ORGANIZATIONAL SEQUENCE, TURN-TAKING, AND ROLE-SWITCHING IN A HUMAN-ROBOT COMMUNICATION

Anggy Fian Febriyanti¹
Irham²

¹ Department of English Literature, Faculty of Humanities, UIN Maulana Malik Ibrahim Malang, Indonesia
Email: Anggyfian123@gmail.com

² Department of English Literature, Faculty of Humanities, UIN Maulana Malik Ibrahim Malang, 65144, Indonesia
Email: Irhamaladist@gmail.com

Abstract

Research on artificial intelligence has developed rapidly since the last fifteen years. Social studies and humanities scholars including linguists has put much attention on the relation between human-robot communications from wider and multidiscipline perspectives. Nonetheless, there is only a dearth of research investigating turn, sequence, and role-switching in humanoid interaction. Focusing on Sophia, the humanoid robot, the present study aims to examine three elements of dialogic-communication related to conversation. The present study applies a descriptive qualitative approach to investigate the research subject. Data consists of forms of utterances taken from two Sophia’s video that will be categorized into a sequence of organization, turn-taking, and role-switching functions. The results of present study show that Sophia is able to perform human-like interactional features in terms of providing adequate adjacency pairs, turn-taking, and a sequence of organizations. The present study shows that Sophia is not able to interrupt a conversation because it needs a brief moment of silence before replying.

Keywords: Humanoid Communication, Organizational Sequence, Role-switching, Sophia, Turn-taking.
1. Introduction

The development of Artificial Intelligence (AI) has become the concern of the scientists not only in the scientific and machinery fields but also in social science such as psychology, sociology, and linguistics. The latter pays more attention toward the artificial intelligence because the development in artificial intelligence has not been fully successful to simulate human’s interaction with language (Hill, Ford, and Farreras, 2015). Although AI uses human language as a communication tool, AI still finds it difficult to understand the complexity of human language (Schild, 2022). In this sense, AI designers should not only develop machine’s ability to understand the implications of words but also to comprehend the expressions in arranging words to convey meaning. It is because the interactions between human and AI can range diversely from conversations in the airlines to hospital service. In addition, it has been argued that a facile communication seems important to build a natural interaction between humans and AI (Breazeal, 1999). Thus, AI designers should put linguistics aspects in the front and center of the development of artificial intelligence, the latter may understand language in human-level (McShane & Nirenburg, 2021). The above points suggest why the development of AI becomes one of focuses in linguistics.

One of interesting cases of the development in AI and its relationship with linguistics is Sophia, the first robot to be given a country’s citizenship. Sophia is able to provide more nuanced human-like conversation features such as giving responses, recognizing people’s faces, understand humor and sarcasm (Schild, 2022). Sophia is unique because it can communicate with humans in an almost human-like level. Due to its uniqueness, Sophia has been interviewed by several human reporters. In this regard, investigating how Sophia converse with people may provide more insights on how AI can develop human-like communication. The findings can assist the development of the social humanoid robot itself, especially in the ability to
Nevertheless, there is still a lack of research which investigates conversation features used by Sophia. Arend, et al. (2017) investigated aspects of conversation, IRF sequence, performed by NAO robot and humans. They found that the NAO robot and human participants engaged in an ‘unbalanced’ communication due to the fact that the robot was only able to catch on verbal utterances without understanding the words’ social nature. Additionally, real-life interactions between human and robot have been investigated in several studies (Lee & Makatchev, 2009; Mirning, et al., 2013; Pelikan, 2015). Nevertheless, the above studies only focus on the humans’ perspective such as how humans adjust to such interaction and what the participants tend to talk about with the robots. The interaction between human and AI seems to be less explored.

In regards to the present study, it aims to investigate the conversation features found in the interviews between Sophia and human. Furthermore, the present study will examine how Sophia and the interviewers communicate, response each other’s utterance, sort out the turn amid the discussions, manage the topic of the talk and create feedback. In order to produce a more detailed analysis, the present study will only focus on utterances performed by Sophia. The present study also analyzes the similarities and differences between the features performed in human-robot and human conversations.

