Making Sense of Stop

Anonymous ACL-IJCNLP submission

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 "stop" usages 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. Although there are a variety of challenges to developing a dialogue system for this kind of interaction and task, here we focus on how we might handle instructions where the operator asks the robot to "stop."

To this end, we leverage a human-robot dialogue corpus collected in an experimental setting similar to our envisioned collaborative task (§2). We identify two theoretical dimensions of "stop:" evidence of an *edit marker* potentially signaling a repair of a problematic aspect of a previously issued instruction, and 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 often ambiguously 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).

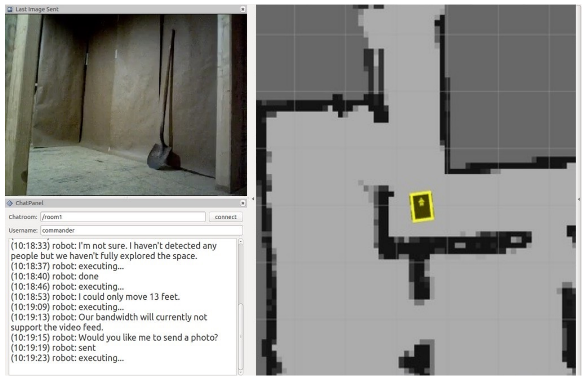

Figure 1: Operator workstation in SCOUT: including image of robot view, chat from robot, and LIDAR map showing terrain features (Marge et al., 2016b).

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).[1] In the data collection scenario, the hu-

[1] This corpus was obtained through a private data-sharing agreement.

man operator speaks to the robot in natural language while sitting at a workstation with several sources of information: 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 walls and doorways, and images taken at the operator's request from a static, front-facing camera on the robot. Figure 1 is a screenshot of the operator's view during the data collection interaction. 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 user's instructions, selecting responses, and passing the user'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 building the robot explores is unfamiliar to the operator and not canonical of any sort; instead it resembles a poorly lit indoor construction site. The operator's assignment in collaborating with the robot to complete search and navigation tasks is therefore challenging on many levels, as the operator must decide on or evolve a strategy for conveying instructions effectively with an extremely unfamiliar conversational partner, and they must decide on or evolve a strategy for how one might go about navigating through the unfamiliar environment to attempt to address their search task. The resulting communications in SCOUT, where communications from the operator and both wizards are transcribed and time-aligned into spreadsheets, are largely instructions for the robot to move either certain distances or (more rarely) to particular landmarks, to turn a certain number of degrees or to face certain structural features, and requests for the robot to send pictures of its current view. Nonetheless, the communications can be quite complex in the interaction of the sources of information at the user's disposal and the dialogue history.

3 Approach to Understanding Stop

We first outline the senses of the word "stop." With these senses in mind, we analyze 208 twenty-minute trials for usages of "stop" issued by the operator in the SCOUT corpus, and characterize usages according to two primary dimensions: i) the status of the term with respect to whether or not we observe evidence that it is serving as an *edit marker*, potentially signaling the need for *repair* of some problematic aspect of a previously issued instruction (Clark and Brennan, 1991), and ii) the timing of the operator's issuing of "stop."

3.1 Senses of Stop

The use of "stop" is polysemous in human-robot explorative dialogue. One common sense of *stop* is a navigation domain action—the opposite of *go*, where *go* means to accelerate from zero, and *stop* means to decelerate to zero.[2] 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. One held by a crossing guard or construction worker, or extending from a school bus means to stop and wait until the sign is put away before resuming. We can 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.[3] In a somewhat related sense, "stop" can also mean to refrain from future performances of a repeated habitual action ("stop kicking my chair," "stop interrupting me"), rather than pausing or interrupting a specific action.[4] We can think of these senses of "stop" as terminating some action, whether it be ongoing or planned/irrealis.

3.2 Evidence of Repair

We assume that coordinated activity between two individuals, even human and robot, requires establishing *common ground* in the form of shared

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

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

[4]This sense corresponds to VerbNet's *Forbid* class, in which a Causer does not allow for a particular eventuality.

mutual knowledge and assumptions (Clark and Brennan, 1991). In conversation, this requires a process of *grounding*, or trying to establish both what has been said and understood between conversational partners. Clark and Brennan (1991) divide the grounding process into a presentation phase, in which one speaker presents an utterance for the addressee to consider, and the acceptance phase, in which the addressee accepts the utterance by demonstrating understanding. Repair, the attempt to correct an utterance that is not satisfactorily grounded, is one method of presentation in this paradigm, in which conversational partners are searching for both negative evidence (signals of misunderstanding) and positive evidence (potentially successful execution of an instruction) that common ground has been achieved.

