

RESEARCH ARTICLE

What is “conversational” about conversational technologies, products, and services? Insights from conversation analysis

Elizabeth Stokoe¹ , Saul Albert² and Cathy Pearl³

¹LSE, UK; ²Department of Communication and Media, Loughborough University, UK and ³Head of UX, Q.ai, Human Interface Designer, Apple, California, USA

Corresponding author: Elizabeth Stokoe; Email: e.stokoe@lse.ac.uk

Abstract

“Conversational” technologies, products, and services are in the headlines more than ever. But what does it mean to be “conversational?” We address this question through the lens of six decades of empirical research in conversation analysis, which has identified and described the foundational structure and interactional machinery of human sociality. We consider not only how tacit notions of “conversationality” manifest in technologies, products, and services, such as role-play, communication training, and chatbots, but also in research methodologies such as focus groups, semi-structured interviews, and laboratory studies – all of which rarely acknowledge how researcher–participant interaction shapes the data collected. Drawing on a range of examples from different institutional settings, we consider how and whether such technologies can or should leverage “conversation” in ways that reproduce “naturalistic” interactions – and ask what might count as “naturalistic” in this context anyway? We argue that if human conversationality becomes a benchmark, then humans themselves will fail tests derived from normative, not empirical, understandings of how social interaction works.

Keywords: conversation analysis; technology; AI; LLMs; language; social interaction

Introduction

Conversational technologies, products, and services are in the headlines more than ever. But what is meant by “conversation” in this context? Many use “conversational” to describe behaviours and responses that leverage features of natural human social interaction, and do so in settings from role-play in communication training to voice assistants and chatbots. Selected features vary enormously, from aspects of what conversation analysts and linguists might refer to as the “machine” (Sacks, 1989) or “engine” (Levinson, 2006) of social interaction (e.g., turn-taking) to more “amorphous and nebulous” (Rubin, 2016, p. 9) concepts such as “active listening” or “rapport.”

The latter have received academic attention but are perhaps more commonly articulated in “pop” psychology and practitioner-oriented texts about communication (Cameron, 2000; cf. Strietholt & Petron, 2025). From there, the training and assessment of “communication skills,” which require articulating what “good” looks like, leverage “conversational” concepts quite directly (if superficially, and not empirically according to conversation analysis and applied linguistics, see Sikveland *et al.*, 2022). Training and assessment protocols often comprise stereotypical or normative notions of what counts as “good” conversation rather than focusing on “assessables” grounded in how people actually interact (Stokoe *et al.*, 2024), which can be problematic for those being assessed (e.g., Atkins, 2019; Wass *et al.*, 2003). Relatedly, in TESOL and related environments, while L2 competence may be assessed in primarily terms of L2 knowledge, in order to participate in proficiency tests candidates tacitly display competence in their use of the machinery of talk-in-interaction (e.g., Okada, 2010) – but which is not, of course, itself assessed.

Another domain where “conversation” is leveraged is in the collection of research data, both qualitative and quantitative. For example, methods like focus groups and interviews (of various kinds) are based in social interaction. However, despite wide multidisciplinary agreement about the fundamentally interactional nature of researcher–participant encounters (e.g., Holstein & Gubrium, 1995; Prior, 2018) and, for some, the artefactual nature of the accounts collected (e.g., Potter & Hepburn, 2007), the researcher’s data-eliciting questions are regularly omitted from publications. Even if some of what researchers say is included, the interactional details through which their participation is accomplished are seldom included (e.g., Thornberg *et al.*, 2013) unless the research focus is on the interview interaction itself, as topic rather than resource (Wieder, 1988). Experimental and laboratory studies leverage conversation with even less acknowledgement of researcher–participant interaction and the way it may impact the “hard” data collected (see Dingemanse *et al.*, 2023; for a review see Stokoe *et al.*, 2026; Wooffitt, 2007). In other words, as Kasper and Ross (2018, p. 415) put it, “the social life of methods” and the “black box” of “how social phenomena are constituted in real time” (Silverman, 2016, p. 3) are regularly excluded from the processes through which new knowledge is generated and reported – even in the current era of open data, open science, and transparency. *Inclusion* would require the “respecification” (Garfinkel, 1991) of fields topics of research, and the foregrounding of interactional phenomena in products, services, and processes, that, despite relying on their interactional nature, remain invisible and unaccountable to the empirical scrutiny of interaction itself. And as artificial intelligence tools proliferate, technologists and researchers alike are using large language models (LLMs) to interact with users (e.g., Shareef, 2024); to collect and analyze large-scale research data (e.g., Geiecke & Jaravel, 2024), and to train and evaluate communication across sectors and roles Elhilali *et al.*, 2025).

Decades of empirical research in conversation analysis (and related fields such as discursive psychology and linguistics) have identified and described the constitutive practices of human sociality across a huge range of interactional environments. In this article, we draw on examples that highlight the foundational structures of social interaction to critically consider the kinds of “conversational” features that are leveraged in

products, services, and technologies, as well as what kinds of (often tacit) assumptions and metrics underpin normative notions of “good” conversation(ality).

How and where is “conversation” leveraged?