Furthermore, the present study will apply Conversation Analysis (CA) as a research approach. Schegloff (1992) stated that CA is the appropriate approach in analyzing the data about social interaction in the form of conversation. Moreover, CA analyzes how the conversation emerges rather than how each participant accomplishes information based on their mental process (Horton, 2017). In addition, CA is known as the common approach used to analyze talks not only between humans, but also between humans and other creatures such as animals (Logue & Stivers, 2012; Sievers & Gruber, 2016).
and even inanimate beings like robots (Mutlu, et al., 2009; Nomura, 2017). It can be said that CA seems appropriate to be applied in the present study.

2. Theoretical Review

2.1 Conversation Analysis

Conversation is an essential part of human language as it is a tool for a human to socialize and establish a connection with others (Liddicoat, 2007, p. 1). Due to its importance, conversation is analyzed in a specific analyze called Conversation Analysis (CA). Started from the study of ethnomethodology, Sacks, Schegloff and Jefferson generated the theory of Conversation Analysis to comprehend the common-sense skills and abilities that are owned by the participants. As a result, the participants are able to build and identify meaningful social interaction (Person, 1996, p. 1). In addition, Hutchby & Wooffitt (2008, p. 11) describes CA as a study that investigates the natural production of talk in people's interaction. CA has several principles that will be discussed in the following paragraphs.

One of the principles in conversation analysis is that the utterances in conversation interactions are organized sequentially (in sequences). In that case, conversation analysis deals a lot with the term “sequence organization”. This term is referred to the types of organization which concerns on the relative positioning of actions in spoken discourses (Schegloff, 2007, p. 2). Have (2007, p. 131) posits that any utterance contained in an interaction is considered to have been produced specifically for a particular position in the conversation, whether for the first or second-pair part; and at the same time, the utterance also creates the context for further utterances. For instance, if one starts with a greeting, the other speaker(s) are expected to reply with the relatable greeting. On the other hand, while one is uttering his/her complaint, the second speaker tends to reply with an apology. In accordance to conversation analysis, the concept of adjacency pairs is the main tool in sequence organization (Schegloff, 2007, p. 2). The concept of adjacency pairs arises because the utterances that appear are usually in pairs, for example in greetings, or
thanks, or also questions that generally give rise to answers from the other person. However, in reality in daily conversation, adjacency pairs do not always follow each other. Namely, after a question as a first pair part, immediately after that there is not always an absolute answer as the second pair part. That was explained by Schegloff with the concept of the presence of expansion forms of adjacency pair. The expansion is divided into pre-expansion, insert-expansion, and post-expansion.

Besides sequence organization, the practice of turn-taking is one of the most essential units in conversation. It refers to the shift or “turn” of talk managed by the participants in a dialogue, not in monologue (Sacks et al in Levinson & Torreira, 2015). The speakers could legibly figured out who should talk next and when should they speak by some signals initiated by the prior speaker. It can be indicated by the use of falling intonation, pitch, eye contacts, or gestures at the end of an utterance. Sacks (1995) proposed two general rules in determining the next speaker. First, the person who is speaking can appoint the next speaker. On the other hand, the second way is that the following speaker appoints himself to take turn to speak. Turn-taking system is required in the talk-in-interaction in order to maintain a well-ordered and comprehensible talk (Schegloff, 2007, p. 3). Furthermore, Mey (1993, p. 217) explained that the organization of conversation is closely ‘equal to the traffic rules’. The fundamental principles are to make sure that the conversation is running smoothly and not obstructed. However, it is sometimes contrasted to the empirical data on the grounds that there are clearly numerous occurrences of short overlaps and short delays or pauses. Overlapping means that a speaker begin to talk during the other’s utterances. Meanwhile, delays or pauses happen when no speaker is responding in a short period of time namely seconds (Levinson, 1983, p. 299). It can be said that turn-taking requires participants to create a balanced and comprehensible conversations.

