which the robot responds with a clarification request, *I'm not sure which way to turn towards the north. Should I turn to the left or the right?*, and the operator clarifies: *To the right ninety degrees, stop there, send picture.* We did not tabulate these cases as involving repair, however, because "stop" itself does not create the opportunity for repair, nor does it supply the repair information.

Evidence of Repair: In contrast, we do see five instances of "stop" where it is part of a *change* repair when there is *other-initiated repair* (Schegloff et al., 1977) by the robot. For example, one operator tells the robot to *Move forward,* in response to which the robot asks, *How far forward should I go?*, and the operator responds, *Until I tell you to stop.* Thus, "stop" in this usage aims to provide the requested stopping point information to repair the original instruction.

4.3 After Grounding, Before Execution

No Evidence of Repair, Halt: We classified two instances of "stop" unrelated to repair in this timing. Both involve fairly complex original instructions that were in the process of being translated by the DM to the RN when the operator then issues the "stop" command, and what follows seems unrelated to the original command, and could not address problematic execution as execution had not begun. For example, an operator asks the robot to *Turn forty-five degrees to your left and go back through the doorway.* After the instruction has been acknowledged by the DM, and the DM is in the process of translating this, the operator says, *Stop, take a picture*.

Evidence of Repair: We found one case of a *fresh start* repair following "stop" in the timing position of after grounding and before execution. In this example, the DM acknowledges the operator's instruction, *Face left ten degrees* with an expression of what the robot will do, *I will move forward 10 feet.* It is clear from this acknowledgment that the original command has been misunderstood. Thus, the operator asks the robot to "stop" immediately to prevent the misunderstood command from being executed. The operator then reformulates their original instruction with the *fresh start* repair, *Turn left ten degrees.*

4.4 During execution

The most frequent timing of "stop" is while the robot is executing a previously instructed action (i.e., operators ask the robot to "stop" after the robot has already begun to move). We hypothesize that this timing is ripe for issuance of "stop" either

[3] The default stopping behavior of the Jackal robot may not be clear to the operator, however, and we could imagine other robots for which it would be even less so, such as a Sphero robot that rolls like a ball.

in the purely halt-motion sense or in the sense that terminates the ongoing action in order to create an opportunity for repair, as the operator is usually monitoring the robot's execution of their instruction on the 2D LIDAR map. This map changes and reveals new terrain features as the robot moves into previously unexplored areas (see Figure 1). The robot does not move quickly, so this is often an extended period of time, lasting anywhere from about 5 seconds to up to a minute, depending upon the complexity of the behavior instructed. During this interval, we observe two primary, plausible motivations for issuing "stop." First, the operator may observe a new terrain feature on the map that is of interest to them because it may be a target object of their search, or a doorway or passageway to a new area; in these cases, we assume that issuing "stop" halts the robot's motion so that the operator can pursue a new intention of further investigating a new feature of interest. Second, the operator may observe that the robot's execution does not meet with the expectations of their original instruction; in these cases, we assume that issuing "stop" allows for an opportunity for a repair strategy.

Our insights into the operator's motivation for issuing "stop," and whether it is intended to create an opportunity for repair, are limited to evidence that the robot did not successfully complete a command, and evidence that what follows "stop" bears syntactic and/or semantic relation to the instruction preceding "stop," indicating that it is an attempt to reformulate or repair the original intention. If there is no overt evidence of misunderstanding or repair in the surrounding context, we assume that "stop" issued during execution is intended to stop motion to transition to a new set of instructions and mark this as a *No repair* case.

No Evidence of Repair, Halt: The usage of "stop" to halt the ongoing execution and transition to a new intent is the most frequent kind of usage in our corpus (95 instances in Table 2). In addition to examining the surrounding context for evidence of misunderstanding, the context can provide evidence that the operator has noted an interesting new feature and is pursuing a new intent to better observe it. This can be especially clear when the operator asks for multiple pictures of an area, and they clearly observe an object of interest in one of the pictures. For example, an operator instructs the robot to *Make a 360 degree turn, take a photo every 45 degrees.* During the robot's execution of this complex behavior, the operator issues *Stop*, followed immediately by *Move toward the red bucket*, where the red bucket is an object pictured in one of the images sent during execution.

Evidence of Repair: Although we have noted that "stop" can be issued in this timing to create an opportunity for repair, the repair that follows is never of a *change* strategy, where the repair itself is an incomplete expression of intent and only carries information about what should be corrected or updated within the reparandum. We hypothesize that because execution has already begun, perhaps operators are inclined to repeat the full expression of the intent or rephrase it fully rather than relying on their interlocutor's memory of the preceding command.

