



Social Interaction. Video-Based Studies of Human Sociality.  
2023 Vol. 6, Issue 1  
ISBN: 2446-3620  
DOI: 10.7146/si.v6i1.137249

## ***Social Interaction***

### ***Video-Based Studies of Human Sociality***

---

#### **How Conversational are “Conversational Agents”? Evidence from the Study of Users’ Interaction with a Service Telephone Chatbot**

---

***Andrei Korbut***

Center for Advanced Internet Studies

#### **Abstract**

The paper considers whether it is possible to view interactions with so-called conversational agents (chatbots, voice assistants, etc.) as a form of conversation. It is argued here that such conversational agents *are* conversational in a proper sense. To justify this conclusion, the analysis of the beginnings of 100 calls to a Russian municipal call center, processed by a chatbot, is conducted. The revealed features of the inquiry formulations, silences, and overlaps at the beginning of the calls show that users deal with the chatbot as a conversational partner and not as a voice user interface. It is proposed that to call an interaction a “conversation,” it is enough that at least one co-participant (the weak participation requirement) is able to understand all the turns in the interaction (the strong analyzability requirement) as part of the ongoing conversation.

*Keywords:* conversational agents, ethnomethodology, conversation analysis, chatbot, call center

## 1. Introduction

The more technologies penetrate everyday life, the more acute becomes the question of their agency (Bennett, 2010; Gibbs et al., 2021; Harbers, 2005; Latour 2005; Slack & Wise, 2005: 137–147; Suchman, 2007; Verbeek 2005). This question is especially timely today, when a range of technologies are available that display capacities one would expect from human beings and that, therefore, make it very easy to attribute some agency to them. In this respect, conversational technologies are a perfect example. There is nothing more habitual to most humans than conversation, and nowadays, when engineers and system designers still struggle to build human-like robots, “conversational agents” have become the gold standard of unobstructive, familiar technology: they seem to be maximally natural. This is one of the reasons—along with economic efficiency—why conversational technologies are extremely popular not only in the media (the movie *Her*, 2013, directed by Spike Jonze, is an illustration) but also among developers—there are a number of voice assistants for mobile phones, computers, smart speakers, smart homes, and cars that are able to communicate with us in a conversational mode.

Developers of such technologies tend to use the term “conversational agents”.<sup>1</sup> A conversational agent is a dialogue system that can not only “understand” users’ natural-language utterances but is also able to respond using natural language. Of course, much of the reason for the name “conversational agent” is marketing—it is called so to be perceived by potential customers as more capable and human-like than, say, voice user interface. More importantly, the term conversational agent also embodies the developers’ hope that they can create interactional partners that humans will relate to in the way they relate to fellow humans. The question is whether “conversation” is a proper description of the way humans interact with such interfaces.

Some researchers suggest that conversational agent is a misleading name. Porcheron et al. (2018) believe that conversational interface is a misnomer because it “confuses interaction with a device *within* conversation with an *actual* conversation” (p. 9; emphasis in the original). They argue that we should distinguish between the interactional embeddedness of voice user interfaces and the conversation. The difference can be illustrated by how humans deal with question–answer sequences. In conversations, the question is a feature of the pair: what makes an utterance a question is the next turn that can be perceived as an answer. In interactions with voice user interfaces, the question is predetermined by design. Porcheron et al. (2018: 8) provide an example in which Alexa, the conversational agent from Amazon, treats the user’s instruction as a question when it is not. Porcheron et al. (2018) suggest that users “routinely treat this as problematic and troublesome output that needs fixing in some way or another, rather than as a response that recasts their own utterance as a question

---

<sup>1</sup> For example, this is how Google presents its chatbot Meena (Adiwardana & Luong, 2020).

(which can be something conversationalists do)” (p. 9). This is a very tempting argument against calling voice user interfaces conversational agents, but the topic deserves a more thorough inspection because we can, of course, find instances of human-human interactions when a co-conversationalist mistakenly takes a partner’s utterance for a question when it is not,<sup>2</sup> and this is perceived as a troublesome matter and is properly corrected instead of being perceived as a “response that recasts their own utterance as a question” (Porcheron et al., 2018: 9).

The emphasis on the “conversation” and not on the “attributed agency” in the present study is justified by the need to consider agency in human-machine communication not only as a “situated construction that is lodged in the production and interpretation of meaning in the developing interaction” (Krummheuer, 2015: 195), but also as something related more to interactional organization than to participants’ affordances. When we consider interaction with conversational agents through the lens of distribution of agency among participants, we still apply the “individualist conception of agency,” just displacing “biological individual with a computational one” (Suchmann, 2007: 240). This is the same logic that system developers rely on when they describe their products as being more or less human-like. To go beyond this framework and see agency as a social achievement, we have to focus on how humans organize their interaction with technological objects. In the case of conversational agents, this requires an analysis of how humans use conversational resources to produce situated order and, hence, whether they consider the ongoing interaction a conversation at all.

To explore whether users of voice user interfaces consider their interactions as conversations, we need to look into the real-world situations of interactions with conversational agents. There is an obvious lack of detailed data on how conversational technologies function “in the wild,” that is, not in laboratory or game settings. The present study provides an analysis of the everyday usage of one such agent: a voice-based telephone chatbot that answers calls in a municipal service call center in a large Russian city. The chatbot considered here helps callers to obtain official information concerning various state services. For callers, a phone call is a familiar, natural activity. When they talk to the chatbot, they have to find a way to accomplish an ordinary, habitual task by coordinating their activity with an unusual and unfamiliar interactional partner. The “naturalness” of the situation is determined by the fact that users have no reason to consider what happens a hoax or a test, although at the same time they cannot take it as business-as-usual because human-chatbot communication in real-world situations is still something new to most callers.

---

<sup>2</sup> As in the case of so-called “rhetorical,” or “reversed polarity,” questions (Koshik, 2005).

In this paper, I analyze the opening sequences of the conversations<sup>3</sup> with the chatbot. This focus is justified, first of all, by the immense importance of the first turns in any conversation: “The ability to open a conversation with another person is . . . fundamental to conversational competence” (Moore & Arar, 2019: 150). There are particular ways of opening and a corresponding distribution of rights and possibilities of talk. For interlocutors, the way the conversation starts may be a window on the omnirelevant properties of the following interaction. The second reason for the focus on the beginning is that in conversations with computer systems like chatbots, the important task that users solve is the identification of the system’s capabilities: “One challenge with conversational interfaces is the *discoverability* of their features” (Moore & Arar, 2019: 161; emphasis in the original). For users, the system’s abilities are not something that can be decided in advance but must be revealed through the conversation itself. The first turns in conversation are the primary place where such work is done. This assessment of the system’s competence is of direct import for users when deciding whether they want to talk to the system in the first place. In this sense, the beginning in human-chatbot conversations can significantly influence the trajectory of the following conversation.

I consider three principal phenomena that reflect the basic structural features of any conversational beginnings. If, “overwhelmingly, one party talks at a time” (Sacks et al., 1974: 700), then the central phenomena of the conversation are: (a) “no gap, no overlap” talk, (b) gaps, and (c) overlaps. These are the three categories into which I sort my findings.

I focus on two opening turns in telephone conversations with the chatbot to analyze what work is done by the callers, how it is done, what difficulties the callers face, and how they overcome those difficulties. To this end, I utilize an ethnomethodological approach and, in particular, conversation analysis (CA). Extensive literature on ethnomethodology (Livingston, 1987; Garfinkel, 2002; Francis & Hester, 2004; Liberman, 2013) and CA (Hutchby & Wooffitt, 2002; ten Have, 2007; Schegloff, 2007; Sidnell, 2010; Sidnell & Stivers, 2012; Garcia, 2013; Clift, 2016) relieves me of the necessity to present the foundations of this approach and its major conceptual and methodological tools. Suffice to say that I focus not on the *structure* of conversation per se, but on the *work* that is done through and as this structure.