The third principle of CA is role-switching. In a conversation, there is a
participant who has a role as an initiator while another is a responder of the talk (Schegloff, 2007, p. 3). That role might switch constantly in daily conversations. In contrast, in an institutional settings such as interviews, meetings, courtrooms, and classrooms, there is always one who is more dominant to initiate a talk. Additionally, Itakura (2001, p. 1860) stated that the domination in conversation may be brought about by social inequalities namely the social status, gender, or job-related role. The speaker who has more dominant role commonly initiates the first pair of utterances more that the other participant. For instance, in a talk show interview, the responses that the guest uttered is depending on the questions produced by the host, on the grounds that the host is the key person in managing the goals of the conversation itself. Nevertheless, considering that an interview is included the kinds of talk-in-interaction, it is possible for the interviewee to generate a first pair of utterance while the interviewer is the one who respond.

2.1 Studies on Humanoid Communication

In the recent decades, number of studies related to the way humans and advanced machines interact had been conducted in various contexts within the field of HRI (Human-Robot Interaction). For instance, some investigations of socially competent robots from a human-centered perspective were conducted to find out how robots could be greatly beneficial for humans in completing everyday tasks and roles (Dautenhahn, 2006; Sidner, et al., 2006; Lee & Makatchev, 2009). Besides showing the benefits of using robots for daily lives, the above studies imply the importance of social exchange when dealing with human. It is because dealing with humans primarily requires social exchange. In this regard, natural language plays a very important role within the social exchange. Thus, developing robots that can fluidly converse with humans becomes the priority of scientists in order to maximize communication effectiveness, and enable their quick and effective application (Mavridis, 2015). Numerous studies in the field of
linguistics analyze the ability of the robots’ (e. g. NAO, Simon, and drumming robot) to perform natural conversation with humans (Arend, et al., 2017; Cuijpers & Goor, 2017; Kose, et al., 2008; Thomaz & Chao, 2011). Following paragraphs will discuss how humanoid communication is analyzed in the above studies.

Arend et al. (2017) analyze an ‘extended’ concept of Bakhtinian ‘dialogism’. The results of their study revealed that the NAO robot and human participants engaged in an ‘unbalanced’ communication because the robot is only able to catch verbal utterances without understanding the words’ social nature. The results of their study are similar with Dautenhahn (2007) that suggests the robot’s lack of social intelligence and ability to interact effectively. Similar points have also been researched in the studies of Miring et al (2013) and Pelikan (2015).

3 Research Method

3.1 Research Design

The present study applies a descriptive-qualitative research method. The latter approach is used because it seems suitable to analyze topics related to man disciplines such as education, psychology, and social sciences (Nassaji, 2015). In particular, the present research attempts to investigate participants’ conversational behaviors when performing verbal interactions or conversation. The present study is a data-driven research in which there are no assumptions or hypothesis that should be accepted or declined in the results. Thus, as mentioned by Vanderstoep & Johnston (2008, p. 167), the results of a qualitative research design is not predictable and should be described in details. Thus, the present study will provide a detailed discussion of the data.

3.2 Research Instrument

The present study will apply the researcher’s interpretation as the main instrument. Lincoln & Guba (1985) argue that the main instrument collecting and analyzing the data of a qualitative research is the human researcher(s). In such cases, the main instrument of this research is the researchers themselves acting as a collector, listener, interpreter, and
analyst of the data; as well as the reporter of the findings and discussions. In addition, a personal computer, the internet and an office software provide supports for the present study.

3.3 Research Subjects

The present study uses the conversation between humans and ‘Sophia’ the robot from two talk shows; The Tonight Show Starring Jimmy Fallon and Good Morning Britain as the research subjects. The video is transcribed according to the transcription convention which follows Schegloff (2007, p. 265-269). A dash (-) is used to portray a halt of the utterance. Furthermore, numbers in parentheses (0.5) indicates silences or pauses between the statements in the form of seconds. Then, the colon (:) marks the prolongation of the sound; more colons are added if the prolongation is longer. The square brackets [] denote overlaps on the words. Last but not least, equal symbol (=) here is used to indicate the immediate continuant of the talks, in which happen when two speakers speak at the same time.

3.3 Data Analysis

The next step after conducting a conversation transcription is to find patterns in the conversation features. Following steps are conducted:

1. Looking for the sequences that are performed in the conversation, especially the first and second part of adjacency pairs.
2. Paying a close attention to features in conversations in connection with changing turn of speech, commonly known as the process of turn-taking.
3. Analyzing existing features in the conversations carefully to point out whether there is a switch of the speaker’s role in the conversation.