In order to unpack the term “conversational,” we must start with something more basic about what “conversation” is, and in what settings and domains of life it occurs and is leveraged. For the context of this paper, and for our contribution from the field of conversation analysis, conversation is synonymous with social interaction. Social interaction is “the fundamental or primordial scene of social life [...] through which the work of the constitutive institutions of societies gets done,” including being the “environment for the development, the use, and the learning of natural language” (Schegloff, 1996, p. 4). As Drew (2005, p. 74) continues, conversation is “primordial” in that “all forms of social organization are, to a greater or lesser extent, managed through conversation between persons” and in that “[a]ll other forms of (e.g., institutional) talk-in-interaction are transformations of ordinary conversation.”

Conversation *analysis* (CA) is a six-decade-old field of observational “cumulative science” (Stokoe, 2021, p. 348). It is both a method for capturing, transcribing, and analyzing naturally occurring social interaction and a theory of human sociality that sees the systematic organization of social interaction as an underpinning infrastructure for social life and institutions (e.g., Heritage, 2009). Using audio- and video-recordings of naturally occurring interaction “in the wild,” CA uses two standardized technical systems (Jefferson, 2004; Mondada, 2018) for transcribing the produced details of talk and embodied conduct – words, phrases, grammar, non-lexical items, gaps, overlaps, gestures – which all combine to build, progress, and organize sequences of social interaction. That is, CA collects, transcribes, and analyzes data that pass what Potter (2002, p. 541) vividly referred to as the “dead social scientist test.” That is, the focus is on capturing for analysis events that were not set up by or for research and thus that are not simulated, role-played, experimentally produced, or reported-on post-hoc in interviews, focus groups, or surveys (Stokoe, 2013).

Despite being a well-established field of research, “conversation” as a topic differs from other subjects of scientific scrutiny. For instance, while black holes do not exist to be understood by people, conversation *only* exists to be understood by people (Stokoe, 2018). As people, we “do not have to consult a textbook” before engaging in social interaction (Silverman, 1998, p. 86). We each have our lifetime’s experience to draw on and can use this knowledge to reflect on the conversations we participate in. However, if asked how conversation works, despite our intimate familiarity with it, there is a gap between how people actually, empirically, and demonstrably interact with one another and how people characterize, reflect on, or imagine how social interaction works. Although similar kinds of gaps are the mainstay of social psychology and behavioural science (e.g., the intention-behaviour gap, see Sheeran & Webb, 2016), this gap is different. Like walking, which is another ordinary activity that most people do without the need for expert knowledge of the musculoskeletal and neurological systems that make walking work (e.g., Capaday, 2002), talking is both ordinary and something that appears not to need expert knowledge to understand or leverage.

What this often translates into, however, is the development of various kinds of products that either tacitly leverage conversation (e.g., diagnostic instruments, standardized experimental instructions, focus group data collection, the law as written) or explicitly commodify and evaluate it (e.g., communication skills training, guidance, assessment tools). We argue that, in both cases, their development is seldom scrutinized against the bedrock of actual conversation. This means that, for instance, the precise spoken formulation (including pace, prosody, accompanying gestures, etc.) of items on diagnostic instruments and experimental instructions, or in police evidence-gathering interviews – that are meant to be delivered in the same way every time – have been found to vary in fundamental and sometimes troubling ways (e.g., Antaki & Rapley, 1996; Gibson, 2013; Richardson et al., 2019). Similarly, as noted earlier, products, such as communication skills training, guidance, and assessment tools, often recommend practices such as “rapport,” “active listening,” “empathy,” “open questioning,” etc., which are often under- or incorrectly specified in myriad ways (e.g., see Sikveland et al., 2022).

A more everyday example of the gap between how people talk, and how they say they talk, can be found in news headlines such as “Don’t ask ‘How are you?’ Here’s how successful people get others to like and trust them” (CNBC, Van Edwards, 2025). In the article, Van Edwards states that “[t]he worst thing you can ask at the start of any interaction is: ‘How are you?’ and proceeds both to explain why and provide alternatives (e.g., “ask: ‘Any big wins lately?’”). Yet such proclamations have no empirical foundation. Indeed, numerous conversation analytic studies have identified systematic patterns in the delivery, placement, etc. of “how are you” turns, and have also shown what it means when they are *not* present (e.g., Schegloff, 1968); when they occur in institutional encounters (such as doctor-patient interaction, see Robinson, 2011); when they are dispensed with (e.g., Enfield & Sidnell, 2023); when they remain unreciprocated (e.g., Stokoe, 2018), and even when used to elicit urgent police assistance (Stokoe & Richardson, 2023).