The paper has the following organization. First, in Section 2, I consider the existing findings from CA and ethnomethodology related to the beginnings of

---

<sup>3</sup> It must be said here that by calling interactions with the chatbot “conversations,” I do not mean to decide in advance whether they are “conversations” for the involved human participants. This last question should be decided on the basis of the analysis of participants’ activities. I call them “conversations” just to indicate that they are “speech exchanges” (Sacks et al., 1974: 696), without making any assumptions about their inner organization. This also means that, as I deal exclusively with conversational data, in this paper the terms “interaction” and “conversation” are interchangeable as descriptive categories, although, of course, it is possible to distinguish them analytically.

telephone calls and communication with artificial agents. Then, in Section 3, I provide a concise description of computational architecture of the chatbot studied. After presenting my dataset (in Section 5), I analyze the three fundamental aspects of the beginnings of conversations with the chatbot: producing “no gap, no overlap” talk (Section 6), gaps (Section 7), and overlaps (Section 8). Finally, in Section 9, I summarize my findings and show that the provided analysis suggests that users communicate with the chatbot as a conversational partner, and not as a voice user interface.

## **2. Findings from Previous Studies**

There are two kinds of literature that I build my analysis on here. The first is the studies of openings in telephone calls; the second is the ethnomethodological and CA studies of interactions with conversational agents.

### **2.1. Openings in telephone calls**

Studies of the beginnings in telephone conversations were initiated by Schegloff (1968, 1979, 1986, 2002a, b) and developed later by others (see, e.g., Hopper, 1992). Schegloff described four core constructional sequences of telephone openings: (a) summons/answer sequences, (b) identification/recognition sequences, (c) greeting sequences, and (d) initial inquiries, with (e) “getting down to business,” added by Pallotti and Varcasia (2008). The value of this description of the canonical opening structure for the present analysis is that it shows that the opening turn, produced by the summoned interlocutor, may consist of several elements, which poses the specific problem for the caller of choosing what part of the previous turn to respond to. It is also important that the path from the summons to the formulation of “the business at hand” tells the caller what conversation they are participating in and what competences the answerer possesses.

These canonical opening sequences, being interconnected and ordered, may have a cultural specificity (see, e.g., Lindström, 1996, on Swedish calls; Houtkoop-Steenstra, 1991, on Dutch; Sifianou, 1989, on Greek; Park, 2002, on Japanese and Korean; Hopper and Chen, 1996, on Taiwanese; Taleghani-Nikazm, 2002, on Iranian; and ten Have, 2002, for a general discussion), but for the study of human-chatbot interaction it is variations of the openings, related to the institutional circumstances of the call, that are more important. As was shown by Danby et al. (2005), Zimmerman (1992), Wakin and Zimmerman (1999), Cromdal et al. (2012), Leydon et al. (2013), and Vinkhuyzen et al. (2006), institutional calls have a particular opening organization that has consequences for the course and ending of the whole call. This organization presupposes various modifications of the canonical structure, for example, the merging of answer and identification in one turn or skipping the greeting sequence. The other

important feature of institutional calls is their topical orientation: people calling, say, to service lines formulate their inquiries as call-center-specific questions or requests that, they feel, are expected from them. As we shall see later, this plays a particularly significant role when the co-conversationalist is a chatbot, not a human.

## 2.2. Ethnomethodological studies of interaction with conversational agents

The most important part of the research literature that the present study relies on is ethnomethodological and CA studies of interactions with artificial conversational agents. These studies began as soon as the conversational agents became real-world entities that people could communicate with. Starting with the groundbreaking study by Suchman (1987) and continued by the detailed analysis of various conversational systems (Luff et al., 1990; Thomas, 1995; Wooffitt et al., 1997), ethnomethodology and CA provided a framework for researching the practices of using and developing conversational agents. These findings are what the present paper's argument – that users interact with artificial agents as conversational partners – is built on. There are, however, some further steps that need to be taken.

The work of Moore at the intersection of the User Experience (UX) and CA deserves special attention. Although the possibility of developing software that can converse as humans do is not indisputable (see Button & Sharrock, 1995), Moore shows “how to *model* natural conversation” (Moore & Arar, 2019: xiv; emphasis in the original) and “how to string bits of natural language together into naturalistic conversational sequences” (Moore, 2018: 182). To do so, he and his collaborators created the Natural Conversation Framework (NCF) that contains a library of modules or patterns of common conversational activities. Each pattern is based on CA findings concerning everyday practices of doing conversation. For example, the opening pattern may consist of greeting, self-identification, organizational identification, and offer of help which Moore & Arar (2019) have described as “the canonical opening for a service encounter” (p. 154). The conclusion that Moore and Arar (2019) come to is that “[e]ven though today's chatbots and voice assistants cannot handle domain-independent *viva voce*, they may be able to understand what the user says and does well enough to answer inquiries, fulfill requests, or troubleshoot problems for all practical purposes” (p. 22). Their practice-based answer to how “conversational” conversational agents are is that these systems can “understand.” They distinguish such “understanding” from the “interpretation” of user's utterances in Natural Language Understanding (NLU) techniques: “Understanding is not the same thing as interpretation. Interpretation is the analysis of the language and the action of an utterance, but *understanding is the demonstration of correct or adequate interpretation of social action within interaction*” (Moore & Arar, 2019: 23; emphasis in the original). While agreeing with Moore and Arar's general conclusion, in this paper I show that “conversational agents” are conversational

not because they are able to understand (i.e., can be viewed as understanding for all practical purposes), but because in every conversation every action is *analyzable* as an action in the ongoing conversation.

Moore's approach aims to reconcile CA studies with the development of conversational computer interfaces and is therefore restricted by the technical properties of such interfaces. This must be supplemented by studies of the actual ways people interact with such systems. Detailed analysis of the real-world encounters with some of these systems can be found in the work of Porcheron and colleagues (Porcheron et al., 2017; Reeves, 2017; Porcheron et al., 2018). Their work "does *not* concern itself with questions about whether a computer that 'talks' as-if-it-were-human can be created and sets aside such concerns, instead orienting to an ethnomethodological perspective of unpacking how interaction with VUI [Voice User Interface] is achieved *within* talk-in-action" (Porcheron et al., 2017: 2; emphasis in the original). I take the same perspective in this paper because it makes the analysis of the interaction with conversational agents a matter of examining the actual details of the ongoing collaborative work by the co-participants. The analysis of conversations with Amazon's Alexa by Porcheron and collaborators contains observations that are helpful in my endeavor. The most important of them is how silence in conversations with such agents becomes a source of trouble and how users invent various strategies to overcome emerging difficulties. However, their analysis differs from mine in some critical respects. First, the requests they analyze are requests to a digital assistant; therefore, the nature of these requests is different from the request to a telephone call-center chatbot. The main difference is that when addressing the digital assistant, the user asks it to *perform* some action, while when addressing the chatbot, the user *inquires* about particular information and does not *request* the chatbot to make an information search. For example, a caller may say: "I would like to know is my labor patent ready or not" but never: "Find information about the readiness of my labor patent." The second difference is that conversations with digital assistants are more often interwoven into other kinds of concurrent actions and may be embedded in different practical situations than calls to the call center. The problems that users face in these settings and the work they have to undertake to make them ordered and accountable are different. For example, in many cases callers have to provide additional information about their inquiries, such as the registration number of the application, and to do this they need to have the relevant documents at hand, which limits their ability to maintain a parallel course of action.

The present analysis of human-chatbot conversations also builds on many findings from ethnomethodological and CA studies of human-robot interaction (HRI). While interaction with robots requires the use of the body and gaze, speech is also involved and, therefore, the corresponding data can be used to enrich the analysis of conversations with a chatbot. Previous ethnomethodological HRI studies have analyzed how humans deal with breakdowns in HRI (Arend et al., 2017), accomplish turn-taking in communication with a humanoid robot (Pelikan

& Broth, 2016), are encouraged to engage with a robot (Pitsch et al., 2009; Gehle et al., 2017), and categorize the identity of participants (Krummheuer, 2016). Of particular interest here is the work by Pitsch and colleagues on the openings in HRI (Pitsch, 2015; Pitsch et al., 2009; Gehle et al., 2017). They show what complex collaborative work should be done by the user and the robot to initiate a conversation. This work can be both concerted and non-concerted, and each influences the engagement of the user. The observations of the study are centered around the use of gaze by the robot to “catch” and maintain the attention of the user. However, the conversational resources considered in the present study are different from the bodily resources available to the participants of HRI. This difference should not be overestimated, though. As the study by Pelikan and Broth (2016) shows, in certain situations users may ignore the non-conversational information conveyed by the robot, such as sound signals. Although many of their observations are confirmed in the present study,<sup>4</sup> it is also evident from their study that users sometimes use the robot’s body as an interactional resource (e.g., searching for the cues concerning their next action by looking at its face). An important question that my data poses is how users manage the sequential order of conversation with a chatbot when they have at their disposal only the resources provided by and produced within the conversation itself. Another interesting question is how the institutional character of the conversation influences its organization from the first turns.

