

Analizing disfluences in HCI

Guillem Gili i Bueno
University of Potsdam
giliibueno@uni-potsdam.de

Abstract

In this document we analyze the disfluences that happen when a human interacts with a robot. Specifically we will focus on self-repair disfluences, that happen in utterances that happen within the same turn of a speaker (in this case the human). We also compare the differences in this behavior when interacting with a robot rather than a human. We try and make sense of the differences from rigorous studies and our results hint a bit at the fact that self-repair may also be affected by alignment.

1 Introduction

The phenomena of *Disfluences* refers to morphosyntactical irregularities that happen in spoken dialogue including fillers("um","eh") and corrections to previous words (Figure 1).

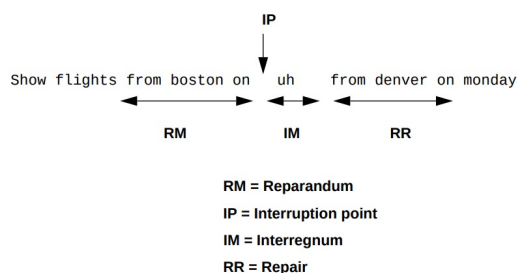


Figure 1: *Disfluency* example extracted from (Shriberg, 1994), one of the most comprehensive works on *Self-repair* in English

When we do *Disfluences*, one of the methods to mitigate confusions is to use *Repair*. In this project I will be always talking about *Self-repair* as the Self-Initiated *Self-repair*, but it is worth mentioning that it can be Other-Initiated. *Self-repair* refers to the act of stopping within the utterance to correct what the same speaker was in the process of saying. In particular I will focus on the replacement and recycling repair types.

Another phenomena we must cover is *Alignment*. *Alignment* is defined in (Pickering and Garrod, 2004):

[...] interlocutors tend to develop the same set of referring expressions to refer to particular objects, and that the expressions become shorter and more similar on repetition with the same interlocutor and are modified if the interlocutor changes [...]

The support for this definition comes from research in priming (evoking an inadvertent change in dialogue) syntactically (Branigan et al., 2000) and lexically (Bock, 1986). And in the same paper where we borrowed the definition for *Alignment*, the following extension is made regarding dialogue specifically:

[...] This occurs by priming mechanisms at each level of linguistic representation, by percolation between the levels so that alignment at one level enhances alignment at other levels, and by repair mechanisms when alignment goes awry. [...]

2 Related work

2.1 Motivations

Self-repair is well-established and classified in many languages, e.g. English (Shriberg, 1994), Japanese (Hayashi, 1994), Indonesian (Wouk, 2005). The studies across these show evidence that *Self-repair* is both inconsistent across different languages and highly conventionalized (e.g. Figure 2,3,4). Further evidence for these properties can be found easily in other papers (Fox et al., 2010).

It would seem likely that the two phenomena, *Alignment* and *Self-Repair* (not present in robotic speech), could alter the human dialogue both in the rate and type of *Self-repair*. On one hand it

.hh maa sonna:: [are ga:::: (1.5) u: **meedonsan ga**] iru
 well like DEM SUBJ uhm maid SUBJ exist
 yoonaa: ie ya nai kara:,
 such family be not because
 because, like, we are not the sort of family to have that,
 uhm a maid

Figure 2: Placeholder repair example extracted from (Wouk, 2005). This type of *Self-repair* does not exist in English

in this building- we finally got [a- **a room**] today
 [I cannot, **I cannot**] stand that.

Figure 3: *Self-repair* examples extracted from (Wouk, 2005). *Self-repair* after a Direct Object recycles a whole constituent (rather than only the verb). This does not happen in Japanese.

is likely that a human tries to *Self-repair* less, due to trying to align to said robotic speech. It is also entirely possible that the effect on *Self-repair* is the opposite, since our second quote from (Branigan et al., 2000) points that *Self-repair* is precisely used when aligning mechanisms go awry. Regarding the type of *Self-repair*, for example, some substitutions may replace with pronouns less due to a perceived minimal common ground with the robot.

Furthermore, *Alignment* is not limited to the aspects listed in the original paper (Pickering and Garrod, 2004), *Alignment* happens in several modalities, quoting from (Branigan et al., 2010):

[...] interlocutors tend to align on accent and speech rate [...] adopt increasingly similar phonetic realizations of repeated words [...] repeat each other in terms of their grammatical choices [...]