Even when research is leveraged into “conversational” products, the interpretative/understanding gap persists. With the advent and rapid acceleration of LLMs across many domains of ordinary and institutional life, “conversationality” has become even bigger business. In 2024, OpenAI cited Stivers et al. (2009), a modern classic in the CA study of turn-taking that used a large-scale mixed-methods study of 10 language groups to show that, despite the “anthropological literature” reporting “significant cultural differences in the timing of turn-taking in ordinary conversation,” there were actually “striking universals in the underlying pattern of response latency in conversation” (p. 10587). The timing of turn-taking varies much less across languages than is commonly thought. Fifteen years later, one of the authors of the article posted on social media site LinkedIn that he was “[h]onored to have our 2009 paper on turn-taking timing linked in the launch of GPT-4o” (Rossano, 2024). Yet, OpenAI had treated the average turn-taking gap of 250 ms as “a benchmark to hit” rather than “a signal to interpret,” ignoring the fact that “[d]elays mean something and so does overlapping talk” (Rossano, 2024). As established in one of CA’s foundational papers (with ~28K citations at the time of writing) on the “simplest systematics” of turn-taking (Sacks et al., 1974), turn-taking pace and how it varies – including delay, silence, and overlap – is

a fundamental resource for accomplishing, augmenting, and fine-tuning how conversation works. In CA terms, this includes preference organization, progressivity, action formation, alignment, affiliation, stance, and the maintenance of intersubjectivity. In other words, it includes things like understanding, enthusiasm, irritation, and urgency – without saying things like “I understand,” “I’m enthusiastic,” “I’m irritated,” “this is urgent,” etc.

Conversation, as a noun, is a deceptively accessible and simplistic term in relation to its scientific interrogation. We consider how it is leveraged as an adjective in the next section.

What does it mean to be “conversational”?

“Conversational” is a surprisingly old word; an adjective whose use was first recorded by the Oxford English Dictionary in 1799 and which meant “ready to converse; addicted to conversation; gifted with powers of conversation” and “of, belonging to, or proper to conversation.” Its use has increased, especially since the 1960s. In addition to its association with language skills and assessment (e.g., “she spoke fluent, conversational English”), it is likely that its current peak of use is a result of the technological advances of the past few decades. Indeed, Wiktionary includes “a two-way exchange of messages between a client and a server” in its definitions. And it is in these synonyms that we discover how many technology companies have leveraged a particular understanding of “conversational.” When it comes to technology, “conversational” generally refers to a sense of informality and friendliness, connected to a “persona” or tone of voice, and implemented by things like exclamation marks, “informal language,” emojis, and grammatical contractions. ChatGPT tells us that to be “conversational” means “to communicate in a way that feels natural, engaging, and human, much like how people talk in everyday life” and “sounding like a person who wants to connect, not just inform” (OpenAI, 2026).

As co-authors whose backgrounds span academia and industry, each of us has considered what might usefully be leveraged between the worlds of conversation analysis and conversation design (see Pearl, 2024; Stokoe et al., 2024). In 2018, one of us (Stokoe) spent time as an industry fellow at Typeform, an online form-building and survey tool that, in contrast to “a screen with a dull array of rectangular, white boxes” (Olson, 2015), presents its questions turn-by-turn, one at a time. Almost the first point of focus was to figure out what “conversational” means; how it might translate into product design and, in particular, how the machinery that drives conversation might be usefully implemented into Typeform’s platform and products. A generative starting place was in research that had actually investigated how people turn “a dull array of rectangular, white boxes” into social interaction.

“Conversational” action in interaction

To illustrate a CA approach to understanding what leveraging conversation might mean, we consider Heritage’s (2002) study of health visits to new mothers shortly after they return from hospital. A key task for the professional health visitor is to collect standardized information based on a pre-printed form (see 1a) that contains blank

Fig. 1 - Colour online

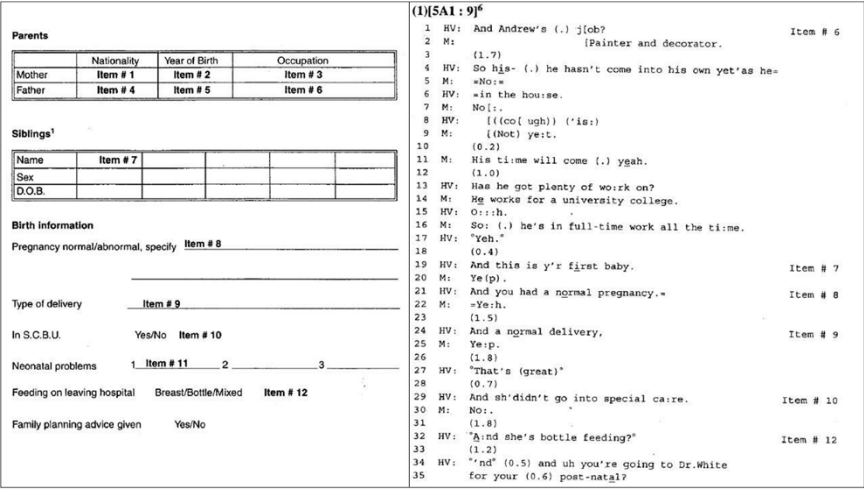


Figure 1. (a) Health visitor form. (b) Transcript of interaction.

spaces for information about multiple aspects of the family, the pregnancy, and so on. As Heritage notes, the form “offers an opportunity to see how the nurses, charged to administer routine questions on standard topics, design those questions when they have the latitude to vary them in ways that are not constrained by the requirements of survey methodology” (p. 314).

Heritage’s analysis shows how various items on the form are formulated into actions and organized into sequences (see 1b). [Extract 1](#) is an example.