This study continues and broadens previous ethnomethodological and CA studies of interactions with conversational agents and robots by providing a detailed analysis of real-world conversations with a chatbot and describing methods that users employ to make these conversations manageable, intelligible, and ordered. I supplement the findings from the previous studies by highlighting the particulars of the interactional work that users do in task-oriented institutional conversations.

### **3. What is a Conversation?**

Before starting the empirical analysis of the interactions with a chatbot, I will clarify what is meant by conversation in this paper.

“What is a conversation?” is a very complex question that cannot be dealt with comprehensively here. However, I can make some points that I rely on in the present study. The most systematic attempts to describe conversation are made in CA, so this is where we should search for the definition of conversation. The problem is that conversation analysts, with their strong distaste for theorizing social phenomena, provide only glimpses into how conversation can be defined. For example, Goodwin and Heritage (1990: 284) suggest that CA uses Goffman’s

---

<sup>4</sup> For example, the observation that overlaps and silences in conversations with a robot are conditioned by problems with projecting and determining the boundaries of the robot’s turns and also that human users solve some of the problems by changing word selection and turn length.



loose definition of conversation as every talk or spoken encounter. This is obviously too broad a definition. We need more precise indications of how such encounters have to be considered by the analyst. A more focused discussion of what conversation is can be found in Harvey Sacks' *Lectures on Conversation* (1992, vol. 2: 36–38). Rejecting the idea that conversation can be defined by referring to how the term “conversation” is used in ordinary talk<sup>5</sup> or by listing some required parts or elements of conversation, Sacks suggests (1992) that “‘doing conversation’ is behaving according to certain sorts of orderly procedures” (p. 37). These procedures are “one party at a time” and “speaker change.” They are formal features of conversation and there are means of providing them. Such features make participant actions *analyzable*, that is, participants understand each other's actions by the use of these features and display this understanding. Participants *should* demonstrate their understanding of the other's prior turn in their current turn: “It is a systematic consequence of the turn-taking organization of conversation that it obliges its participants to display to each other, in a turn's talk, their understanding of other turns' talk. . . . [S]uch understandings are displayed to co-participants, and are an important basis for the local self-correction mechanism of conversation” (Sacks et al., 1974: 728). This is most evident in the case of “adjacent pairs” such as greeting-greeting, complaint-apology, or question-answer.

If we examine the details of the working of “conversational agents,” it seems that they do not display such “understanding” of the human co-participant's turns. There are only some “entities” that are recognizable for a computer system, and these entities are not “greetings,” or “complaints,” or “questions.” They are words reconstructed from the sound patterns detected by a speech recognition module. At the same time, all “conversational agents” are programmed to follow the procedures “one party at a time” and “speaker change.” Of course, these procedures are provided by the programmers, and I will return to this question after the empirical analysis. How to deal with this ambiguity – the simultaneous lack of understanding of the co-participant's actions and adherence to the most formal procedures of the conversation by the conversational agent?

To make the issue more manageable, it is useful to introduce two distinctions. First, we can distinguish between two expectations of conversational participation. The *strong participation requirement* is the expectation that all co-participants must demonstrate their understanding of the others' actions in conversation. The *weak participation requirement* is the expectation that at least one co-participant must demonstrate their understanding of others' actions in conversation. It seems that most of the criticism of the name “conversational agent” is based on the strong participation requirement: if a computer system

---

<sup>5</sup> However, in the foundational paper of conversation analysis, Sacks et al., (1974: 696) describe “conversation” as a member of the set called “speech exchange systems,” the other member being “interviews, meetings, debates, ceremonies.” It seems from this characterization, that for them “conversation” is something mundane, non-institutional, which corresponds to how this term is used in many ordinary situations.

does not demonstrate understanding of the human co-participant's utterances and silences, interaction with it cannot be called "conversation." And vice versa, if a computer system demonstrates such understanding, as Moore suggests, we can call it conversation. Later, I show that when analyzing interactions with chatbots and similar agents, it is more relevant to use the weak participation requirement and, in this case, they *can* be called "conversational."

The second distinction we can draw is between analyzable and non-analyzable conversational actions. The *strong analyzability requirement* is the expectation that all actions in a conversation must be analyzable by at least one co-participant. The *weak analyzability requirement* is the expectation that some actions in a conversation must be analyzable by at least one co-participant. I suggest that for interaction to be called "conversation" it has to satisfy the strong analyzability requirement. If it only satisfies the weak analyzability requirement, it cannot be called "conversation." In the following analysis, I show that interactions with chatbots *are* conversations in this sense.

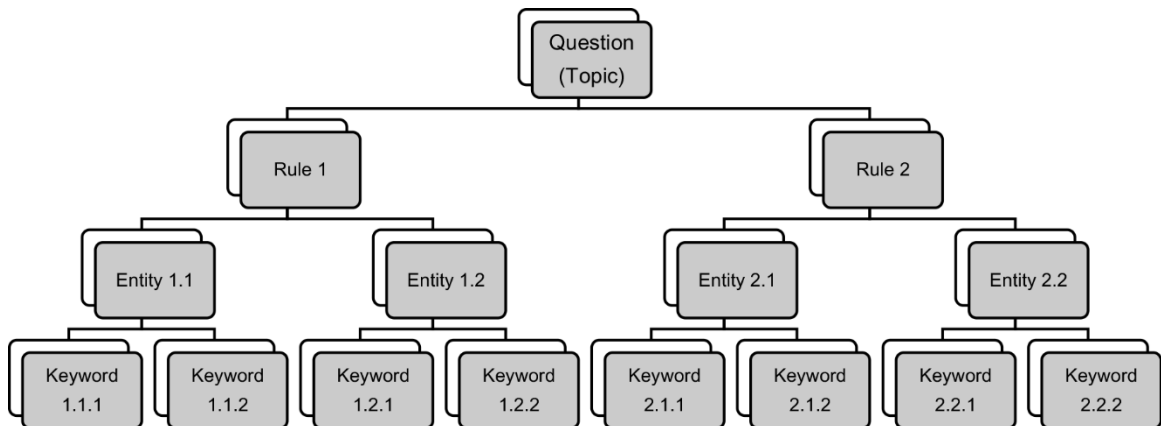
#### **4. The Chatbot**

The chatbot analyzed here was introduced in 2015 in a major Russian city's municipal call center as a configuration of several technologies, the main ones being speech recognition and speech synthesis. The chatbot answers calls from local residents and provides official information on particular topics, such as the status of the applications for different governmental services (for example, passport change), municipal plans concerning the building they live in, and telephone numbers and addresses of municipal services. At the end of 2019, there were 62 questions grouped into 25 threads. The most popular inquiries are about the status of the documents that callers had applied for at centers for governmental services (about 19%) and about the contacts of the asset management companies (about 15%). When the caller's question is out of the chatbot's topical range or there is some problem with communication, the chatbot forwards the caller to a human operator.

The call center where the chatbot operates is one of the largest in the city; as of the end of 2019, it processed about 3,000,000 calls per month. About 500,000 of them were processed by the chatbot.

To understand the interaction with the chatbot, we need to understand some of its inner workings. The general functional scheme is as follows:

Figure 1.



The chatbot recognizes separate words in the caller’s speech and, using specific rules, matches them to lists of keywords. From the keywords identified, it then decides which entity, rule, and topic the particular request can be matched to. For example, if a user says: “I want to know my electricity bill arrears,” the topic will be “Incorrect charges of Unified Payment Document.” Here, the keywords are “I,” “know,” “electricity bill,” “arrears”; the entities are “I,” “housing costs,” “know,” “arrears”; and the rules are “know housing costs” and “know arrears”. When the topic is determined, the chatbot accesses the database and provides an answer. If necessary, it asks additional questions to clarify the inquiry. To give a sense of how this happens, here is a typical conversation.