Where repair follows "stop" issued during execution, these are always *fresh start* repairs (18 cases in Table 2). This fits with the picture painted by related work, as it may only become obvious that the operator's original instruction requires repair while the operator observes some mismatch between their intent and the robot's execution. The mismatch may arise because the robot seems to have misinterpreted the instruction, or it may arise because the operator realizes that their own production was flawed in some way (e.g., the operator realizes they meant for the robot to turn left, but had said right).

For example, Table 3 illustrates an exchange in which it is clear the operator's high-level goal is to find shovels, and they are exploring different strategies to achieve this. They issue a command to *move forward ten feet* (#70), and then interrupt with "stop," so that the operator can then ask again if the robot sees any shovels. Note there is some ambiguity: "stop" could be seen as halting motion to shift (back) to the intention of asking about shovels, or it could be seen as a marker of the *fresh start* repair to come, motivated by the apparent mismatch between the stopped instruction underway and the operator's desired outcome of execution. In this case, there is evidence in line #75 that the robot was not able to successfully carry out the original instruction of moving ten feet, and instead could only move nine feet. So we have one piece of evidence that the execution may not have matched the operator's intent, calling for repair. Furthermore, there is a lack of confidence in the production of the operator, evidenced by the failed attempt to issue a command that the robot accepts in line #68, as

	Floor 1		Floor 2	
#	**Operator**	**DM > Operator**	**DM >RN**	**RN**
68	keep moving until you see your next shovel			
69		I think you are more familiar with shovels than I am.		
70	move forward <pause .41> ten feet			
71		processing...		
72			move forward 10 feet	
73		I will move forward as far as I can, ok?		
74	stop			
75				done, it was nine
76		done		
77	do you see any shovels near you?			

Table 3: The failed execution of the command underway (line #76) when stop is issued, as well as the speaker's multiple attempts at producing an instruction that reflects their intent, provide evidence that "stop" (line #74) provides the opportunity for a *fresh start* repair in line 77.

well as the long pause of .41 seconds in line #70. It is clear that this operator is struggling to determine how to produce an instruction that will achieve progress towards their goal of finding shovels, so we can also take this as evidence that perhaps the operator is dissatisfied with how the execution of line #70 is matching up with their goal, again calling for repair. Thus, we can conclude that "stop" here creates the opportunity for the *fresh start* repair following it in line #77.

4.5 After Execution

No Evidence of Repair, Halt: We observe one usage of "stop" issued after execution is complete, but before grounding the termination of the action (i.e., after the RN wizard has indicated "done" but before the DM wizard has had a chance to pass this message back to the operator) and eight instances where "stop" is nearly concurrent with, but after grounding termination (i.e., after the DM has indicated "done"). In these cases, there is no evidence in the dialogue that follows that the original instruction was not understood or completed satisfactorily, so we have no evidence that "stop" is signaling the need for repair. We hypothesize that the operator may be trying to remind the robot of where to stop (Walker, 1993), or reinforcing the successful grounding and execution of the action by overtly stating that the robot should stop upon completion. These cases also apply vacuously in terms of execution, because there may not be time in the multi-wizard communication setup for the DM to pass the "stop" command along to the RN before the execution is complete anyway. The dialogue in Table 1 reflects this, showing that the operator issues "stop" after the RN has deemed the execution complete, but before the DM is able to acknowledge or successfully ground termination within the conversational floor with the operator.

5 Considerations from Related Work

How we distill the analysis of "stop" here into particular design recommendations can be informed by related work in dialogue systems and human-robot interaction. Howard et al. (2021) focus on approaches to symbol grounding—mapping natural language to the robot's behaviors and physical surroundings—but the language handled is limited and does not include "stop." It does handle instructions beginning with *Instead* that interrupt another instruction, which cues the robot to pause and transition to execution of the new instruction. We do find hundreds of instances of "stop" in the BladeMistress corpus (Leuski et al., 2012) of virtual human interaction, and our preliminary analysis finds instances of the halt-motion and terminate-action senses, as well as requests to refrain from repeating actions in the future; thus, we are exploring further comparison. In contrast, most other research on dialogue systems focuses on chatbots and smart assistants, therefore, we do not see analysis of interactions involving instructions like "stop," or any physically grounded behaviors. Nonetheless, we can gain insights into how to handle repair.

Much work in NLP broadly has focused on taking disfluent inputs and returning cleaned up grammatical strings, but this practice ignores the fact that repairs often draw upon portions of the disfluent utterance for full interpretation. Hough and Purver (2012) recommend that instead of expunging disfluent utterances as junk, systems should exploit the aborted syntactic categories to supply optional rules for cleaned up parses.