In the same paper (Branigan et al., 2010), several other motivations for *Alignment* (passive) adaptations are mentioned that may be relevant in HCI:

[...] cognitive economy that they offer, since they do not require the speaker to model her interlocutor [...] might also serve a different strategic purpose, related to the social relationships between the participants in the interaction [...]

It is possible that one may not find the motivations listed when interacting with a robot. On another hand, the paper does also compile a lot of evidence on alignment happening when a person interacting in a Wizard of Oz paradigm believed

tteyuuka koko denwa [kaket- **kakete**] kite sa,
 I:mean here telephone ca- call come FP
 'I mean, (they) ca- called us here,'

Figure 4: *Self-repair* example extracted from (Wouk, 2005). *Self-repair* in Japanese of the direct constituent (verb), rarer in English.

the other agent to be a robot (less pronoun usage, happened in the reverse Wizard of Oz paradigm as well). Other syntax and lexical alignment is mentioned to happen, even to a higher degree in HCI (Branigan et al., 2003).

2.2 Data

Data availability is limited when researching dialogue datasets online. Not only that, manual analysis for this data is challenging: costly in time and effort, requires education prior to being done and revision is very advisable.

My first intention on this study was to look at data from other existing publications and try to draw a quantitative comparison on the data obtained from our HCI samples to those. Given the limited amount of samples in experimentation I don't think it is valid drawing the comparison another original paper; regardless I will go over it for these reasons:

- The initial study proposal (HCI and its effects on self-repair quantitatively) is valid and interesting.
- If in the future someone decides to do this study (myself included) with larger amounts of data and more rigorous conditions, I would hope to lighten their work on research.
- Some of the observations on the data papers may be applicable to the data I obtain or explain some behavior.

The data I intended to use is in Replacement and Recycling in English according to syntactic category (Fox et al., 2010) and can be seen in Figure 5. In this case the most common recycle destination (where we jump back for recycling) by a large margin are subject pronouns, followed by determiners, wh-words and prepositions. On the other hand the replaced item has a more even distribution (except wh-words, auxiliars and existentials) but Nouns appear to be much more commonly replaced. We can also see that Function words predominate as recycle destination and replaced item, where as

content words are rarer. It would be expectable to see at least see the most common substitutions happen often in our study.

3 Task Formalisation

3.1 Context

The objective is to study the occurrence of Self-Repair. This was a deliberate choice, since Disfluencies are relatively common in dialogues and the amount of data available and time to dedicate is limited. In the design, we noticed the following requirements:

- **Naturality** We want the dialogue to resemble an average human-to-human conversation; since the HCI is already rigid enough we want to at least minimize the amount of issues that could happen in the conversation
- **Bi-lateral communication** It is important that both the human and robot interact in meaningful ways, as to elicit the alignment properly if it were to happen. One-sided conversations will not work for us.

Given those constraints, we decided on a **role-play situation where the human would get interview advice from a robot**. This has several advantages:

- The speaker has a goal unrelated to the phenomena we want to study which should help with naturality.
- Role-playing parenthesis happen naturally, which help with clarification requests (another team member required them).
- Balanced turns should happen more or less naturally.

It is worth noting that we initially considered using an LLM combined with some incremental model (Schlangen and Skantze, 2011) for simplicity of development and naturality. However, this option was discarded as the overhead from including new functionalities would have added much more work, and it was decided to do a more hardcoded implementation. We still think that this would be an interesting idea, but beyond the scope for this project.

3.2 Implementation

We used the Furhat environment for implementing our robot agent, as is required per the subject. The interaction was designed always allowing to request dialogue options to the robot, and upon selection of an option the robot asked an open question (to elicit a bit of conversation from the participant) and then gave a bit of trivia (2-3 sentences). There were 2 broad fields (Resumé and Job Interview) and each field had 3 interactions where slots could be filled. Upon successfully requesting one of the interactions there was a 1/5th chance of the robot doing a clarification request. The agent also reminded after every slot filling which slots on the broad field were filled, to remind the user of what had been said. The full interaction diagram can be seen in Figure 6.

While the interaction with the robot itself was less natural than we would have wanted it do, for the purpose of eliciting natural Self-Repair this implementation proved to be enough. We would have preferred more naturality in the conversation to distract the human from our research, but from listening to the audios the goal the participants chose was to explore the dialogue options. This proved to be an equally valid distraction.