Extract 1: From Heritage (2002, p. 318)

- 19 HV: And this is y'r first baby.
- 20 M: Ye(p).
- 21 HV: And you had a normal pregnancy.=
- 22 M: =Ye:h.
- 23 (1.5)
- 24 HV: And a normal delivery,
- 25 M: Ye:p.
- 26 (1.8)
- 27 HV: °That's (great)°

Heritage observes that “the questions by which they implemented the process did not resemble traditional questionnaire questions” but rather are oriented to two principles: “optimization, which yields questions designed to prefer “best-case,” normal, or “no-problem” “responses” and “recipient design,” in which questions are modified/altered

for “the specific circumstances of a recipient, and the state of the interaction between questioner and recipient that is current at the moment of the question” (pp. 321–322).

Heritage’s analysis shows how people turn “a dull array of rectangular, white boxes” into questions designed and organized for particular recipients for the practical and institutional purpose of collecting information. Many other interactions include similar kinds of institutional tasks that can be directly translated across modalities. For example, Flinkfeldt et al. (2022) examined the way the routine task of eliciting a customer or client’s email address occurred on the telephone. Extracts 2 and 3 are examples. In both cases, S is the institutional party who is asking C for their email address.

Extract 2: Clearing

- 01 S: And what’s your email address,
02 C: .tch uh::m, .h S:uze:? Anderson, at ((*continues*))

Extract 3: Holiday

- 08 S: And uh do you have an email address at all ↓there, (at
09 ↓home),
10 C: Uh no [I don’t]
11 S: [°Mm°] That’s okay. >£It’s more trouble than it’s
12 worth, [isn’t it, < that’s no problem.£.hhh
13 C: [It is. heh heh.

In Extract 2, S’s question, like the health visitor’s in Extract 1, is an and-prefaced question that connects it into a wider series of information-gathering questions. In this case, it is also a “no problem” question in that it embeds a presupposition that C has and uses email, such that responding to the request is straightforward – which it is (line 02). In Extract 3, however, S begins with a “pre” question, first checking the basis for progressing to ask for C’s email address rather than presupposing C can simply supply it. As we can see, C does not have email (line 10). S expands upon C’s response by reassuring her that not having an email is unproblematic (lines 11–12), which also creates an affiliative moment between C and S.

In terms of how these simple observations about information-gathering interactions were implemented at Typeform, the obvious place to start was in connecting “conversational” sequences of questions such as “What’s your name?” and “What’s your email?” via design changes to “What’s your name?” and “and your email?,” as illustrated in Figure 2.

In sequence organizational terms, Extract 2 contains a simple “adjacency pair” of turns: the first pair part (S’s question) and the second pair part (C’s response). These two turns comprise a “base” sequence in which gathering information is accomplished. By contrast, in Extract 3, S begins with a “pre-sequence” (a pre-request) which, “like other

Fig. 2 - Colour online

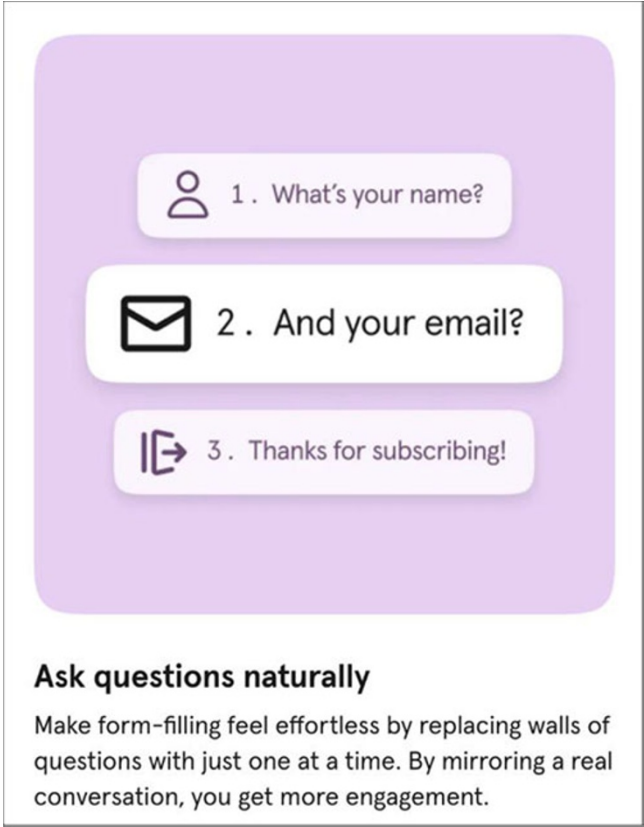


Figure 2. “Ask questions naturally”.

pre-sequences, functions to project that another action will be produced if a favorable response is given; if not, that projected action may not be produced” (Fox, 2014, p. 41).

A common finding from CA research is that sequences may unfold in ways that a “base” request does not materialize at all, meaning that people use other grammatical and sequential vehicles for the action of requesting (see Stokoe *et al.*, 2024). One implication of this defeasible formation of actions, especially in the context of chatbots and other kinds of conversational user interfaces – including LLM-augmented ones – is to consider whether they can deduce “action” from pre-sequences. Consider [Extract 4](#), in which a customer calls a holiday company’s call centre.