**Fragment 1**

```

1  ЧТВ  здравствуйте (0.4) вас приветствует
    СНВ  hello (0.4) I am
2      виртуальный оператор (.) XXXXXX XXXXXXXXXXXX
    a virtual operator (.) of XXXXXX XXXXXXXXXXXX
3      XXXXXX XXXXXXXXXXXXXXXXXXXX XXXXXX. (1.1)
    XXXXXXXXXXXXXXXXXXXXXXXX XXXXXX. (1.1)
4      какой у вас вопрос.
    what is your question.
5      (2.4)
6  ЗВН  хачу узнать гатоу ли мой патэнт.
    CAL  wanna know whether my labor permit is ready.
7      (3.2)
8  ЧТВ  назовите (0.3) пожалуйста (.) номер заявления
    СНВ  please (0.3) tell the number of labor permit
9      на патент (0.2) выданного в миграционном центре
    application (0.2) provided in the migration center
10     (1.7)
  
```

11 ЗВН <ноль ноль ноль (0.4) сорок пять (0.5) сорок  
**CAL <zero zero zero (0.3) forty five (0.5) forty**  
12 (0.3) триста двадцать>  
**(0.3) three hundred and twenty>**  
13 (4.3)  
14 ЧТВ патент готов. пожалуйста совершите платеж  
**СНВ patent is ready. please make**  
15 по (.) НДФЛ в миграционном центре. (0.3)  
**(.) a PIT payment in the migration center. (0.3)**  
16 всего доброго. (0.2) до свидания. (0.5)  
**all the best. (0.2) goodbye.**  
((PIT stands for Personal Income Tax))

Here, the chatbot calls itself a “virtual operator.” This name was introduced, along with some other changes, in 2018 to make the caller’s experience more “natural” and the callers themselves more satisfied with the call center’s handling of their inquiries. I will indicate just two changes which will be important for the following discussion. First, in the past when the call was answered, the first thing the caller heard was an announcement: “Hello, you’ve called the XXXXX call center. You are being served by a robot. For quality control purposes, the conversation may be recorded. Please, state your question clearly and speak after the beep.” Now, when a caller summons, the first thing they hear is: “Hello. I am a virtual operator of the XXXXX call center. What is your question?” Secondly, previously at the beginning and in some other places in the course of the call (for example, after the chatbot asks a caller to make an action, such as provide the registration number of the application) there was a beep after which the caller was to speak. There are no beeps anymore. I consider the impact of these changes on human-chatbot interaction below. Suffice to say that now both the “old” and “new” chatbots have one rule when dealing with the caller: they wait for four seconds before taking their next turn. The other feature worth mentioning is that both the “old” and “new” chatbots react in the same way to the caller’s silence after the chatbot’s turn: if the caller does not talk for four seconds, the chatbot says “Please, don’t be silent,” and, if the caller continues to be silent, the chatbot adds “Speak.”

## 5. The Data

The dataset for this study is 100 recordings of conversations with the new chatbot (complete set of all calls received by the call center in several minutes on one day in May 2019), which was introduced at the end of 2018. I use data only from the new chatbot because it was specifically designed to overcome some of the major shortcomings of the previous version. The calls last from twenty seconds to several minutes. All recordings were transcribed using Gail Jefferson’s notation

system (see the transcription conventions at the end of the paper)<sup>6</sup> and were anonymized to exclude the possibility of identifying either the callers or the service line (all the identifying information was either concealed or changed). The recordings were provided by the governmental organization that runs the call center.

I start my analysis of the features of the conversations with the chatbot by examining the inquiry formulations in the “no gap, no overlap”<sup>7</sup> mode.

## 6. Inquiry Formulation

Inquiries are a widespread phenomenon in ordinary conversations: “The inquiry pattern is perhaps the canonical, conversational sequence pattern” (Moore & Arar, 2019: 90). In service calls, inquiries are the central feature of the interaction. The “normal” sequence of the beginning of a service call presupposes that after the introductory phase the caller presents their inquiry and then the operator provides an answer or asks a clarification-seeking question. When the operator is an artificial agent, this sequence may undergo some specific transformations. In this section, I consider the most prominent of the transformations.

Let us first analyze a conversation in which the inquiry is formulated in a similar way to how inquiries are formulated in conversations with human operators.

### Fragment 2

1 ЧТВ здравствуйте (0.4) вас приветствует  
СНВ hello (0.4) I am  
2 виртуальный оператор (.) XXXXXX XXXXXXXXXXXX  
a virtual operator (.) of XXXXXX XXXXXXXXXXXX

---

<sup>6</sup> The English translations of the transcripts are by me and, to some degree, they inevitably distort the real picture of the analyzed interactions. For example, the Russian language has more direct ways of making conversation formal, compared to English. However, in the present paper I focus only on the organizational properties of conversations, and not on the interactional effects of grammatical features of the used language. The latter aspect deserves a special analysis. An example of such analysis, based on the comparison between Russian and Brazilian Portuguese, can be found in Boden & Guimaraes (2012).

<sup>7</sup> An important qualification has to be made: “no gap, no overlap” is understood here as a relational phenomenon. Fragments in Section 6 show no overlaps, but they also show some silences after the chatbot’s first turn, which can be considered “gaps.” However, our data show that there are also silences after human operators’ first turns in this call center, although they are usually shorter and there are more cases when such silences are absent. This means that the presence and the absence of the gaps have to be considered in relation not only to the ordinary conversations, but also to the institutional settings where participants may have more freedom in extending inter-turn silences without turning them into “gaps.” Or, to put it differently, “normative gap durations may be differently calibrated across languages, cultures . . . and activity contexts” (Hoey, 2020: 15).



6 (1.4)  
 7 ЗВН прописка  
**CAL residence permit**  
 8 (4.2)

The inquiry in line 7 is very different from the inquiry in lines 6–9 of Fragment 2. The former is just a single word and has no signs of hesitation. It seems more like a “search inquiry” in a search engine. The difference can be explained by the divergent understanding of what the chatbot asks from the caller. The caller in Fragment 2 sees the chatbot’s “What is your question?” as a request for a particular “problem” she wants to ask the call center about, but “question” here can also mean the general “topic” – the class of questions the particular question falls into. In the latter case, the chatbot may be perceived by the caller not as an “operator” that answers the inquiry, but as a “receptionist” that decides where the inquiry should be forwarded.

As these two fragments show, the institutional context of the interaction with the chatbot is not insignificant for the callers. The caller and the chatbot take a specific institutionalized position within these conversations, but this institutional context can play its role only in, and as, particular conversational sequences. The sequences have their own properties that contribute to the institutional understanding of what is going on at each moment of talk or silence. Among these properties, the most prominent ones are prosodic.

The chatbot’s speech has recognizable “robotic” characteristics: it is hearably “artificial.” Some of these characteristics cannot be conveyed with CA notation, but others can: in the previous examples, we saw quite long intra-turn pauses in transition-relevance places, plus the chatbot’s speaking is noticeably “flat,” with almost no pitch modifications, stresses, accents, or significant changes in intonation. As I was told by the developers of the system, this was, in part, done on purpose, because Russian law prohibits “deceiving” customers about whether they are speaking to a chatbot or a human. The other reason is the imperfection of speech synthesis technology. Whatever the reasons for making the chatbot’s voice recognizably robotic, callers orient to this feature of the chatbot’s talk to perform institutionally relevant actions and receive the desired information. Consider the following two fragments.

#### Fragment 4

1 ЧТВ здравствуйте (0.4) вас приветствует  
**СНВ hello (0.4) I am**  
 2 виртуальный оператор (.) XXXXXX XXXXXXXXXXXX  
**a virtual operator (.) of XXXXXX XXXXXXXXXXXX**

3           XXXXXX XXXXXXXXXXXXXXXX XXXXXX. (1.1)  
**XXXXXX XXXXXXXXXXXXXXXX XXXXXX. (1.1)**  
4           какой у вас вопрос.  
              **what is your question.**  
5           (1.5)  
6   ЗВН    .hh: (0.5) <ПОКАЗАНИЯ ПО ВОДЕ>  
          **CAL .hh: (0.5) <WATER VALUES>**  
7           (4.0)

### Fragment 5

1   ЧТВ    здравьствуйте (0.4) вас приветствует  
          **СНВ hello (0.4) I am**  
2           виртуальный оператор (.) XXXXXX XXXXXXXXXXXX  
              **a virtual operator (.) of XXXXXX XXXXXXXXXXXX**  
3           XXXXXX XXXXXXXXXXXXXXXX XXXXXX. (1.1)  
              **XXXXXX XXXXXXXXXXXXXXXX XXXXXX. (1.1)**  
4           какой у вас вопрос.  
              **what is your question.**  
5           (2.1)  
6   ЗВН    ↑ГОТОВНОСТЬ ↓ДОКУМЕНТОВ  
          **CAL ↑documents ↓readiness**  
7           (3.9)

In lines 6 of both fragments, the callers undertake similar work: they form their utterances in such a way as to make them specifically accessible to the chatbot. They manage the details of their talk (speed, volume, intonation) to facilitate the chatbot’s task of recognizing both their speech and their inquiry (the same strategy is used in interactions with embodied robots, as Pelikan and Broth [2016: 4927] show). The changes in the content of the turn, seen in Fragment 3, are accompanied here by the prosodic work that makes callers’ speech more similar to the chatbot’s than to usual human speech. Such prosodic features of the callers’ talk are clearly recipient-designed (Sacks et al., 1974: 727). The chatbot appears in these utterances as an interlocutor with specific hearing abilities.