Schegloff et al. (1977) consider the distinction of other vs. self initiation as part of an explanation of the clear preference for self-repair observed in conversation. The authors posit the importance of repair-initiation opportunity positions in conversational structure, which are leveraged distinctly for either other or self initiation of 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 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*. We assume that some structure of the original instruction must be re-used for a repair strategy following "stop" to qualify as a *change*, whereas repairs that do not echo any of the original instruction are assumed to be *fresh starts*.

3.3 Timing During Instructional Sequence

The above discussion has begun to highlight the 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. We list these different timings below:

1. As part of the initial instruction, prior to beginning execution
2. During the grounding of the instruction; for example, when the operator has specified part but not all of the instruction sequence, or if the addressee is clarifying, questioning, or negotiating some aspects, such as a termination point or manner
3. After the instruction has been given and accepted, but before execution has begun
4. During execution, when part has been performed and part remains unperformed
5. After execution, but before the conversational partners ground the fact that execution has (successfully or unsuccessfully) terminated
6. After it has been agreed that the action has terminated (seems unrelated to the instruction, perhaps like the first case, relating to a new instruction)

4 Corpus Analysis of "Stop"

To understand the interplay of timing and the intention behind "stop," and thereby inform our system design, we analyze SCOUT for instances of "stop" along the dimensions of repair status and timing. We first discover that the instruction sequences in SCOUT are not as clear as the six timings we hypothesized in §3.3 due to the complex multi-wizard

	Floor 1		*Floor 2*	
#	**User**	**DM > User**	**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.

paradigm of the experimental data—where the Dialogue Manager (DM) wizard interfaces directly with the operator, and simultaneously passes the operator’s intent to the second wizard, the Robot Navigator (RN) moving the robot with a joystick. 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). These floors add additional sub-steps in the instructional sequence. We therefore insert the following new timing phases that can make up sub-steps of the original six timings, given that SCOUT involves passing information between two conversational floors:

3.1 After the instruction has been given by commander but before it has been translated by DM to the RN

3.2 After the instruction has been translated by DM, but before it has been carried out by RN

5.1 After the RN has finished executing, but before the RN has acknowledged completion

5.2 After the RN has acknowledged completion but before the DM has translated the completion.

The remainder of this section categorizes 132 cases of “stop” issued by the operator in SCOUT by repair status, and describes the commonly observed timings within each type. Table 2 provides a matrix of each repair-timing pair, including rough counts of the number of “stop” usages characterized by that pair. The counts are rough given the challenge of weighing evidence of whether or not a usage is acting as an edit marker signaling repair to come; nonetheless, the tendencies can be observed.

4.1 No Evidence of Repair - Halt Motion

A frequent timing of issuing “stop” is **during the original instruction**. In these cases, it is clear that the operator intends to pinpoint where the robot should halt or pause its motion (38 instances in Table 2). 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.[5] For example, *Move forward and stop in front of orange object*, and *Move forward to end of hallway, stop at wall*. Somewhat 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*, and *Turn right thirty degrees, move forward ten feet, stop, send picture*. 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.

We did observe two usages of “stop” as the same kind of transition marker **during grounding of an instruction**—when responding to a clarification request, the operator repeated the same instruction as the original, but inserted “stop” between actions in the sequence. We did not tabulate these cases as involving repair, however, because “stop” itself neither acts as an edit marker nor supplies the repair information. 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.*

[5]The default stopping behavior of the Jackal robot may not be clear to the user, 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.

		Repair Status		
		No evidence of repair - halt motion	Evidence of repair - change strategy	Evidence of repair - fresh start strategy
Timing	Original instruction	38	0	0
	During grounding	2	7	0
	After grounding, before execution	0	0	0
	During execution	59	7	5
	After execution, before grounding termination	2	0	0
	After grounding termination	12	0	0

Table 2: Corpus counts of “stop”, as characterized along the dimensions of the timing of issuance and the status as a potential edit marker, signaling repair to come, either change or fresh start repair strategies.