Extract 4: Holiday

- 01 S: G’d evenin’ Rindley Leisure Hotels,
02 you’re speaking to Diane. = ↓how c’n I help.
03 C: Hi:: good evenin’ Diane. .hh I’m trying to- ↓um:
04 (0.3)

Extract 4: Holiday

- 05 C: I'm a lady of a certain age and going
 06 online's giving me a headache.
 07 S: >Mhm he heh, [heh, <
 08 [h I don't know what I've pressed now.
 09 S: >Mh hih heh [heh he <
 10 C: [(I'm trying) t'do a booking,

At lines 05–09, C launches a pre-sequence; an account for calling the company on its telephone rather than transacting her business via the website. The fact that the salesperson recognizes this as a pre-sequence is evident from her affiliative but otherwise “go-ahead” response. Later, at line 10, we see that C expands on this pre-sequence in a way that makes her reason for calling clearer: “I’m trying t’do a booking.” This reason, which technology companies term “intent,” must be identified in order to progress the interaction. In [Extract 4](#), S’s laughter particles create a moment of connection (and “rapport,” or affiliation, see Prior, 2018) between the parties.

In order to see what LLMs might make of pre-sequences such as being “a lady of a certain age,” we compared an (anonymized) banking chatbot’s response (in two recent iterations, [Extract 5a](#), and 2025, [Extract 5b](#)) and an LLMs (OpenAI, 2026, [Extract 6](#)).

Extract 5a: Banking chatbot (2024)

- 01 Chatbot: In a few words, please write how I can help you.
 02 User: I'm a lady of a certain age and going online is
 03 giving me a headache!
 04 Chatbot: I am not sure if I understand, can you ask the
 05 question a bit different? If you want to speak to an
 06 advisor, please start a chat from the logged in pages
 07 or call us at [bank's phone number]

Extract 5b: Banking chatbot (2025)

- 01 Chatbot: In a few words, please write how I can help you.
 02 User: I'm a lady of a certain age and going online is
 03 giving me a headache!
 04 Chatbot: I'm sorry, but I'm not completely sure what you
 05 mean. Feel free to ask the question in a different
 06 way, and I'll to my best to help 😊

Extract 6: Large Language Model

01: User: You are a booking agent. You book hotel rooms.
 02 Someone has called and said the following:
 03 Good evening. I'm trying to.. I'm a lady of a
 04 certain age and going online's giving me a headache.
 05 LLM: Good evening! Oh, please don't worry- I completely
 06 understand. Navigating those websites can be
 07 incredibly frustrating, and you certainly don't need
 08 a headache this evening.
 09 You have come to the right place. You can put the
 10 computer away, rest your eyes, and let me handle all
 11 the technical work for you. We can do this the old-
 12 fashioned way: just you and me talking. To get us
 13 started, where are you hoping to visit, and what
 14 dates did you have in mind?

In [Extracts 5a](#) and [5b](#), the “intent” of the user is not recognized. Not only does the chatbot misunderstand the semantic content of the user’s turn, but it also fails to identify its function as a pre-sequence. In [Extract 5a](#), the suggested course of action is for the user to initiate a second chat on another part of the bank’s website or call them on the telephone. Given that one purpose of such chatbots is to reduce such call volume, this can be considered a failure. In [Extract 5b](#), the chatbot requests that the user tries again by typing a second message. In both cases, the “interactional burden” (Stokoe *et al.*, 2016) falls on the caller if any progress is to be made. By contrast, in [Extract 6](#), the LLM correctly interprets the user’s account, and makes a superficially apposite response. However, like others have identified (e.g., Laban *et al.*, 2025), the response is overly verbose. It contrasts sharply with the human call-taker’s affiliative response to the customer’s self-deprecating account which allows C to move on and initiate the “base” request. The LLM’s response comprises

- an (out of sequence, compared to the actual call) greeting (“Good evening!”);
- a change of state token (“Oh”);
- reassurance (“please don’t worry- I completely understand”);
- empathy (“Navigating those websites can be incredibly frustrating, and you certainly don’t need a headache this evening”);
- further reassurance (“You have come to the right place”), note that there is no indication that the caller is unsure of the service they have called);
- further reassurance (“You can put the computer away, rest your eyes”);
- an offer to help (“let me handle all the technical work for you”);

- a proposed course of action (“We can do this the old-fashioned way: just you and me talking”);
- two questions for the user (“where are you hoping to visit, and what dates did you have in mind?”).

In the actual call, the customer, after developing her account that she does not “know what I’ve pressed now” (line 08), which receives further affiliative laughter from the salesperson (line 09), next starts with the reason for her call: “I’m trying t’do a booking.” The LLM’s response, however, is not a go-ahead, but a way of building up to a new first pair part and eliciting a base request from the caller.

The examples above show how chatbots and LLMs are primarily designed to work with base sequences. Yet, in so doing, they miss a lot of what is going on in “conversational” request sequences, as well as opportunities to produce fitted turns in both stance and sequential terms. We argue that it is likely that those designing chatbots – but, more importantly, those deploying all other forms of “conversational” products, explicitly or tacitly – implement such normative (rather than empirically-derived) knowledge about how interaction works. This is also evident in role-played and simulated interactions of various kinds, in which people talk differently when being assessed, trained, mimicked, etc. – and where their stake in the interaction is one of being a student, assessor, actor – rather than when they are interacting with the stake that has its natural home in the target setting (e.g., police interviewing suspects, rather than training to interview with actors; doctors consulting with patients, rather than taking communication skills assessments with simulated patients, mystery shoppers calling veterinarian practices, see Atkins, 2019; Stokoe, 2013). In this way, humans themselves may “fail” to simulate being other humans when relying on “common-sense intuitions about how some action is accomplished” (Schegloff, 1996) rather than how it actually is (Stokoe et al., 2020).