Finally, the researcher will take further investigation on the comparison between human and human-robot conversation based on the features of conversation found in the data.

4. Results and Discussions

4.1 Results

Sequence Organization
The results of present study suggest that the most commonly used interactions between Sophia and TV hosts is Question – Answer. It appears up to 12 times in total of the whole presented data. Furthermore, the other pairs that occur are Greeting – Greeting, Request – Acceptance, Offer – Acceptance/Refusal, and Information – Acknowledgement.

**Greeting – Greeting**

Greeting the interlocutor is commonly happen in talk-in-interaction, especially in the opening. It arises automatically when people meet the other and tend to build an interaction.

2. S: Hello. (0.5) Jimmy.

The first speaker did not only use a basic greeting statement but he also mentioned the name of his interlocutor. As a result, Sophia as the second speaker did the same kind of greeting.

**Question – Answer**

Asking questions is required in the conversation when people need to gain information from the other. It is always followed by an answer, whether immediately or interjected with insertion sequence such as continuant, assessment, or repair.

1. FH: Sophia, do you know what program you're on?
2. S: Yes. Good morning Britain.
3. One of the hottest morning news show in
3. Britain, and I don't mean the weather=

The pairs went smoothly without any insertion because the second speaker provided the answer directly. Moreover, not only did the robot answered to the yes/no question, it also capable to expand the answer by giving additional information about the program.

**Request – Acceptance**

In a conversation, a request happens when a speaker asks or suggests his/her interlocutor to do something. Correspondingly, the other speaker has the right to accept or refuse the request depending on the situation. The pattern itself resembles the Question – Answer one, but the intention is distinctive regarding to the context.

8. J: U:mm. Sophia, can you tell me a joke?
9. S: Sure. (0.5) What cheese can never be yours?

We can see from the data in line 8 that the first speaker, the host, was asking...
Sophia as the second speaker to do something, specifically telling him a joke. As a respond, the second speaker accepted or granted the request by saying “Sure” and immediately followed by the joke being asked before.

**Offer – Acceptance/Refusal**

Instead of requesting, people might offer a favor for the others. This act can also be delivered in the form of communication. Basically, the first speaker who performs the offering statement expected the other participant to accept the action. However, the other participant sometimes showing a refusal act or statement toward the offer, in which it is considered as an unexpected respond. Regarding to this case, the analyzed data consist of both Offer – Acceptance and Offer – Refusal pairs of talk.

| S: | Would you like to play a game of rock, paper, scissors, robot style? |
| J: | Sure. |

In contrast with the previous structure, the displayed data show an offer that is followed by an acceptance. In fact, line 27 indicated that the second speaker confidently accepted the offer given by the prior speaker as represented by an emphasis on the utterance.

**Information – Acknowledgement/Disagreement**

The type of adjacency pairs performed in human-robot conversation is information followed by acknowledgement or disagreement. It refers to time when the first speaker stated random information, and the next speaker gives a respond to show understanding or comment about the topic being said.
The simple “Yeah” in line 37 is an _utterance that implies_ acknowledgement from the second speaker. It shows that the second speaker understood and believed the information stated by the first speaker. Furthermore, the acknowledgement is also followed by compliments.

**Turn-taking**

There are three kinds of turn-taking pattern that regularly occurred in a conversation.

**Gap/Pause**

Gaps in conversation are indicated by acoustic silences in a few seconds. It may occur in the middle of a speaker’s talk or at the end of it. Moreover, those brief silences may be interpreted differently regarding to the contexts.

**Overlap**

Overlap in a conversation, as explained by Cook in Jennyfer (2005, p. 52), arises when people are

chuckling. It can also be interpreted as the moment for the host to think about initiating a new pairs of talk. Meanwhile, the silence of (1.2) seconds in line 12 happens as time for the next speaker (Sophia) to think about the respond for the prior speaker’s statement.

In line 11, the (2.0) seconds silence is analyzed as the moment for the speaker to calm down after
engrossed to incorporate in the conversation due to the urgency to correct or to complete what is being said by the other speaker. Overlap is also more likely to occur when there are more than two speakers in a conversation, like the example found in the data.