The analyzed fragments, however, show only one part of the work done by the participants at the beginning of the human-chatbot telephone conversations. These conversations may start in “no gap, no overlap” mode, but this is not always the case. Silence after the chatbot’s first turn is quite common, as well as overlaps after this silence. Let us examine first the silences, as they constitute one the most noticeable phenomena in our data.



## 7. Silence at the Beginning

After the chatbot’s initial turn, if the caller remains silent (or produces sounds not recognized as speech by the system) for 4 seconds, the chatbot says “Please, don’t be silent,” and then waits another 1.5 seconds. The silence deserves a detailed examination because of its frequency (it is present in 26% of conversations in my dataset). As mentioned, at the end of 2018 a new version of the chatbot was introduced, the major differences being the absence of the “announcement” that was made before the conversation proper and the absence of a “beep” as a way of turn transfer. The reason behind these changes was that they were supposed to make a telephone conversation with the chatbot, particularly its beginning, more “natural.” One of the noticeable features of the old chatbot was the long silences after the announcement. The first explanation that emerges when one analyzes these older calls is that callers have some difficulty with recognizing that it is their turn to speak after the announcement because the announcement is considered as a preamble rather than a part of the conversation, and callers wait until the proper conversation starts (with the chatbot being the first speaker).<sup>8</sup> But it turned out that the change in how the conversations with the chatbot open did not change the overall picture: the silence after the chatbot’s first turn is still in place. This invites a closer examination of the silence.

Silence in general, of course, is a widespread conversational phenomenon. The most relevant feature of silence for the present study is that it can indicate, and helps in finding, troublesome aspects of the conversation (Roberts et al., 2006; Etehadieh & Rendle-Short, 2016; Schegloff, 1992; Pomerantz, 1984; Gardner, 2004). But this does not determine the form that silence takes. It seems that the silence at the beginning of conversations with the chatbot is a complex phenomenon that can have different forms. We may reveal them on the basis of what happens after the silence.

The first reason for silence may be that callers, despite the changes made to the chatbot, have difficulties with understanding that it is their turn. Consider, for example, the following fragment:

### Fragment 6

```
1  ЧТВ  здравствуйте (0.4) вас приветствует  
   СНВ  hello (0.4) I am  
2      виртуальный оператор (.) XXXXXX XXXXXXXXXXXX  
   a virtual operator (.) of XXXXXX XXXXXXXXXXXX  
3      XXXXXX XXXXXXXXXXXXXXXXXXXX XXXXXX. (1.1)  
   XXXXXX XXXXXXXXXXXXXXXXXXXX XXXXXX. (1.1)
```

---

<sup>8</sup> I have been told by the developers of the chatbot that they reached the same conclusion after analyzing the records of the calls.

4           какой у вас вопрос.  
               **what is your question.**  
 5           (1.9)  
 6 ЗВН а:лѐ (2.7) м (0.5) алѐ:::=  
       **CAL he:llo (2.7) hm (0.5) hello:::**  
 7 ЧТВ =я вас слушаю (.) говорите  
       **CHB I listen to you (.) speak**  
 8           (0.5)

After the silence (line 5), the caller uses “hello” to check the presence of an interlocutor or the functioning of the line. The pause of 2.7 seconds after the first “hello” and the prolongation of “o:::” in second “hello” show that the caller tries to “summon” the chatbot by providing it the opportunity to talk and, when this fails, by strengthening the summoning. This attempt to get a response from the chatbot is done because the caller is in the position of a listener-without-speaker: she holds that it is the other party’s turn and waits for it.

Fragment 6 suggests that at least in some cases the long silence after the chatbot’s “What is your question?” can be explained by the caller waiting for the chatbot to start talking, because silence in conversation has the property of “attributability” (Schegloff & Sacks, 1973). It might be difficult to find direct evidence of these in the data because the continuation of the chatbot’s talk after the silence can save the caller from the necessity of checking (e.g., via “hello”) whether the chatbot is there and why it is silent. In other cases, we have firmer grounds for analyzing the work that is done by the caller when they do not talk after the chatbot’s first turn. In the following example the caller runs into problems with inquiry formulation:

### Fragment 7

1 ЧТВ здравствуйте (0.4) вас приветствует  
       **CHB hello (0.4) I am**  
 2 виртуальный оператор (.) XXXXXX XXXXXXXXXXXX  
       **a virtual operator (.) of XXXXXX XXXXXXXXXXXX**  
 3 XXXXXX XXXXXXXXXXXXXXXXXXXX XXXXXX. (1.1)  
       **XXXXXXXXXXXXXXXXXXXXXXXX XXXXXX. (1.1)**  
 4 какой у вас вопрос.  
       **what is your question.**  
 5           (2.4)  
 6 ЗВН .hh (1.0) hh  
       **CAL .hh (1.0) hh**  
 7 ЧТВ пожа [луйста не молчите  
       **CHB ple[ase don't be silent**

8 ЗВН [( )]  
**CAL [( )]**  
9 (0.6)  
10 ЗВН .hh мне нужно соединиться с XXXXXXXXXXXX эмфэце  
**CAL .hh I need to be connected to XXXXXXXXXXXX MFC**  
11 .hh для- э узнать паспорт готов или нет  
**.hh for- ehm to find out is my passport ready or not**  
12 (4.4)  
((MFC stands for Multifunctional Center))

Evidence of the difficulty that the caller faces starts to emerge in line 6. The hearable inhale followed by an exhale after a 1-second pause shows that the caller is listening but does not formulate her inquiry. The reason becomes clearer in lines 10 and 11 when she struggles with formulating her inquiry: the caller inhales hearably in the transition-relevance place, then starts a new clause but cuts off her talk right after the first word and, after a short hesitation, produces a new clause. This struggle suggests that the caller has a problem with her inquiry, which may begin to emerge with her initial hesitation. If this is so, we can suppose that there are other cases when difficulties with inquiry formulation are the reason for the silence after the chatbot's first turn.

Another reason for callers' initial silence may be the anticipation of the chatbot's "failure", as demonstrated in the following fragment.

#### Fragment 8

1 ЧТВ здравствуйте (0.4) вас приветствует  
**СНВ hello (0.4) I am**  
2 виртуальный оператор (.) XXXXXX XXXXXXXXXXXX  
**a virtual operator (.) of XXXXXX XXXXXXXXXXXX**  
3 XXXXXX XXXXXXXXXXXXXXXXXXXX XXXXXX. (1.1)  
**XXXXXXXXXXXXXXXXXXXXXXXX XXXXXX. (1.1)**  
4 какой у вас вопрос. (5.0)  
**what is your question. (5.0)**  
5 пожалуйста не молчите (2.3) говорите  
**please don't be silent (2.3) speak**  
6 (2.3)  
7 ЗВН °блин чет я я (через) (0.6) чо буду с этой  
**CAL °jeez what I I (through) (0.6) am I to talk to this**  
8 м=ж:елезой разговаривать (0.8) (акагна)-  
**hm=p:iece of iron (0.8) (howda)-**  
9 как оператору дозвониться°  
**how to get through to the operator°**  
10 (1.9)

After a long silence and two incentives from the machine, the caller produces his comment concerning the chatbot (he is addressing someone else present in the background) in a low voice, thus making evident that he is evaluating whether it is worth talking to the machine at all. He does not want to start the conversation and would like to talk to a human operator but sees no way to get through to them. Such a deadlock explains why he stays in the conversation but does not speak. This combination of skepticism about the abilities of a chatbot and seeing no way out of the conversation makes the caller extremely hesitant as his utterances in lines 7–9 show: he repeats “I,” pauses, cuts himself off, and struggles while producing the beginning of “piece.”