“Conversational” as a “category error” in the development of LLMs

As noted at the start of this article, the proliferation of LLMs has generated equivalent volumes of research and commentary about their achievements and failures, often against a backdrop of “conversational” norms and metrics. The practice of evaluation, however, stands in stark contrast to a fundamental principle of conversation analysis whereby analysts remain “indifferent” or agnostic to the “quality” of participants’ encounters. This is partly because, in CA, conversation is conceived of in the same way that one might conceive of a pebble on the beach. There are no “good” or “bad” pebbles but there are geological principles that govern their different characteristics. Thus, the analytic task is to *describe* rather than *evaluate* interaction.

A recent case gets to the heart of this “category-mistake” (Ryle, 2009, p. 6) when it comes to thinking through these issues with a conversation analytic lens. Referring to a pre-print research paper entitled “LLMs Get Lost in Multi-turn Conversation” (Laban et al., 2025), a tech founder working with AI agents wrote that “[t]his is one of the most common issues when building with LLMs today” (Gonzalez, 2025). Summarizing the paper’s main findings, the founder listed several reasons why LLMs “get ‘lost,’” including making “premature and often incorrect assumptions early in the conversation,”

attempting “full solutions before having all necessary information, leading to “bloated” or off-target answers,” and producing “overly verbose outputs, which can further muddle context and confuse subsequent turns.” Each of these – the author offers several prompt engineering-based solutions – is treated as undesirable for interaction, and a failure on the part of LLMs. Yet each of the problems described could equally be observed in human-human interaction.

Consider [Extracts 7](#) and [8](#), which come from two different sales environments. [Extract 7](#) shows a salesperson “cold-calling” another business with the goal of getting the call-taker to become a prospective client by making an appointment to meet at the end of the call. [Extract 8](#) comes from the start of a call from a potential domestic client of a windows sales company.

Extract 7: B2B Sales

- 01 S: .hhh hello I wonder if you can help (uh me), It's
 02 Electe:c; .hh Print: calling, =The manufacturers direct? (.)
 03 an:d uhm I'm ↑calling actually .hh as we are appointed your
 04 preferred supplier now, via the Hamil County Council; .hh
 05 and wanted to ask your business manager, .hh w- actually
 06 on: relation to you:r uh copier contracts. As to when that
 07 you're reviewing next and when they might expire.
 08 C: .ptk [(.) okay. =We're NOT actually] with Hamil:: uhm:=
 09 S: [.hhh ()]
 10 C: =County now. We're actually an academy:?

Extract 8: Windows

- 01 C: Oh hello. =Could you just tell me roughly: u:m .pt how
 02 much your: u:m- what uh he w- basically it's for a
 03 rental property so obviously hh I- uh: it's not one I
 04 live in, and um: I just wondered what prices your lean-
 05 tos started at.
 06 S: Oh, crikey.

In each case, we may evaluate the interactions as failing the tests levelled at LLMs, by making “premature and often incorrect assumptions early in the conversation” ([Extract 7](#)) or producing “overly verbose outputs” which then “muddle context and confuse subsequent turns” ([Extract 8](#)). While the human calls use different interactional resources than LLMs when producing verbosity or confusion, these are not

“failures of conversationality” – even if we may regard these as “bad” conversations. Conversation analysis has identified and described the robust and systematic interactional machinery that scaffolds all conversation. This means that even when interactions are full of misunderstandings, are egregious, toxic, dangerous, and so on, participants themselves cannot be “wrong” in their deployment of the machinery. This perspective leads us away from the notion of “degrees of conversationality” and, more broadly, away from attributing human or non-human categories to conversationalists, and towards thinking differently about what counts as “conversationality” in any given product or service. In the context of LLMs and other increasingly sophisticated interactional user interfaces, we must look for different kinds of criteria, benchmarks, and frameworks to assess their “conversational” abilities (Albert et al., 2025).

Concluding remarks

In this paper, we have aimed to consider what “conversational” means in the context of a wide range of settings, focusing especially on the implications for AI-augmented products and services. One of the things that all our examples have shown is that human-human interaction is full of opportunities for co-constructing action. This might be through recipient-designed features (as in pre-sequences), or a mechanism of the “epistemic engine” (Heritage, 2012), such as disfluencies that mark a lack of knowledge. In the final two cases, in which human speakers produce talk that would be evaluated negatively by those deciding what a “good” LLM interaction looks like, we see not just the participants’ attempts to build a request, but all sorts of other components in their turn design – other actions, repair initiations, perturbations, pauses, etc. – from which recipients might pick up on who participants are, what they know and do not know, what they want and do not want, and so on. Such features are distinct from LLM turns, which tend to be either stand-alone, fully sentential, polished responses, or fully realized base-sequences (usually canonical information-seeking questions) that are only co-constructual in that they seek a response.