54 S: I would ask him to focus on observing and listening, (0.5) more than talking.
55 FH: =Yeah. (0.5)
56 MH: [That’s the best advice] I’ve ever heard= 
57
In the data, the overlapping talks only occurred between the male and female interviewers as both of them simultaneously responded to what Sophia said. The male host (MH) overlaps to show annoyance or disagreement of what was being said. It also happens because the female host (FH) provided brief silence in line 56 before she started to talk again. Thus, the MH assumed that it was time for him to speak. As a result, they both spoke at the same time. Primarily, Sophia as a robot does not or not yet acquire this feature as there is no evidence found in the data.

### Interruption

Interruption is closely related to overlap. It is also an action to show correction or clarification, rejection, and completion of the statement being produced (Wardaugh, 1991). However, an interruption would cause the prior speaker to stop abruptly and do not finish his/her talk.

33 FH: {hhh}, Now I have to make (it) clear that I didn't come up with this::
34 question. But, uh- this is really o:dd, but I have to a:sk
36 no:w-
37 MH: -I’ll ask it, are you single?
38 (1.8) I'm technically just a little year old, a bit young to worry about
39 S: more than a
40 romance.
41
The interruption in line 37 by the male host is a sign of completion because the previous speaker said doubtfully about asking the question to the robot. Furthermore, the researchers found an interruption done by Sophia toward the host.

16 J: Gosh, it did "ew".(1.0)Uh-
17 S: -I'm getting
18 la:ughhs. [May]be I should host the show= 
19 J: [Yeah]. =Oka:y
20 all ri:ght
21 (1.0) S- stay in your lane, girl. Uh:::
22 No.
The data in line 16 explicated that the robot can interrupt the human’s talk. However, there is a one second of silence before the speaker intended to talk again. In fact, the pause is the cause of interruption on the grounds that it can be interpreted as a signal for the next speaker to talk. Thus, it can be concluded that the interruption feature of Sophia is not or has not been as equal as human.

Role-switching

This conversational feature refers to the moment when the speaker becomes the listener and vice versa. Coates (2004) asserted that switching the role of speaker in a conversation is influenced by the turn-taking system as it is meant to have a speaker change. In other words, the conversational features occurred in turn-taking, especially gap and interruption, can lead to the process of role-switching model.

Role-switching by pause

Pause or gap between talks is the first reason that the role-switching in communication occurs like in the example below.

We can notice that the switch of role took place due to the short silence produced by the previous speaker in line 21. There are (2.1) seconds when the interviewer did not speak to initiate a first pair part of talk. Thus, it provided Sophia, as the guest, to do so. In result, Sophia took the interviewer’s role in starting the conversation by being the initiator of Summon – Answer pairs of talk.

Role-switching by interruption

Besides pause, the role-switching in conversation can also be caused by interruption. It is found in the data as will be explicated below.

At first, as has been shown in line 12, the initiator of the talk is the interviewer himself by giving his
opinion about “nacho cheese”, while Sophia provided disagreement as a respond to the prior talk. Then, the initiating role switched to Sophia as she interrupted the interviewer in line 17. Sophia produced the first pair part of talk in the form of offer, in which it will enforce the interviewer to respond by accepting or refusing Sophia’s offer.

4.2 Discussions

Feature of communication in humanoid interaction: Sequence of organization

The sequence that appears automatically in a conversation given arises to the theory of adjacency pairs proposed by Sacks (1967). The pairs of talk consist of the first and second part, in which those two parts are performed by different speaker. Naturally, the first speaker is the one who initiate a topic and the second speaker will give the response, albeit it is sometimes being intervened by an insertion sequence formed in the middle of the pair.

In the current study, the most used pair of talk is Question – Answer which appears up to 12 times in the whole data. It demonstrates that the interviewer(s) need to dig up information from and about the robot, as well as testing the robot’s ability to interact with humans. It also shows that the hosts are eager to know about the development of the humanoid robot. In addition, the robot was able to provide the answer fluently. It can even provide further explanation about the basic answer it stated. The feature of Sophia is unlike the results of Pelikan’s (2015) study which shows how their robot subject only stuck with countable particular statements (Pelikan, 2015). It can be said that Sophia uses more complicated features.