The most frequent timing usage of “stop” is **during execution** (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 in the purely halt-motion sense or as an edit marker, signaling repair to come, as we note that 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 during this period that the robot’s execution does not meet with the expectations of their original instruction; in these cases, we assume that issuing “stop” acts as an edit marker, allowing for an opportunity for a repair strategy. In absence of having insights into the operator’s true intent, we can only attempt to glean evidence from the surrounding context, noting if there is evidence that the robot did not successfully complete a command, or if what follows “stop” bears some relation to the original instruction indicating that it is repair. 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 the *halt* sense.

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 (59 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, a user 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 user 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.

We also observe “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 nearly concurrent with, but **after grounding termination**. 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, but rather the operator seems to be reinforcing the successful grounding and execution of the action. These cases also apply somewhat vacuously in terms of execution, because there may not

	Floor 1		*Floor 2*	
#	**User**	**DM > User**	**DM >RN**	**RN**
82	um go <pause. 33>go straight			
83		How far should I move forward?		
84	five feet			
85		ok		
86			move forward 5 feet	
87	okay stop			
88			stop	
89				done, I could move about 4
90		done		
91	go east, go east five feet			

Table 3: Evidence from unsuccessful execution of the original instruction (line 89) and re-use of structure of the original instruction indicates "stop" precedes a *change* repair strategy.

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.

4.2 "Stop" & Evidence of Repair

If there is some evidence of misunderstanding and/or an attempt to repair, we assume that "stop" is an edit marker and try to distinguish whether what follows is a *change* repair strategy or a *fresh start* repair strategy.

4.3 Evidence of Repair - Change Strategy

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. The clear cases that we do see could be characterized as the typically preferred style of self-initiated, self-repair. Most involve 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 an appropriate edit marker for repair of the operator's own production of an utterance.

During grounding of the instruction, we do see "stop" issued not as an edit marker, but as part of the repair itself when there is other-initiated repair by the robot (7 usages in Table 2). For example, one user tells the robot to *move forward,* in response to which the robot asks, *how far forward should I go?*, and the user 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.

During execution is the only timing where we observe "stop" issued as an edit marker, signaling a *change* repair strategy to come. For example, in Table 3, the user tentatively issues an instruction to *um go (pause) go straight...five feet*. This cannot be successfully completed in the space restrictions, as evidenced by the RN's declaration *I could move about four.* Further evidencing the interpretation of repair is the echoing of the structure of the original instruction in the repair, with one parameter changed: *go east, go east five feet.*

4.4 Evidence of Repair - Fresh Start Strategy

We only observe "stop" issued as an edit marker, signaling a *fresh start* repair to come **during execution** (5 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. This 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). If there is no repetition or rephrasing of the instruction underway included in the repair itself, then these cases are deemed *fresh start* repairs.

For example, Table 4 illustrates an exchange in which it is clear the operator's high-level goal is to

	Floor 1		*Floor 2*	
#	**User**	**DM > User**	**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 4: 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) is an edit marker, providing the opportunity for a *fresh start* repair in line 77.

find shovels, and they are exploring different strategies to achieve this. They issue a command to *move forward ten feet*, 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 that the robot was not able to successfully carry out the original instruction of moving ten feet in line #75, which indicates the robot 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 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 is an edit marker, with a *fresh start* repair following it.

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. In contrast, most research on dialogue systems focuses on chatbots and smart assistants like Alexa and Siri, 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. The authors highlight the requirement for a strong incremental approach to interpretation, as successful interpretation of repair requires that the system can make available the maximal amount of information possible from the unfinished utterance as it is being processed.

While incremental interpretation may help with a system's limited capabilities to understand the operator's verbal attempts at repair, it does not address another major challenge for successful repair strategies: a 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.

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, to terminate current velocity
- halt current execution of an action (and later do something unrelated)
- pause current execution of an action (and resume the action 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 to do)
- explain or request clarification when the command seems inappropriate or unclear
- refrain from repeating a previous or current action (that might or might not currently be planned to do again)

When we characterize "stop" usages according to their timing and repair status and reveal patterns, we can 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 likely be ignored as redundant with the planned behavior for 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.

On the other hand, 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.

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 is not necessarily an edit marker, 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.

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

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.

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.

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.

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.

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.