4 Experiments

At the beginning of the experiment the participants were told the general context for their conversation with the robot (the robot giving interview advice) and that they could request their options at any time. Figure 7 shows the setup. We reduced the interaction for Furhat to 2 meters, which in the experiment room ensured nobody else was captured by it. The human interacting with furhat was facing the windows (on a 2nd floor and blocked immediately by another building) to try and minimize the distractions that could affect him/her. All our project members were present during the recording, we tried interacting minimally with the speaker and only intervene in some specific instances where the speaker was using an unrecognizable words by the language detection algorithm ("Say cv instead of resume"). Our assistants' voice was never captured by Furhat, and we only addressed the participant. Transcriptions only include the proper Furhat interaction, not necessarily the introduction to the task even though that is in the audio. We encouraged participants on open questions to talk a bit further on open questions.

English	Destination of recycling	Replaced item	Total
Function words			
Subject pronoun	48 (43%)	6 (17%)	54
Determiner	13 (12%)	4 (11%)	17
Wh-word	10 (9%)	0 (0%)	10
Preposition	10 (9%)	3 (8%)	13
Aux	5 (4%)	0 (0%)	5
Existential	2 (2%)	0 (0%)	2
Copula	2 (2%)	3 (8%)	5
Connective	2 (2%)	1 (3%)	3
Other	4	5	9
Content words			
Adverb	2 (2%)	2 (5%)	4
Verb	5 (4%)	4 (11%)	9
Adjective	4 (4%)	2 (5%)	6
Noun	5 (4%)	7 (19%)	12
Total	111	36	147

Figure 5: *Self-Repair* substitutions by type (Fox et al., 2010)

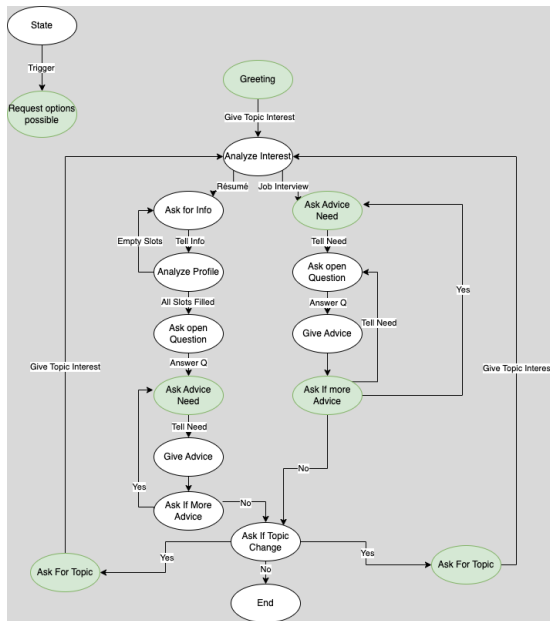


Figure 6: Flow diagram for our agent

A consent form was given prior to the interaction and an evaluation form given at the end, these can be found within the project in the "forms" folder (GiliBueno, Accessed on 2023-09-18)¹. The data from the experiment is also accessible in the repo at the moment of writing this report, but we cannot make it public as is stated in the consent form.

Only audio was collected for the purpose of our research since no team member decided to work on gestures. This was collected from a computer right near Furhat to ensure audio was received intensely enough and slight testing was done prior to it.

¹This repo is private, request can be granted asking me or another contributor.

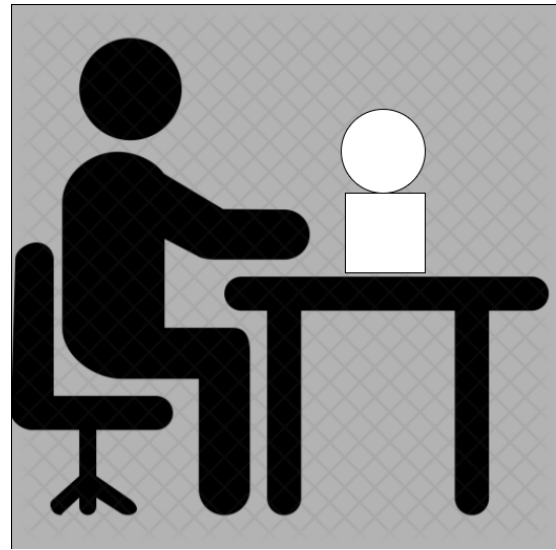


Figure 7: Set up for the interaction

5 Data Analysis

I have excluded the data from 2 participants due to not meeting minimum validity criteria. During Participant #1's conversation, some technical issues happened with the code and the dialogue looped, making the overall conversation quite jarring and long without really much interaction and a lot of repetition. After that the code issues were fixed. Participant #6 was extremely confused during all the interaction; which resulted in an extremely disconnect dialogue and has been excluded as well. This leaves us with a total of 5 dialogues and slightly below 30 minutes of recording.