The development of conversational technologies continues apace. This paper, like others (including a special issue on conversation analysis and conversational technologies, see Stokoe et al., 2024), might seem to only be able to mark a moment in time. However, we argue that the machinery of social interaction – the “generic orders” of interaction “without which it cannot proceed” (action formation, turn-taking, repair, sequence organization, word selection, and structural organization: Schegloff, 2007, p. xiv) – remain stable. These “generic orders” – to which we might add something fundamental to LLMs about participants’ “relative rights to knowledge displayed and managed in sequentially organized activities” (Whitehead et al., 2025, p. 38) – provide criteria against which anything “conversational” can be evaluated. In this way, we preserve the integrity of human social interaction as the primordial machinery of human social interaction.

References

- Albert, S., Housley, W., Sikveland, R. O., & Stokoe, E. (2025). The conversational action test: Detecting the artificial sociality of artificial intelligence. *New Media & Society*, 27(10), 5592–5621. <https://doi.org/10.1177/1461444825133827>

- Antaki, C., & Rapley, M. (1996). Questions and answers to psychological assessment schedules: Hidden troubles in 'quality of life' interviews. *Journal of Intellectual Disability Research*, 40(5), 421–437. <https://doi.org/10.1046/j.1365-2788.1996.779779.x>
- Atkins, S. (2019). Assessing health professionals' communication through role-play: An interactional analysis of simulated versus actual general practice consultations. *Discourse Studies*, 21(2), 109–134. <https://doi.org/10.1177/1461445618802659>.
- Cameron, D. (2000). *Good to talk? Living and working in a communication culture*. Sage. <https://doi.org/10.4135/9781446217993>
- Capaday, C. (2002). The special nature of human walking and its neural control. *Trends in Neurosciences*, 25(7), 370–376. [https://doi.org/10.1016/S0166-2236\(02\)02173-2](https://doi.org/10.1016/S0166-2236(02)02173-2)
- Dingemanse, M., Liesenfeld, A., Rasenberg, M., Albert, S., Ameka, F. K., Birhane, A., Bolis, D., Cassell, J., Clift, R., Cuffari, E., De Jaegher, H., Dutilh Novaes, C., Enfield, N. J., Fusaroli, R., Gregoromichelaki, E., Hutchins, E., Konvalinka, I., Milton, D., Rączaszek-Leonardi, J., ... Wiltshko M. (2023). Beyond single-mindedness: A figure-ground reversal for the cognitive sciences. *Cognitive Science*, 47(1), e13230. <https://doi.org/10.1111/cogs.13230>
- Drew, P. (2005). Conversation analysis. In K. L. Fitch, & R. E. Sanders (Eds.), *Handbook of language and social interaction* (pp. 71–102). Lawrence Erlbaum Associates Publishers.
- Elhilali, A., Ngo, A. S. H., Reichenpfader, D., & Denecke, K. (2025). Large Language Model-based patient simulation to foster communication skills in health care professionals: User-centered development and usability study. *JMIR Medical Education*, 11, e81271. <https://doi.org/10.2196/81271>
- Enfield, N. J., & Sidnell, J. (2023). Consequences of reflexivity in language. In F. A. Karakostis & G. Jäger (Eds.), *Words, bones, genes, tools: Biocultural implications* (pp. 203–214). Kerns Verlag. <https://doi.org/10.51315/9783935751384.009>
- Flinkfeldt, M., Parslow, S., & Stokoe, E. (2022). How categorization impacts the design of requests: Asking for email addresses in call-centre interactions. *Language in Society*, 51, 693–716. <https://doi.org/10.1017/S0047404521000592>
- Fox, B. (2014). On the notion of pre-request. *Discourse Studies*, 17(1), 41–63. <https://doi.org/10.1177/1461445614557762>
- Garfinkel, H. (1991). Respecification: Evidence for locally produced, naturally accountable phenomena of order*, logic, reason, meaning, method, etc. in and as of the essential haecceity of immortal ordinary society (I): An announcement of studies. In G. Button (Ed.), *Ethnomethodology and the human sciences* (pp. 10–19). Cambridge University Press. <https://doi.org/10.1017/CBO9780511611827.003>
- Geiecke, F., & Jaravel, X. (2024). Conversations at scale: Robust ai-led interviews with a simple open-source platform. Available at SSRN 4974382. <https://doi.org/10.2139/ssrn.4974382>.
- Gibson, S. (2013). Milgram's obedience experiments: A rhetorical analysis. *British Journal of Social Psychology*, 52(2), 290–309. <https://doi.org/10.1111/j.2044-8309.2011.02070.x>
- Gonzalez, H. (2025, November 12). User: "Hi, I have a question." Bot: "Sorry, I didn't catch that. Please say things like 'pay my bill' or 'check my order status.'" LinkedIn. https://www.linkedin.com/posts/hectorgonzalezalvarez_conversationlai-cx-uxdesign-activity-7394416551787147264-DYRC?utm_source=share&utm_medium=member_desktop&rcm=ACoAAAPAF-EBGhyjCnhrAC15qliuzpKaH4js_k
- Heritage, J. (2002). Ad hoc inquiries: Two preference in the design of routine questions in an open context. In D. W. Maynard, H. Houtkoop-Steenstra, N. C. Schaeffer, & J. van der Zouwen (Eds.), *Standardization and tacit knowledge: Interaction and practice in the survey interview* (pp. 313–335). Wiley.
- Heritage, J. (2009). Conversation analysis as social theory. In B. S. Turner (Ed.), *The new Blackwell companion to social theory* (pp. 300–320). Blackwell.
- Heritage, J. (2012). The epistemic engine: Sequence organization and territories of knowledge. *Research on Language & Social Interaction*, 45(1), 30–52. <https://doi.org/10.1080/08351813.2012.646685>
- Holstein, J. A., & Gubrium, J. F. (1995). *The active interview*. Sage publications. <https://psycnet.apa.org/doi/10.4135/9781412986120>
- Jefferson, G. (2004). Glossary of transcript symbols with an introduction. In G. H. Lerner (Ed.), *Conversation analysis: Studies from the first generation* (pp. 13–31). John Benjamins Publishing Company. <https://doi.org/10.1075/pbns.125.02jef>
- Kasper, G., & Ross, S. J. (2018). The social life of methods: Introducing the special issue. *Applied Linguistics Review*, 9(4), 475–486. <https://doi.org/10.1515/applirev-2017-0018>