Another major challenge for successful repair strategies is the lack of transparency about the state of a system's understanding (Li et al., 2020). Without some sense of what the system has and has not understood, the operator is left guessing how to repair an utterance that fails to ground successfully, which can be very frustrating (Beneteau et al., 2019; Cho and Rader, 2020). Thus, a body of research has examined patterns and preferences in repair strategies, generally indicating that people prefer a system that can help with repair by somehow pinpointing where and how an utterance has failed and suggesting one that will succeed (Li et al., 2020; Ashktorab et al., 2019; Myers et al., 2018). Complementing this, Bohus and Rudnicky (2008) find that a strategy of simply moving on from the problematic instruction was most preferred in their studies, echoing prior evidence from Wizard-of-Oz studies that show human operators often do not signal non-understanding, instead opting to try to advance the task in some other way. We can see these tendencies reflected in the SCOUT data, where failures to ground are most often addressed with a *fresh start* repair.

6 Conclusions & Future Work

Given the various usages of "stop" that we have analyzed here, the question arises, what should the robot do in each of these cases when given the instruction to "stop"? The robot must decide whether to...

- perform a "stopping" action, terminating current velocity
- halt current execution (and later do something unrelated)
- pause current execution (and resume later)
- pause execution and resume a slightly altered action after a correction has been specified
- ignore the command as redundant with what has already been done (or already planned)
- explain or request clarification when the command seems inappropriate or unclear
- refrain from repeating a previous or current action (that may or may not be planned)

Characterizing "stop" usages according to their timing and repair status reveals patterns that we can use to begin to make some design recommendations. First, within the navigation domain at least, when "stop" is issued as part of the original instruction with a location of stopping (e.g., *Stop at the cone*), it indicates where the robot should halt velocity, but can be ignored as redundant with the planned behavior for execution for certain mobile robots. When "stop" is issued as part of the original instruction between individual segments of a multi-step command (e.g., *Turn left, stop, take a picture*), it indicates where the robot should halt the execution of one step and transition to another and could aid in recognition of individual steps that require sequential execution. Similarly, when "stop" is issued nearly concurrently with successful execution and termination of the original command, it likely indicates feedback from the operator, helping to ground successful completion of the operator's command, as such it may also be ignored with respect to execution.

And finally, when "stop" is issued anywhere after execution of the original command is underway but still incomplete, this should flag the potential need for repair. Although issuing "stop" during execution does not necessarily mean that repair will follow, the robot should recognize the potential for this by deploying a policy of comparing the instruction following "stop" to the original instruction underway and potentially concatenating sources of information from both instructions to gain a fuller picture of the operator's intent and ideally prevent ongoing and future miscommunication.

We will begin to explore implementing these design considerations in our own architecture, but determining the right strategy for handling repair generally remains elusive. To address this, we are currently annotating SCOUT for other types of repair. We must situate our understanding of "stop" with respect to other edit markers and repair strategies (as well as other motion and aspectual verbs) to create a general solution, bringing robots that much closer to efficiently establishing common ground with their human conversational partners.

References

Jens Allwood, Joakim Nivre, and Elisabeth Ahlsen. 1992. On the semantics and pragmatics of linguistic feedback. *Journal of Semantics*, 9.

Zahra Ashktorab, Mohit Jain, Q Vera Liao, and Justin D Weisz. 2019. Resilient chatbots: Repair strategy preferences for conversational breakdowns. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pages 1–12.

Collin F Baker, Charles J Fillmore, and John B Lowe. 1998. The berkeley framenet project. In *COLING 1998 Volume 1: The 17th International Conference on Computational Linguistics*.

Erin Beneteau, Olivia K Richards, Mingrui Zhang, Julie A Kientz, Jason Yip, and Alexis Hiniker. 2019. Communication breakdowns between families and alexa. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pages 1–13.

Dan Bohus and Alexander I Rudnicky. 2008. Sorry, i didn't catch that! In *Recent trends in discourse and dialogue*, pages 123–154. Springer.

Claire Bonial, Mitchell Abrams, Anthony L Baker, Taylor Hudson, Stephanie M Lukin, David Traum, and Clare R Voss. 2021. Context is key: Annotating situated dialogue relations in multi-floor dialogue. In *Proceedings of the 25th Workshop on the Semantics and Pragmatics of Dialogue*.

Janghee Cho and Emilee Rader. 2020. The role of conversational grounding in supporting symbiosis between people and digital assistants. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW1):1–28.