There are only a total of 9 *Self-Repair* examples. Minimal context has been kept to show from where the disfluency originates. Participants 2,3,4,5 and 7 made respectively 2,1,0,5 and 1 *Self-Repairs*. These can be read in the Appendix A.

Figures 10 and 12 show an articulation error, which are usually quite rare. I would assume the increased pressure from a limited window to respond to the robot (after some tries where the robot interrupted the user) are the cause for these.

Regarding the destinations of recycling: Figures 8,11,13,14,15, all show examples where the destination is a wh-word, Figures 13,16 show a pronoun and Figure 9 shows a preposition. If we look back at the statistical data we have (Fox et al., 2010) we can see that these are 3 of the top 4 categories and we have no other examples that use another recycling point. While we would be expecting the Pronoun to be the most prevalent category, the setup for the interaction (requesting options being is very prevalent tool for the human interacting with the robot) probably explains the wh-words being much more common. In Figure 8 it is worth noting that in the first of two consecutive repairs the repair destination happens on a verb; but it is a modal verb with a role equivalent to an wh-word.

Regarding the replaced item, we really only have examples on Figures 8 (Modal verb, Wh-word), 13 (Verb) and 16 (Subject Pronoun). The data once again matches our reference paper (Fox et al., 2010), but differs in the prominence of Wh-words. I once again interpret this as a result of the setting of the conversation.

In my opinion, the unusual prevalence of wh-words unlike in our literature (Fox et al., 2010) is due to the combination of two general factors (besides the setup). First, in English self-repair tends to have as a destination word the beginning of a syntactic structure rather than the last word; when this happens it is out of convenience rather than convention. Second, I would imagine in HCI (or at the very least in interactions with Furhat) full-utterance repairs are much more prevalent. Quite often the robot will not be listening or will focus only in the sentence that it can recognize. Noticing this, participants may have a much higher preference for recycling the sentence altogether. This could be seen as an alignment in the sense that participants may be trying to match the "only full utterances" of our robot and suggests that alignment also happens in conventionalized situations in general.

6 Conclusion

We have seen some evidence that the setup is extremely relevant to the types of repairs we can ex-

pect in an interaction. We also have some data that suggests that alignment does happen even in robotic speech; even when the motivating factors behind it do not happen. The fact rigorous evidence (Fox et al., 2010) points to alignment happening across HCI and that combined with the results of this study suggest the conventionalized Self-Repair mechanics to be another aspect of alignment. Further research needs to be done in this aspect to confirm this point.

7 Limitations and Ethical Considerations

During data collection, a lot of irregularities or bad practices happened that I will go over. This does not make the research on other papers and the observations put together less valid, but the data obtained on our team's experiments should be taken with a grain of salt. As much as I have tried to be rigorous in its analysis, there are factors that make the data obtained not rigorous enough for an official research:

- All the participants (2 through 7) were familiar with the robot technology, knew how to develop an interaction itself. A conversation where you are aware of the limitations is heavily biased.
- The amount of participants and data is extremely low. While I have related the instances of self-repair to our data, the samples are just too few to have any statistical significance.
- Some participants had some exposure to the experiment prior to participating. This was inevitable, as the room we were using for experimentation linked the other group's experiment and the outside.
- Our team encouraging on talking more when responding to open questions makes the data less rigorous as well.
- While we did our best implementing the interaction, the final result was quite jarring to work with. Repeating multiple times the same sentences, being interrupted in unusual situations and navigating the interaction with the robot itself. Those 3 factors for sure fatigued or annoyed the participants, and that may bias the results. It also probably shifted the conversation from a dialog to maze-solving,

or something akin to talking to an interactive book. In hindsight this may be a better idea for the setting of an experiment, as this also puts a goal unrelated to phenomena to study and may be more fun for the participant.

