Artificial Intelligence for Human-Robot Interaction
AAAI Technical Report FS-17-01

Theoretical Concerns for the Integration of Repair

Sean Trott, Federico Rossano
University of California, San Diego, Cognitive Science
sttrott@ucsd.edu, frossano@ucsd.edu

Abstract

Human conversation is messy. Speakers frequently repair their speech, and listeners must therefore integrate information across ill-formed, often fragmentary inputs. Previous dialogue systems for human-robot interaction (HRI) have addressed certain problems in dialogue repair, but there are many problems that remain. In this paper, we discuss these problems from the perspective of Conversation Analysis, and argue that a more holistic account of dialogue repair will actually aid in the design and implementation of machine dialogue systems.

Introduction

One of the goals of Artificial Intelligence and Human-Robot Interaction is facilitating human communication with machines through language. The field has come a long way since early efforts (Winograd, 1971; Weizenbaum, 1966). Robots can now understand basic commands, ask for clarification, and correctly interpret certain indirect requests (Williams et al, 2015). However, there is considerable work to be done to enable naturalistic conversations with robots.

Notably, human dialogue – particularly spoken dialogue – is difficult to understand. Listeners must contend with input that is ill-formed and fragmentary (Garrod and Pickering, 2004). Utterances in real-time conversations are peppered with disfluencies, interruptions, and repairs.

Given the noisiness of this input, it seems remarkable that humans can engage in conversation at all – and perhaps impossible to build robots that can do the same. And yet, humans *do* converse with each other. Our conversations occur at remarkable speeds, with gaps between dialogue turns lasting an average of only 0.2 seconds (Levinson, 2016), with some cross-cultural variability (Stivers et al, 2009); furthermore, repairs happen fluidly in the stream of dialogue, without listeners constantly halting the conversation to check for clarification.

In this paper, we focus on how the larger system governing the structure and organization of conversation facilitates its flexible and efficient use by humans (Schegloff, Jefferson, and Sacks, 1977), and what insights this system can offer the HRI community. One of the core insights is that language use is *orderly* (Sacks, 1992), and this order usefully constrains the problem space.

Specifically, we will focus on the occasion of *repair* in conversation. Repair is a "self-righting mechanism for the organization of language use in social interaction" (Schegloff, Jefferson, and Sacks, 1977), employed by both speakers and listeners to address problems in speaking, hearing, or understanding. Here, we will focus on the structure of *self-initiated repair* in particular, but *other-initiated repair* will be partially addressed in the Discussion section.

We will outline the main problems posed by self-repair, including those which have not yet been addressed in the HRI literature. Ultimately, we will argue that an understanding of repair as a *conversational resource* (Hayashi et al, 2013), as opposed to a "hitch" in the system, will help design machines that can interact more successfully with humans. We argue this point using insights from Conversation Analysis. Conversation Analysis is a micro-analytical approach that aims to describe the organizing principles of how people talk during social interactions. Its method is observational, mostly qualitative, and inductive, and relies on audio and video recordings of naturally occurring interactions both in ordinary and institutional settings.

Previous Work

Previous HRI work has made considerable progress in addressing certain problems in understanding dialogue repair and disfluencies. This section will discuss the problems that current systems have addressed.

Problem 1: Integrating disfluencies in speech

A *disfluency* is a break in the flow of human speech. Common examples in English include non-lexical fillers such as "um" and "uh", or lexical fillers such as "like".

Most existing solutions simply identify and remove these disfluencies, and do not attempt to interpret their import for upcoming utterance planning or sequence organization of turns at talk, with the exception of Gervits (2017). This makes the assumption that disfluencies function solely as "fillers", and do not add other information to the utterance.

Non-lexical disfluencies

Non-lexical disfluencies pose a challenge to language understanding systems because they can occur at various points of an utterance, and their meaning or function as it relates to the larger utterance at hand is not immediately clear.