This situation shows once more how important the beginning of the conversation is. From the first turn of the chatbot, the caller makes particular inferences, however vague, about what can and cannot be expected from it. Pitsch (2016) refers to the same feature when observing human-robot interaction in a museum: “[V]isitors build hypotheses about relevant subsequent actions and the robot’s interactional capabilities based on the communicational resources used by the robot in the opening phase” (p. 589).

We saw that silence at the beginning of the conversation with the chatbot can be produced in different ways. To account for these, we have to analyze what goes on right after the silence and, in some cases, during it. The caller’s work is observable in the details of the sequences of talk and non-talk. The meaning of silence for the participants, both human and non-human, is determined by its position in the sequential order of conversation. Thus, we have to examine more thoroughly a widespread consequence of the initial silence in conversations with the chatbot: the overlaps.

## 8. Overlaps

In general, the overlaps in various parts of the conversations with the chatbot do not create interactional problems for the callers. But at the beginning of the conversation, all overlapping talk is problematic because overlapping is conditioned by the silence after the chatbot’s initial turn. A typical example is Fragment 9.

### Fragment 9

- 1 ЧТВ здравствуйте (0.4) вас приветствует  
СНВ hello (0.4) I am  
2 виртуальный оператор (.) XXXXXX XXXXXXXXXXXX  
a virtual operator (.) of XXXXXX XXXXXXXXXXXX

3           XXXXXX XXXXXXXXXXXXXXXX XXXXXX. (1.1)  
**XXXXXX XXXXXXXXXXXXXXXX XXXXXX. (1.1)**  
4           какой у вас вопрос. (4.9)  
**what is your question. (4.9)**  
5           по[жалуйста                   не молчите]  
**pl[ease don't be silent]**  
6   ЗВН       [° (водительское удостове]ние)° (3.6)  
**CAL       [° (driver's           licen]se)° (3.6)**

The almost simultaneous start of the caller and the chatbot is a consequence of the previous silence. Whatever the reason for the caller's silence (he may have been distracted by something, he may have been deciding what to say and how to say it, or he may have been certain that it was the chatbot's turn), the longer the silence becomes, the more obvious it is for him that something has to be said.

The condition that causes overlapping at the beginning of the conversation may be the unpredictability of the turn's start. Consider the following fragment.

#### Fragment 10

1   ЧТВ   здравствуйте (0.4) вас приветствует  
**СНВ   hello (0.4) I am**  
2           виртуальный оператор (.) XXXXXX XXXXXXXXXXXX  
**a virtual operator (.) of XXXXXX XXXXXXXXXXXX**  
3           XXXXXX XXXXXXXXXXXXXXXX XXXXXX. (1.1)  
**XXXXXX XXXXXXXXXXXXXXXX XXXXXX. (1.1)**  
4           какой у вас вопрос.  
**what is your question.**  
5           (1.6)  
6   ЗВН   .hh телефон поликлиники. (1.2)  
**CAL   .hh phone number of the clinic. (1.2)**  
7           общий телефон поликлиники. (2.4)  
**public phone number of the clinic. (2.4)**  
8           .hh (1.1) позвонить в поли[клинику.]  
**.hh (1.1) to call to the cli[nic. ]**  
9   ЧТВ   [по   дан]ному вопросу  
**СНВ   [on th]is question**  
10          вы сможете получить информацию  
**you can obtain information**

In lines 6–8, the caller formulates the answer to the question “What is your question?” three times. She stops the flow of formulations only after her turn is

overlapped by the chatbot's turn. Two features are noticeable in this string of formulations. First, there are rather long pauses after the first and second formulations—the second pause being twice as long as the first. Second, all three formulations are variants of the same inquiry. The caller is trying to find a “correct” or an “understandable” wording, taking the chatbot's silence as evidence of the problematic character of the previous wording: the second utterance clarifies which phone number she needs, and the third simplifies the inquiry. However, the first aspect—pauses between formulations—is more important for overlap. The caller cannot project at what moment the robot will enter the conversation and therefore perceives the absence of the chatbot's turn not as evidence that the chatbot is waiting for her to finish, or is processing what she has said, but as an indication of the “inadequacy” of her inquiry. In other words, for the caller the absence of the chatbot's turn at the transition-relevance place demonstrates not just an absence of actions on the chatbot's part (caused, for example, by computer error) but the chatbot's difficulties with understanding her inquiry. Not knowing exactly what the problem is with her inquiry, the caller continues the series of formulations until the overlap.

## 9. Conclusions

The analyzed data show that in their telephone interactions with the chatbot, users design their initial turns and respond to the chatbot's actions to make the ongoing interaction understandable as a conversation. The chatbot's contributions are considered to be conversational turns that are caused by, and require, the application of the formal conversational procedures “one party at a time” and “speaker change.” Of course, there are cases in my data when users refuse to talk to the chatbot and request to be forwarded to a human operator, but even then, they make their requests in an orderly fashion. Although I analyzed only the beginning of the conversations, the examined fragments contain many interactional phenomena that can be found in other parts of the conversations. These data suggest that, for users, the chatbot *is* a conversational agent, however incapable and troublesome.

The present analysis shows that interactions with chatbots satisfy the weak participation requirement and strong analyzability requirements as introduced in Section 3: the caller understands the chatbot's action as conversationally meaningful and intelligible and displays this understanding in their actions, and this is enough for the interaction to proceed. *All* actions of both parties are analyzable for the human user in the studied fragments, and although they are analyzable only for *one* co-participant, such interactions can still be considered conversations. The chatbot is a conversational agent because the human user understands its contributions as conversational actions and because the chatbot follows “one party at a time” and “speaker change” procedures as a displayed feature of its actions. That it is *only* the human co-participant who understands

every action in the conversation should not be considered as a restricting condition that prohibits the application of the term “conversations” to such interactions. On the contrary, it should be considered as an *enabling* condition that justifies this name, providing that the computer system demonstrates its orientation to the most formal properties of conversation.

This does not mean that there are no troubles in the conversations with chatbots or other conversational agents. As my data show, there are a number of problems when interacting with chatbots, but these problems are conversationally manageable, that is, human users can use conversational techniques to solve them. A conversational agent can (from the user’s point of view) “interrupt,” “ignore” co-participants, “misinterpret” their actions, and so on. But all this can also be found in human-human conversations. There are no specific “conversational-agentic” problems particular to interactions with artificial partners. These are the usual, normal conversational problems. The fact that, when interacting with chatbots and other conversational agents, humans face these problems much more often, and they are more severe, does not make such interactions “non-conversations,” it only makes them “difficult conversations.” Humans have to navigate such conversations (and time and again they truncate or try to avoid them), but they use *conversational* means to do so. The task of the analyst in this case is not to find out how conversational agents fail, that is, fall short of human conversational abilities, but to discover how humans make sense of interactions with such agents. To call these agents “conversational” is to draw attention to their actual place in human-computer interaction: a place of the co-participant that humans deal with conversationally.

Of course, it is possible to justify the “conversationality” of conversational agents by saying that their conversational abilities are not actually theirs but are provided by programmers who implement their understanding of how communication works into computer systems. The possible argument here is that, in fact, we are dealing with a human-human interaction mediated by complex computer technologies. The problem with this argument is that in real-world interactions with conversational agents, they are, for users, self-dependent co-participants in the ongoing communication. It is conversational agents’ abilities that are evaluated by human users and to which humans have to adapt their actions. Recipient design is done with respect to the conversational agent and not its creator(s). Of course, users (at least, the majority of them) know that chatbots and similar systems are programmed. But they do not know how they are programmed or what they can and cannot do. Users have to find this out through the actual interaction, and their findings can be very different from what developers expect from the users of such systems. To know how the agent is actually programmed helps the researcher to understand why and how the agent acts the way it does but does not help with understanding how these actions are made accountable by humans in actual conversational situations.

Thus, when using ethnomethodological conversation analysis to consider data from human-chatbot conversations, we can see that artificial conversational agents are more “agentic” than the term “voice user interface” presupposes. Although chatbots’ agency is not *ontological* in the sense that we can attribute the abilities to “think,” “remember,” “comprehend,” etcetera, to the chatbot, their agency is *interactional*. One can “exchange speech” with chatbots using the sequential organization of conversation as a tool for producing local interactional order. When facing a problem while communicating with such artificial agents, human users do not respond to the problem by, for instance, just repeating their actions, as would be the case if they considered it a computer user interface. They turn to conversational methods of repairing the troublesome interactions. In such conversations, the agency of the participants, be they artificial or natural, emerges from their oriented-to-each-other interactional contributions and not from their closeness to “fully human” capabilities.