Herbert H Clark and Susan E Brennan. 1991. Grounding in communication.

Herbert H. Clark and Catherine R. Marshall. 1981. Definite reference and mutual knowledge. In Aravind K. Joshi, Bonnie L. Webber, and Ivan A. Sag, editors, *Elements of Discourse Understanding*. Cambridge University Press.

Herbert H. Clark and Edward F. Schaefer. 1989. Contributing to discourse. *Cognitive Science*, 13:259–294.

Herbert H. Clark and Deanna Wilkes-Gibbs. 1986. Referring as a collaborative process. *Cognition*, 22:1–39.

Felix Gervits, Anton Leuski, Claire Bonial, Carla Gordon, and David Traum. 2021. A classification-based approach to automating human-robot dialogue. In *Increasing Naturalness and Flexibility in Spoken Dialogue Interaction*, pages 115–127. Springer.

Kilem Gwet. 2001. Handbook of inter-rater reliability. *Gaithersburg, MD: STATAXIS Publishing Company*, pages 223–246.

Peter Heeman and James Allen. 1994. Detecting and correcting speech repairs. *arXiv preprint cmp-lg/9406006*.

Julian Hough and Matthew Purver. 2012. Processing self-repairs in an incremental type-theoretic dialogue system. In *Proceedings of the 16th SemDial workshop on the semantics and pragmatics of dialogue (SeineDial)*, pages 136–144.

Thomas M Howard, Nicholas Roy, Jonathan Fink, Jacob Arkin, Rohan Paul, Daehyung Park, Subhro Roy, D Barber, Rhyse Bendell, Karl Schmeckpeper, et al. 2021. An intelligence architecture for grounded language communication with field robots. In *Field Robotics, 2021*. Field Robotics.

Anton Leuski, Carsten Eickhoff, James Ganis, and Victor Lavrenko. 2012. The BladeMistress corpus: From talk to action in virtual worlds. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 4060–4067, Istanbul, Turkey. European Language Resources Association (ELRA).

Willem JM Levelt. 1983. Monitoring and self-repair in speech. *Cognition*, 14(1):41–104.

Toby Jia-Jun Li, Jingya Chen, Haijun Xia, Tom M Mitchell, and Brad A Myers. 2020. Multi-modal repairs of conversational breakdowns in task-oriented dialogs. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, pages 1094–1107.

Stephanie Lukin, Felix Gervits, Cory Hayes, Pooja Moolchandani, Anton Leuski, John G Rogers III, Carlos Sanchez Amaro, Matthew Marge, Clare Voss, and David Traum. 2018. Scoutbot: A dialogue system for collaborative navigation. In *Proceedings of ACL 2018, System Demonstrations*, pages 93–98.

Matthew Marge, Claire Bonial, Brendan Byrne, Taylor Cassidy, A. William Evans, Susan G. Hill, and Clare Voss. 2016a. Applying the Wizard-of-Oz technique to multimodal human-robot dialogue. In *RO-MAN 2016: IEEE International Symposium on Robot and Human Interactive Communication*.

Matthew Marge, Claire Bonial, Kimberly A Pollard, Ron Artstein, Brendan Byrne, Susan G Hill, Clare Voss, and David Traum. 2016b. Assessing Agreement in Human-Robot Dialogue Strategies: A Tale of Two Wizards. In *International Conference on Intelligent Virtual Agents*, pages 484–488. Springer.

Chelsea Myers, Anushay Furqan, Jessica Nebolsky, Karina Caro, and Jichen Zhu. 2018. Patterns for how users overcome obstacles in voice user interfaces. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–7.

Emanuel A Schegloff, Gail Jefferson, and Harvey Sacks. 1977. The preference for self-correction in the organization of repair in conversation. *Language*, 53(2):361–382.

Karin Kipper Schuler. 2005. *VerbNet: A broad-coverage, comprehensive verb lexicon*. University of Pennsylvania.

David Traum, Cassidy Henry, Stephanie Lukin, Ron Artstein, Felix Gervits, Kimberly Pollard, Claire Bonial, Su Lei, Clare Voss, Matthew Marge, et al. 2018. Dialogue structure annotation for multi-floor interaction. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.

David R. Traum. 1994. *A Computational Theory of Grounding in Natural Language Conversation*. Ph.D. thesis, Department of Computer Science, University of Rochester. Also available as TR 545, Department of Computer Science, University of Rochester.

Marilyn A Walker. 1993. *Informational redundancy and resource bounds in dialogue*. Ph.D. thesis, Graduate School of Arts and Sciences, University of Pennsylvania.