Both Gervits (2017) and Cantrell et al (2010) use a first-pass regular expression filter to remove fillers like "um" from an utterance; the resulting "cleaned" utterance can then be fed into a language parser. This is not unlike the solution discussed in Bastianelli et al (2014), in which "wild cards" (such as "erm") are added to a language model's grammar, to allow their insertion in various parts of an utterance.

Assuming: 1) a speech recognizer correctly converts audio input to textual representations of these non-lexical fillers; and 2) these fillers do not convey useful information about an utterance; this seems to be a successful solution to the problem.

Lexical disfluencies

Disfluencies can also take the form of lexical fillers. The HRI literature typically distinguishes lexical fillers from non-lexical fillers in that lexical fillers have another function besides serving as a filler word. For example, the word "like" can be used either as a filler, as in (a) below, or as a comparative preposition, as in (b):

(a) This is, like, my favorite coffee shop.
(b) This is like my favorite coffee shop.

A system should be able to interpret the "like" in (a) as a lexical filler bearing little to no significance on utterance meaning, and the "like" in (b) as comparing the indexical ("this") to the speaker's favorite coffee shop.

Cantrell et al (2010) use a trigram statistical model to differentiate between competing analyses (e.g. "like" as filler vs. comparative preposition). If "like" is ungrammatical or statistically unlikely, it can be interpreted as a filler. Unfortunately, disambiguating between the usages in (a) and (b) would be difficult with this solution; unless the model makes use of timing or prosody information, (a) and (b) would likely have the same score.

Kruijff et al (2010) describe a sophisticated model for handling disfluencies (both lexical and non-lexical) through contextual information. Here, "context" is represented by a situation model including both visual features of the environment and previous utterances in the discourse, forming a "cross-modal *salience model*" (Kruijff et al, 2010). This model can be used to compute probability distributions over which words are likely to be heard next, and thus identify a word's most likely function.

Again, assuming: 1) text-to-speech is correct; and 2) the goal is simply to identify and remove lexical fillers; these disambiguating solutions are successful approaches to the problem.

Problem 2: Integrating same-turn self-repairs

In addition to disfluencies like filler words, utterances also contain attempts at *repair*, such as:

(c) Then take a right turn – **I mean** a *left turn*…

This poses several challenges. First, a system must recognize that a repair has occurred at all. Second, the system must recognize which information is being corrected ("a right turn"), and ignore that in its eventual parse. Finally, a system must recognize which information is meant to *replace* the incorrect information, and make a substitution such as to produce a grammatical sentence.

Cantrell et al (2010) describe an implemented system that addresses all three of these challenges. The system recognizes the occurrence of a repair using a grammar of *repair markers* (such as "I mean"). Then, the system checks whether the word or phrase immediately after this marker can be substituted into the syntactic slot preceding the marker. If both phrases are of the same type (E.g. both are noun phrases), a replacement can be performed.

Assuming: 1) a self-repair contains an explicit marker; 2) the repair is initiated on the same turn as the repairable item; and 3) the repair operation is *replacement*; this solution will be successful. As we shall see below, not all of these assumptions hold.

Outstanding Problems

The implementations described above solve a certain class of problems in dialogue repair and speech disfluencies, but make several limiting assumptions. Below, we discuss the problems that remain and where possible, propose design solutions. We propose that the added complexity can actually *help* dialogue systems – not just with the integration of self-repair, but with other problems in the domain of language and interaction.

In this section, we have attempted to avoid field-specific jargon. However, several important terms should be clearly defined here. First, a *turn-at-talk* describes the time during which one speaker holds the floor. Second, a *TCU*, or *turn-constructional unit*, is a self-contained unit of speech during a turn-at-talk; depending on context, this could be a word, a clause, or sentence (Sacks, Schegloff, and Jefferson, 1974). Finally, a *transition-relevance place* is the point at which another speaker could begin their turn, or the current speaker could continue with their own (Sacks, Schegloff, and Jefferson, 1974).