To sum up, this paper provides evidence that we can justifiably call conversational agents “conversational.” However, I am far from being sure that the question is solved once and for all. We need to conduct further studies of other interactional phenomena that can be found in various places in the conversations with such artificial interlocutors (for example, in the closing parts). And there is an obvious need for more detailed analysis of naturalistic data concerning interactions with conversational agents in the real world. But I am sure that, as conversational agents are here to stay, it is more productive to study how humans make sense of them instead of focusing on how these agents differ from humans.

## References

- Adiwardana, D., & Luong, Th. (2020, January 28). *Towards a conversational agent that can chat about . . . anything*.  
<https://ai.googleblog.com/2020/01/towards-conversational-agent-that-can.html>
- Arend, B., Sunnen, P., & Caire, P. (2017). Investigating breakdowns in human chatbot interaction: A conversation analysis guided single case study of a human-chatbot communication in a museum environment. *International Journal of Mechanical, Aerospace, Industrial, Mechatronic and Manufacturing Engineering*, 11(5), 839–845.  
<https://doi.org/10.5281/zenodo.1130169>
- Bennett, J. (2010). *Vibrant matter: A political ecology of things*. Duke University Press.



- Bolden, G. B., & Guimaraes, E. (2012). Grammatical flexibility as a resource in explicating referents. *Research on Language and Social Interaction*, 45(2), 156–174. <https://doi.org/10.1080/08351813.2012.673861>
- Button, G., & Sharrock, W. W. (1995). On simulacrum of conversation: Toward a clarification of the relevance of conversation analysis for human-computer interaction. In P. J. Thomas (Ed.), *The social and interactional dimensions of human-computer interfaces* (pp. 107–125). Cambridge University Press.
- Clift, R. (2016). *Conversation analysis*. Cambridge University Press. <https://doi.org/10.1017/9781139022767>
- Cromdal, J., Persson-Thunqvist, D., & Osvaldsson, K. (2012). “SOS 112 what has occurred?”: Managing openings in children’s emergency calls. *Discourse, Context & Media*, 1(4), 183–202. <https://doi.org/10.1016/j.dcm.2012.10.002>
- Danby, S., Baker, C. D., & Emmison, M. (2005). Four observations on openings in calls to kids help line. In C. D. Baker, M. Emmison, & A. Firth (Eds.), *Calling for help: Language and social interaction in telephone helplines* (pp. 133–151). John Benjamins. <https://doi.org/10.1075/pbns.143.10dan>
- Etehadieh, E., & Rendle-Short, J. (2016). Intersubjectivity or preference: Interpreting student pauses in supervisory meetings. *Australian Journal of Linguistics*, 36(2), 172–188. <https://doi.org/10.1080/07268602.2015.1121529>
- Francis, D., & Hester, S. (2004). *An invitation to ethnomethodology: Language, society, and social interaction*. SAGE. <https://doi.org/10.4135/9781849208567>
- Garcia, A. C. (2013). *An introduction to interaction: Understanding talk in formal and informal settings*. Bloomsbury. <https://doi.org/10.5040/9781350284821>
- Gardner, R. (2004). On delaying the answer: Question sequences extended after the question. In R. Gardner & J. Wagner (Eds.), *Second language conversations* (pp. 246–266). Continuum. <https://doi.org/10.5040/9781474212335.0016>
- Garfinkel, H. (2002). *Ethnomethodology’s program: Working out Durkheim’s aphorism*. Rowman & Littlefield.
- Gehle, R., Pitsch, K., Dankert, T., & Wrede, S. (2017). How to open an interaction between chatbot and museum visitor? Strategies to establish a focused encounter in HRI. In *HRI’17: Proceedings of the 2017 ACM/IEEE*

*International Conference on Human-Chatbot Interaction (Vienna, Austria, March 6–9, 2017)* (pp. 187–195). ACM.  
<https://doi.org/10.1145/2909824.3020219>

Gibbs, J. L., Kirkwood, G. L., Fang, C., & Wilkenfeld, J. N. (2021). Negotiating agency and control: Theorizing human-machine communication from a structural perspective. *Human-Machine Communication*, 2, 153–171.  
<https://doi.org/10.30658/hmc.2.8>

Goodwin, Ch., & Heritage J. (1990). Conversation analysis. *Annual Review of Anthropology*, 19, 283–307.  
<https://doi.org/10.1146/annurev.an.19.100190.001435>

Harbers, H. (Ed.). (2005). *Inside the politics of technology: Agency and normativity in the co-production of technology and society*. Amsterdam University Press. <http://library.oapen.org/handle/20.500.12657/35139>

Hoey, E. M. (2020). *When conversation lapses: The public accountability of silent copresence*. Oxford University Press.

Hopper, R. (1992). *Telephone conversation*. Indiana University Press.

Hopper, R., & Chen, Ch.-H. (1996). Languages, cultures, relationships: Telephone openings in Taiwan. *Research on Language and Social Interaction*, 29(4), 291–313. [https://doi.org/10.1207/s15327973rlsi2904\\_1](https://doi.org/10.1207/s15327973rlsi2904_1)

Houtkoop-Steenstra, H. (1991). Opening sequences in Dutch telephone conversations. In D. Boden, & D. H. Zimmerman (Eds.), *Talk and social structure* (pp. 232–250). University of California Press.

Hutchby, I., & Wooffitt, R. (2002). *Conversation analysis: Principles, practices and applications*. Polity Press.

Koshik, I. (2005). *Beyond rhetorical questions: Assertive questions in everyday interaction*. John Benjamins. <https://doi.org/10.1075/sidag.16>

Krummheuer, A. L. (2015). Technical agency in practice: The enactment of artefacts as conversation partners. *PsychNology Journal*, 13(2–3), 179–202. [http://www.psychology.org/File/PNJ13%282-3%29/PSYCHNOLOGY\\_JOURNAL\\_13\\_2\\_KRUMMHEUER.pdf](http://www.psychology.org/File/PNJ13%282-3%29/PSYCHNOLOGY_JOURNAL_13_2_KRUMMHEUER.pdf)

Krummheuer, A.L. (2016). Who am I? What are you? Identity construction in encounters between a teleoperated robot and people with acquired brain injury. In A. Agah, J.-J. Cabibihan, A. M. Howard, M. A. Salichs, & H. He (Eds.), *Social chatbotics: Proceedings of the 8th International Conference ICSR 2016 (Kansas City, MO, USA, November 1–3, 2016)* (pp. 880–889). Springer. [https://doi.org/10.1007/978-3-319-47437-3\\_86](https://doi.org/10.1007/978-3-319-47437-3_86)

- Latour, B. (2005). *Reassembling the social: An introduction to actor-network-theory*. Oxford University Press.
- Leydon, G. M., Ekberg, K., & Drew, P. (2013). "How can I help?": Nurse call openings on a cancer helpline and implications for call progressivity. *Patient Education and Counseling*, 92(1), 23–30. <https://doi.org/10.1016/j.pec.2013.02.007>
- Lindström, A. (1996). Identification and recognition in Swedish telephone conversation openings. *Language in Society*, 23(2), 231–252. <https://doi.org/10.1017/S004740450001784X>
- Liberman, K. (2013). *More studies in ethnomethodology*. State University of New York Press.
- Livingston, E. (1987). *Making sense of ethnomethodology*. Routledge & Kegan Paul.
- Luff, P. K., Gilbert, N., & Frohlich, D. (Eds.). (1990). *Computers and conversation*. Academic Press. <https://doi.org/10.1016/C2009-0-21641-2>
- Moore, R. J. (2018). A Natural Conversation Framework for conversational UX design. In R. J. Moore, M. H. Szymanski, R. Arar, & G.-J. Ren (Eds.), *Studies in conversational UX design* (pp. 181–204). Springer. [https://doi.org/10.1007/978-3-319-95579-7\\_9](https://doi.org/10.1007/978-3-319-95579-7_9)
- Moore, R. J., & Arar, R. (2019). *Conversational UX design: A practitioner's guide to the Natural Conversation Framework*. ACM. <https://doi.org/10.1145/3304087>
- Pallotti, G., & Varcasia, C. (2008). Service telephone call openings: A comparative study on five European languages. *Journal of Intercultural Communication*, 17. <http://www.immi.se/intercultural/nr17/pallotti.htm>
- Park, Y.-Y. (2002). Recognition and identification in Japanese and Korean telephone conversation openings. In: K. K. Luke & T.-S. Pavlidou (Eds.), *Telephone calls: Unity and diversity in conversational structure across languages and cultures* (pp. 25–47). John Benjamins. <https://doi.org/10.1075/pbns.101.06par>
- Pelikan, M. H. R., & Broth, M. (2016). Why that Nao? How humans adapt to a conventional humanoid robot in taking turns-at-talk. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 4921–4932). ACM. <https://doi.org/10.1145/2858036.2858478>
- Pitsch, K. (2015). Ko-Konstruktion in der Mensch-Roboter-Interaktion: Kontingenz, Erwartungen und Routinen in der Eröffnung. In E. Gülich, U.,