References

- J Kathryn Bock. 1986. Meaning, sound, and syntax: Lexical priming in sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12(4):575.
- Holly P Branigan, Martin J Pickering, Jamie Pearson, and Janet F McLean. 2010. Linguistic alignment between people and computers. *Journal of pragmatics*, 42(9):2355–2368.
- Holly P Branigan, Martin J Pickering, Jamie Pearson, Janet F McLean, and Clifford Nass. 2003. Syntactic alignment between computers and people: The role of belief about mental states. In *Proceedings of the 25th annual conference of the cognitive science society*, volume 31. Lawrence Erlbaum Associates Hillsdale, NJ, USA.
- Holly P. Branigan, Martin J. Pickering, Andrew J. Stewart, and Janet F. Mclean. 2000. [Syntactic priming in spoken production: Linguistic and temporal interference](#). *Memory & Cognition*, 28(8):1297–1302.
- Barbara A. Fox, Yael Maschler, and Susanne Uhmman. 2010. [A cross-linguistic study of self-repair: Evidence from english, german, and hebrew](#). *Journal of Pragmatics*, 42(9):2487–2505. How people talk to Robots and Computers.
- GiliiBueno. Accessed on 2023-09-18. FurhatInterviewRobot github repository. <https://github.com/xarxaxdev/FurhatInterviewRobot>.
- Makoto Hayashi. 1994. A comparative study of self-repair in english and japanese conversation. *Japanese/Korean Linguistics*, 4:77–93.
- Martin J. Pickering and Simon Garrod. 2004. [Toward a mechanistic psychology of dialogue](#). *Behavioral and Brain Sciences*, 27(2):169–190.
- David Schlangen and Gabriel Skantze. 2011. A general, abstract model of incremental dialogue processing. *Dialogue & Discourse*, 2(1):83–111.
- Elizabeth Ellen Shriberg. 1994. *Preliminaries to a theory of speech disfluencies*. Ph.D. thesis, Citeseer.
- Fay Wouk. 2005. The syntax of repair in indonesian. *Discourse Studies*, 7(2):237–258.

A Appendix

R: Interesting. What kind of advice about writing a cv were you looking for?
H: How to write
R: I didn't get that, sorry. Try again.
H: How to write
R: I still didn't get that, sorry. Could you rephrase?
H: Hm... Can I ask my - How are - What are my options

Figure 8: Self-Repair Example 1 on Participant 2

H: I would like to know how to , uh ,
<overlap> write the content <overlap>
R: <overlap> Still couldn't <overlap> get that. Let's start over. What kind of advice about writing a cv were you looking for?
H: I wanna know about - about the content of my resume.

Figure 9: Self-Repair Example 2 on Participant 2

R: I see. How do you personally prepare for an interview?
H: Um, I guess I would look up the company details, aand, uh, try to anticipate the questions.

Figure 10: Self-Repair Example 1 on Participant 3

H: Well with what can you help me with?
R: I didn't get that, sorry. Try again.
H: What uh, what are my options what eh what-what do you - what can you help me with?

Figure 11: Self-Repair Example 1 on Participant 5

R: How many years of experience do you have in your field?

H: em, workwise eh I got my first job in 2018 so this is 5 years now alm-almost yeah

Figure 12: Self-Repair Example 2 on Participant 5

R: So you have a master's degree, have worked in 2 roles and have 5 years of working experience. What are your concerns when it comes to writing a CV?

H: No, I've- I have a bachelor's degree, you misunderstood me ah, sorry.

Figure 13: Self-Repair Example 3 on Participant 5

H: Um, I don't know. I mean, what advice can you offer?

R: I didn't get that sorry. Try again.

H: Em, what-what-what tips do you have for writing a cv? What are my options?

Figure 14: Self-Repair Example 4 on Participant 5

R: I still didn't get that, sorry. Could you rephrase?

H: What are - what are my options?

Figure 15: Self-Repair Example 5 on Participant 5

R: So you have a Bachelor Degree, have worked in 1 roles and have 3 years of working experience. What are your concerns when it comes to writing a CV?

H: I don't know what the companies want to see on the cv, I don't know which one I should - which of my abilities, which of my education and what is the most important thing that I should display in it.

Figure 16: Self-Repair Example 1 on Participant 7