Problem 3: Repairs do more than correct errors

Self-initiated repairs frequently occur in the absence of a noticeable or obvious error (Kitzinger, 2013; Schegloff, 2007). Often it is only from the repair itself that the listener learns there was a problem with what was previously said.

This means that a dialogue system cannot rely solely on using explicit errors (whether phonetic, lexical, or syntactic) to identify the repairable item. More importantly, repairs can serve other purposes than the correction of speech. For example, repair can "fine-tune" the intended interpretation of a dialogue turn, or even pre-emptively address a projected misinterpretation down the line (Kitzinger, 2013):

(1) *H*: This girl's fixed up on a **da- a blind date**.

In (1), no noticeable error has been produced by *H*. However, *H* predicts that the word "date" will be insufficient and lead to a misinterpretation on the part of the listener. Thus, *H* cuts off her expression at the midpoint and inserts a more specific expression: "*blind* date".

The Bright Side

This seems to complicate matters, but the silver lining is that these repairs can now be used to infer the intentions of the speaker. Assuming the repair can be identified and integrated using an enriched model of repair operations and technology (see: *Problem 4* and *Problem 5*), the listener (e.g. the robot) could theoretically learn additional information about what the speaker is trying to say.

For example, in (1), the listener can infer that *H* believes "date" to be an insufficient descriptor for their intended interpretation; it is somehow significant that the date is a *blind* date. This aids with interpreting both the utterance at hand, as well as future utterances in the discourse through the lens of this repair.

Problem 4: Operations other than replacing

Speakers employ at least ten different *repair operations* for addressing problems in their turn at talk (Schegloff, 2013). An *operation* refers to the mechanism by which a speaker alters their turn in some "interactionally consequential way" (Schegloff, 2013). The most common operation is replacing (Kitzinger, 2013), in which a speaker substitutes "for a wholly or partially articulated element of a TCU-in-progress another, different element" (Schegloff, 2013). For example:

(2) Pick up the **green** box – the **blue** box.

This is the operation assumed by most systems for integrating repair, such as Cantrell et al (2010). But speakers also make use of a number of other repair operations: inserting, deleting, searching, parenthesizing, aborting, sequence-jumping, recycling, reformatting, and reordering (Schegloff, 2013).

Thus, the mechanism for integrating a repair must be contingent on the operation used, meaning: 1) a system will have to identify which of the ten operations is being employed; and 2) the system could require as many as ten different integration procedures, one for each operation.

The Bright Side

The dizzying array of possible repair operations makes the problem appear insurmountable – but fortunately, as with many aspects of language, there is a systematicity to the situations in which these operations are employed.

The different repair operations can often be associated with different repair *technology* – the practice by which the repair is performed (see *Problem 5* below). Although *inserting* and *parenthesizing* can be used to address future trouble in the dialogue, they are deployed with different technology. *Parenthesizing* repairs, as in (3) below, are often composed of a clausal TCU, and can be "interpolated into a turn-constructional unit and contained there" (Schegloff, 2013):

(3) *M*: So, boy when Keegan come in he – **y'know how he's gotta temper anyway** – he just…

Systems have already been implemented to deal with the *replacing* operation. While writing procedures for the other operations will certainly be challenging, it does not seem as insurmountable when one considers each separately. For example, *inserting*, as in (1) above, consists of adding "one or more elements into the turn-so-far, recognizable as other than what was on tap to be said next" (Schegloff, 2013); thus, rather than searching for a word or phrase to *replace*, a system should recognize which word or phrase to *insert*, and where. Other operations, such as *parenthesizing*, might require back-channeling information in the form of verbal feedback (e.g. "uh-huh") or gesture (e.g. a nod).

And as mentioned previously, these repairs provide additional information about the speaker's intentions. Both *inserting* and *parenthesizing* are used to address projected trouble down the line, albeit in different ways: *inserting* suggests that the speaker felt the original expression to be insufficient, while *parenthesizing* can be used to check for understanding (e.g. by requesting back-channeling) or emphasize a piece of information in a story, as in (3) above.