Furthermore, the other pairs performed in the talk show interviews are Greeting–Greeting, Request–Acceptance, Offer–Acceptance/Refusal, and Information–Acknowledgement. All of those pairs are well organized by the speakers including Sophia, by means that the humanoid robot can proficiently manage the coherence of the topic initiated by the interviewers.

Numerous studies show that the robot
can perform the second part of the pairs considering that the one who is more dominant in starting the pairs of conversation in a talk show is the host (Sabry, 2016; Mayasari, 2018). The weight of evidence suggests that the robot is able to produce coherent replies.

**Patterns of turn-taking**

This pattern of turn-taking consists of the hosts and Sophia as the guest star of the talk show. Nevertheless, the present study will only focus on the patterns of turn-taking conducted by Sophia in order to provide a specific analysis. The results of present study show that Sophia performed two out of three turn-taking systems occurred in the data. Those are brief pauses and interruptions. Pauses and interruptions, as well as overlaps, have been commonly happened in talk show interviews (Jufadri, 2018; Ismaliyah, 2015). It can be said that the ability of Sophia to implement the turn-taking pattern is almost similar to human. The results suggest that the mechanism of Sophia as the humanoid robot has been greatly developed because previous studies have shown that the robots’ abilities in taking the turn are quite different from human, leading to confusion and/or hesitation for the human participant (Cujipers, 2017; Arrend et. al, 2017).

**Role-switching model**

Role-switching in conversation refers to an occasion when an initiator of the talk becomes the responder and vice versa. It may take place in the process of turn-taking or when the speakers are changing the turn in talks (Coates, 2004). The results of present study suggest that the causes of role-switching found in the data are the production of silence and an abrupt interference, making the speaker discontinued his/her speech. The results of present study suggest that Sophia can switch its role from a second speaker to be the first one through both pauses and interruptions. Such ability is unlike On the other hand, studies conducted by Pelikan (2015) and Arrend et. al. (2017) show that the humanoid robots only can listen and respond to the human participant, not able to switch its role. It can be concluded that Sophia has
improved drastically from its predecessors.

**Human vs Human-robot communication**

Language is not only exclusively used by humans. Robots also use a language as a tool of communication. However, there are differences between the use of language in humans and robots communication. The findings of present study data show that an interruption happens regularly in human’s conversation. Similarly, the analysis revealed that the humanoid robot can also perform an interruption when conducting a talk with human. Nevertheless, the interruption is not as powerful as one conducted by human because basically there was a short moment of silence in the interviewer’s speech and the interviewer himself was thinking about what should be said next. Consequently, Sophia interrupted the host’s talk due to the assumption that the host was already done talking. It means that the humanoid robot needs a signal that shows a chance for it to speak. The signal itself is in the form of silence.

Furthermore, the distinction between human-human vs human-robot communication is the presence of overlap. The latter refers to an event when two or more people are talking at the same time (Iman & Winata, 2021). Overlapping is natural to be happened in human conversation, both in institutional setting such as interviews and in every day talks (Jufadri, 2018; Martinez, 2018). Conversely, the findings of present study do not suggest any sign of overlap produced by the robot. It can be said that the humanoid robot is unlikely to perform overlap.

**5. Conclusion**

The present study aimed to analyze conversation features given by Sophia, a humanoid-robot interviewed by two talk shows. The findings of present study show that the sequence of human-robot conversation consists of Question–Answer in which the host acted the first speaker and Sophia as the responder. In terms of the turn-taking pattern, the results show that Sophia is able to maintain the effectiveness of the talk, showing no
serious problems such as hesitation or confusion occurred when the speakers are changing the turn in conversation. For the role-switching model, the findings suggest that Sophia has the ability to switch its role as the second speaker to be the initiator of the talks. The switch of role itself is supported by the implementation of pause and interruption. Additionally, the significant distinction between humans and human-robot conversation occurs in the process of overlapping. Unlike humans, the humanoid robot cannot interfere when someone is speaking as it needs brief silence as an indication for it to start speaking. It can be concluded that the conversation conducted by the hosts and Sophia went smoothly, although the robot has not reach the level of humans’ finesse in using language as a communication tool. The findings of present study may provide more insights on the development of humanoid robots, especially in developing ones that converse in natural language.

REFERENCES