Krafft, & U. Dausendschön-Gay (Eds.), *Ko-Konstruktion in der Interaktion: Die gemeinsame Arbeit an Äußerungen und anderen sozialen Ereignissen* (pp. 229–258). Transcript. <https://doi.org/10.1515/9783839432952-013>

Pitsch, K. (2016). Limits and opportunities for mathematizing communicational conduct for social robotics in the real world? Toward enabling a robot to make use of the human's competences. *AI & Society*, 31(4), 587–593. <https://doi.org/10.1007/s00146-015-0629-0>

Pitsch, K., Kuzuoka, H., Suzuki, Y., Süssenbach, L., Luff, P., & Heath, Ch. (2009). "The first five seconds": Contingent stepwise entry into an interaction as a means to secure sustained engagement in HRI. In *The 18th IEEE International Symposium on Chatbot and Human Interactive Communication (Toyama, Japan, September 27-October 2, 2009)* (pp. 985–991). IEEE. <https://doi.org/10.1109/ROMAN.2009.5326167>

Pomerantz, A. (1984). Pursuing a response. In J. M. Atkinson & J. Heritage (Eds.), *Structures of social action* (pp. 152–163). Cambridge University Press. <https://doi.org/10.1017/CBO9780511665868.011>

Porcheron, M., Fischer, J. E., Reeves, S., & Sharples, S. (2018). Voice interfaces in everyday life. In *CHI'18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (Montreal QC, Canada, April 21–26, 2018)* (pp. 640:1–640:12). ACM. <https://doi.org/10.1145/3173574.3174214>

Porcheron, M., Fischer, J. E., & Sharples, S. (2017). "Do animals have accents?": Talking with agents in multi-party conversation. In *CSCW'17: Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (Portland, Oregon, USA, February 25-March 1, 2017)* (pp. 207–219). ACM. <https://doi.org/10.1145/2998181.2998298>

Reeves, S. (2017). Some conversational challenges of talking with machines. In *Talking with Conversational Agents in Collaborative Action: Workshop at the 20th ACM Conference on Computer-Supported Cooperative Work and Social Computing (CSCW'17) (Portland, Oregon, USA, February 25–March 1, 2017)*. <https://doi.org/10.1145/3022198.3022666>

Roberts, F., Francis, A., & Morgan, M. (2006). The interaction of inter-turn silence with prosodic cues in listener perceptions of "trouble" in conversation. *Speech Communication*, 48(9), 1079–1093. <https://doi.org/10.1016/j.specom.2006.02.001>

Sacks, H. (1992). *Lectures on conversation*. Blackwell. <https://doi.org/10.1002/9781444328301>

- Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4), 696–735. <https://doi.org/10.1353/lan.1974.0010>
- Schegloff, E. A. (1968). Sequencing in conversational openings. *American Anthropologist*, 70(6), 1075–1095. <https://doi.org/10.1525/aa.1968.70.6.02a00030>
- Schegloff, E. A. (1979). Identification and recognition in telephone conversation openings. In G. Psathas (Ed.), *Everyday language: Studies in ethnomethodology*. Irvington, pp. 23–78.
- Schegloff, E. A. (1986). The routine as achievement. *Human Studies*, 9(2–3), 111–151. <https://doi.org/10.1007/BF00148124>
- Schegloff, E. A. (1992). Repair after next turn: The last structurally provided defense of intersubjectivity in conversation. *American Journal of Sociology*, 97(5), 1295–1345. <https://doi.org/10.1086/229903>
- Schegloff, E. A. (2002a). Opening sequencing. In J. E. Katz & M. Aakhus (Eds.), *Perpetual contact: Mobile communication, private talk, public performance* (pp. 326–385). Cambridge University Press. <https://doi.org/10.1017/CBO9780511489471.026>
- Schegloff, E. A. (2002b). Reflections on research on telephone conversation: Issues of cross-cultural scope and scholarly exchange, interactional import and consequences. In K. K. Luke & T.-S. Pavlidou (Eds.), *Telephone calls: Unity and diversity in conversational structure across languages and cultures* (pp. 249–281). John Benjamins. <https://doi.org/10.1075/pbns.101.16sch>
- Schegloff, E. A. (2007). *Sequence organization in interaction: A primer in conversation analysis*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511791208>
- Schegloff, E. A., & Sacks, H. (1973). Opening up closings. *Semiotica*, 8(4), 289–327. <https://doi.org/10.1515/semi.1973.8.4.289>
- Sidnell, J. (2010). *Conversation analysis: An introduction*. Wiley-Blackwell.
- Sidnell, J., & Stivers, T. (Eds.). (2012). *The handbook of conversation analysis*. Wiley-Blackwell. <https://doi.org/10.1002/9781118325001>
- Sifianou, M. (1989). On the telephone again! Differences in telephone behaviour: England versus Greece. *Language in Society*, 18(4), 527–544. <https://doi.org/10.1017/S0047404500013890>

- Slack, J. D., & Wise, J. M. (2005). *Culture and technology: A primer*. Peter Lang.
- Suchman, L. (1987). *Plans and situated actions: The problem of human-machine communication*. Cambridge University Press.
- Suchman, L. (2007). *Human-machine reconfigurations: Plans and situated actions*. Cambridge University Press.  
<https://doi.org/10.1017/CBO9780511808418>
- Taleghani-Nikazm, C. (2002). Telephone conversation openings in Persian. In K. K. Luke & T.-S. Pavlidou (Eds.), *Telephone calls: Unity and diversity in conversational structure across languages and cultures* (pp. 87–109). John Benjamins. <https://doi.org/10.1075/pbns.101.08tal>
- ten Have, P. (2002). Comparing telephone call openings: Theoretical and methodological reflections. In K. K. Luke & T.-S. Pavlidou (Eds.), *Telephone calls: Unity and diversity in conversational structure across languages and cultures* (pp. 234–248). John Benjamins.  
<https://doi.org/10.1075/pbns.101.15ten>
- ten Have, P. (2007). *Doing conversation analysis: A practical guide*. SAGE.
- Thomas, P. J. (Ed.). (1995). *The social and interactional dimensions of human-computer interfaces*. Cambridge University Press.
- Verbeek P.-P. (2005) *What things do: Philosophical reflections on technology, agency, and design*. Pennsylvania State University Press.  
<https://doi.org/10.1515/9780271033228>
- Vinkhuyzen, E., Whalen, M., & Szymanski, M. (2006). Security, efficiency, and customer service in calls to a financial services organization. *Revue Française de Linguistique Appliquée*, 11(2), 53–68.  
<https://www.cairn.info/revue-francaise-de-linguistique-appliquee-2006-2-page-53.htm>
- Wakin, M. A., & Zimmerman, D. H. (1999). Reduction and specialization in emergency and directory assistance calls. *Research on Language and Social Interaction*, 32(4), 409–437.  
[https://doi.org/10.1207/S15327973rls3204\\_4](https://doi.org/10.1207/S15327973rls3204_4)
- Whalen, M., & Zimmerman, D. H. (1987). Sequential and institutional contexts in calls for help. *Social Psychology Quarterly*, 50(2), 172–185.  
<https://doi.org/10.2307/2786750>

Wooffitt, R., Fraser, N. M., Gilbert, N., & McGlashan, S. (1997). *Humans, computers and wizards: Human (simulated) computer interaction*. Routledge.

Zimmerman, D. H. (1992). Achieving context: Openings in emergency calls. In G. Watson, & R. M. Seiler (Eds.), *Text in context: Contributions to ethnomethodology* (pp. 35–51). SAGE.