Problem 5: Repairs without explicit markers

Sometimes repairs contain explicit markers, such as "I mean", delineating the repairable item and the repair. Such markers are useful to dialogue systems, because they provide a way to identify that a repair has occurred, and organize the sentence to facilitate the integration of repair.

As Cantrell et al (2010) note, however, utterances without such markers will lead to "failed semantic parses", since the system relies on the marker in its parsing procedure. Unfortunately, many repairs do *not* contain markers such as "I mean" (Schegloff, 2013; Kitzinger, 2013). Thus, a dialogue system must have other affordances for identifying and integrating repairs.

The Bright Side

The good news is twofold:

First of all, there is a limited set of practices. The space of possibilities is by no means infinite, and is described in great detail in the conversation analytic literature (Schegloff, 2013; Wilkinson and Weatherall, 2011; Kitzinger, 2013). For example, Schegloff (2013) writes that *parenthesizing* is generally deployed in the form of a clausal TCU within a larger TCU. This means that the first step – **recognizing** that a repair has occurred – can be achieved by expanding the search space of repair indicators.

Second, as noted in the section on repair operations, the different technologies are deployed for different operations. While this correspondence is by no means one-to-one, it does provide a heuristic by which to order potential integration procedures. This means that the second step – **integrating** the repair – can be achieved, at least in part, by mapping different operations to their most frequent repair technologies.

Problem 6: Repair initiated later than same-TCU

Most self-initiated repairs are initiated on the same TCU as the repairable item, as in (1). But a speaker can initiate a self-repair in at least three points in a conversation (Schegloff, Jefferson, and Sacks, 1977):

1. The same turn as the *trouble source* (the item to be repaired).
2. The *transition space*, or the space just before the listener's next turn.
3. Third turn, or the turn after the listener's next turn.

Self-repairs on the same turn as the trouble source require the listener only to integrate information across a single TCU. Third turn self-repairs, on the other hand, require the listener to integrate the repair across multiple turns, as in (4) below (adapted from Schegloff, Jefferson, and Sacks (1977)):

(4) *L*: I read a very interesting story **today**.
M: What's that?
L: Well **not today, maybe yesterday**, it's called Dragon Stew.

To understand *L*'s repair, *M* must keep in mind the original utterance across two turns, and determine which word or phrase the repair refers to.

The Bright Side

Fortunately, there is also systematicity in the timing of repair initiation. Like same-TCU self-repairs (and unlike other-initiated repairs), we know that the trajectory of initiation to completion for transition space and third turn are "overwhelmingly successful within the turn in which they are initiated" (Schegloff, Jefferson, and Sacks, 1977).

More importantly, we also know that operations with which transition-space and third-turn repairs are performed, and the technologies with which they are deployed, are very similar to those used for same-TCU self-repair (Kitzinger, 2013). This simplifies our task considerably, because the solutions for transition-space and third-turn repairs can thus mirror the solutions for same-TCU repair. Crucially, the only difference is that a listener (or dialogue system) must widen the space of dialogue states to search. When a dialogue system recognizes that a repair has been initiated (using the insights from *Problem 4* and *Problem 5*), it must not only search the current utterance for the trouble source, but the utterance two turns ago. This entails that a dialogue system should always be tracking at least three dialogue turns at any given time.

Discussion

In this paper, we set out to contribute two things:

1. A description of the problems in self-repair that current dialogue systems do not address.
2. A deeper understanding of how one might solve these problems.

We have presented each problem separately, but it should be apparent that they are all related – they are all part of the same *system* of repair.

It is also hopefully clearer now that there is "order at all points" in language use, even the structure of conversations (Sacks, 1992). Although it may seem that we are merely adding to the pile of issues that dialogue systems must solve, viewing these problems as part of a larger system should help in constructing a more generalizable solution. Notably, a better understanding of repair could help with other language understanding tasks as well, such as intention recognition and discourse modeling.

Of course, actually implementing solutions to these problems will still be very challenging. We hope, however, that insights from those studying the detailed structures of human conversations will aid in building systems to better integrate self-repair.

By focusing on self-repair, we have admittedly neglected the issues that arise in *other-repair*. Most language understanding research considers other-repair only from the perspective of clarification requests, but other-initiated and self-initiated repair are, again, both part of the same system for organizing repair in conversation. There is a wealth of research on other-initiated repair (Schegloff, Jefferson, and Sacks, 1977; Schegloff, 1997; Schegloff, 2000), including the way it is deployed across cultures and languages (Dingemanse and Enfield, 2015; Dingemanse et al, 2015). The latter will be particularly helpful as HRI research attempts to expand its scope beyond monolingual, English-speaking dialogue systems.

References

Bastianelli, E., Castellucci, G., Croce, D., Basili, R., & Nardi, D. 2014. Effective and robust natural language understanding for human-robot interaction. In *Proceedings of the Twenty-first European Conference on Artificial Intelligence* (pp. 57-62). IOS Press.

Cantrell, R., Scheutz, M., Schermerhorn, P., & Wu, X. 2010. Robust Spoken Instruction Understanding for HRI. In *Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction* (Vol. 1).

Dingemanse, M., & Enfield, N. J. 2015. Other-initiated repair across languages : towards a typology of conversational structures. *Open Linguistics*, 96–118. http://doi.org/10.2478/opli-2014-0007

Dingemanse, M., Roberts, S. G., Baranova, J., Blythe, J., Drew, P., Floyd, S., Gisladottir, R., Kendrick, K., Levinson, S., Manrique, E., Rossi, G., Enfield, N. J. 2015. Universal Principles in the Repair of Communication Problems. *Plos One*, 1–15.

Garrod, S., & Pickering, M. J. 2004. Why is conversation so easy?. *Trends in cognitive sciences*, *8*(1), 8-11.

Gervits, F. 2017. Disfluency Handling for Robot Teammates. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 341-342). ACM.

Hayashi, M., Raymond, G., & Sidnell, J. 2013. Conversational repair and human understanding : an introduction. *Conversational Repair and Human Understanding*.

Kitzinger, C. 2013. Repair. In *The Handbook of Conversation Analysis*.

Kruijff, G. J. M., Lison, P., Benjamin, T., Jacobsson, H., Zender, H., Kruijff-Korbayová, I., & Hawes, N. 2010. Situated dialogue processing for human-robot interaction. *Cognitive Systems*, 311-364.

Levinson, S. C. 2016. Turn-taking in human communication–origins and implications for language processing. *Trends in cognitive sciences*, *20*(1), 6-14.

Sacks, H., Schegloff, E. A., & Jefferson, G. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language*, 696-735.

Sacks, H. 1992. *Lectures on conversation* (Vol. I). Cambridge MA: Blackwell.

Schegloff, E. 2013. Ten operations in self-initiated, same-turn repair. In *Conversational Repair and Human Understanding*.

Schegloff, E. A. 1997. Practices and actions: Boundary cases of other-initiated repair. *Discourse Processes*, *6950*.

Schegloff, Emanuel; Jefferson, Gail; Sacks, H. 1977. The Preference for Self-Correction in the Organization of Repair in Conversation. *Linguistic Society of America*, *53*(2), 361–382.

Schegloff, E. A. (2000). When 'others' initiate repair. *Applied linguistics*, *21*(2), 205-243.

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., ... & Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, *106*(26), 10587-10592.

Williams, T., Briggs, G., Oosterveld, B., & Scheutz, M. 2015. Going Beyond Literal Command-Based Instructions: Extending Robotic Natural Language Interaction Capabilities. In *AAAI* (pp. 1387-1393).

Winograd, T. 1971. *Procedures as a representation for data in a computer program for understanding natural language* (No. MAC-TR-84). MASSACHUSETTS INST OF TECH CAMBRIDGE PROJECT MAC.

Weizenbaum, J. 1966. ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, *9*(1), 36-45.