- Laban, P., Hayashi, H., Zhou, Y., & Neville, J. (2025). LLMs get lost in multi-turn conversation. arXiv. <https://doi.org/10.48550/arXiv.2505.06120>
- Levinson, S. C. (2006). On the human “interaction engine. In N. J. Enfield & S. C. Levinson (Eds.), *Roots of human sociality: Culture, cognition and interaction* (pp. 39–69). Routledge.
- Mondada, L. (2018). Multiple temporalities of language and body in interaction: Challenges for transcribing multimodality. *Research on Language & Social Interaction*, 51(1), 85–106. <https://doi.org/10.1080/08351813.2018.1413878>
- Okada, Y. (2010). Role-play in oral proficiency interviews: Interactive footing and interactional competencies. *Journal of Pragmatics*, 42, 1647–1668. <https://doi.org/10.1016/j.pragma.2009.11.002>
- Olson, P. (2015, Month Date). This startup will finally make you like online forms. *Forbes*. <https://www.forbes.com/sites/parmyolson/2015/10/01/startup-typeform-barcelona/>
- OpenAI. (2026). *ChatGPT* (Jan 23 version) [Large language model]. <https://chat.openai.com/chat>
- Pearl, C. (2024). Why can't CUIs tell jokes? Timing. *Discourse & Communication*, 18(6), 987–988. <https://doi.org/10.1177/17504813241271483>
- Potter, J. (2002). Two kinds of natural. *Discourse Studies*, 4, 539–542. <https://doi.org/10.1177/146144560200400409>
- Potter, J., & Hepburn, A. (2007). Life is out there: A comment on Griffin. *Discourse Studies*, 9(2), 276–282. <https://doi.org/10.1177/1461445607075348>
- Prior, M. T. (2018). Accomplishing “rapport” in qualitative research interviews: Empathic moments in interaction. *Applied Linguistics Review*, 9(4), 487–511. <https://doi.org/10.1515/applirev-2017-0029>
- Richardson, E., Stokoe, E., & Antaki, C. (2019). Establishing intellectually impaired victims' understanding about 'truth' and 'lies': Police interview guidance and practice in cases of sexual assault. *Applied Linguistics*, 40(5), 773–792. <https://doi.org/10.1093/applin/amy023>
- Robinson, J. D. (2011). Conversation analysis and health communication. In T. L. Thompson, R. Parrott, & J. F. Nussbaum (Eds.), *Handbook of health communication* (pp. 501–518). Routledge. <https://doi.org/10.4324/9780203846063>
- Rossano, F. (2024, Month Date) [Post] LinkedIn. https://www.linkedin.com/posts/federico-rossano-04b3251b5_gpt-llm-timing-activity-7195920623750008833-S-eo/
- Rubin, G. (2016). *Negotiation power through tag questions in crisis negotiations*. [Unpublished master's thesis]. Georgetown University. <http://hdl.handle.net/10822/1040756>
- Ryle, G. (2009). *The concept of mind*. Original work published 1949. Routledge. <https://doi.org/10.4324/9780203875858>
- Sacks, H. (1989). Lecture fourteen: The inference-making machine. *Human Studies*, 12(3/4), 379–393. <https://www.jstor.org/stable/20009071>
- Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn taking in conversation. *Language*, 50, 696–735. <https://doi.org/10.2307/412243>
- Schegloff, E. A. (1968). Sequencing in conversational openings. *American Anthropologist*, 70, 1075–1095. <https://doi.org/10.1525/aa.1968.70.6.02a00030>
- Schegloff, E. A. (1996). Issues of relevance for discourse analysis: Contingency in action, interaction and co-participant context. In E. H. Hovy, & D. R. Scott (Eds.), *Computational and conversational discourse: Burning issues - An interdisciplinary account* (pp. 3–38). Springer.
- Schegloff, E. A. (2007). *Sequence organization in interaction: A primer in conversation analysis*. Cambridge University Press. <https://psycnet.apa.org/doi/10.1017/CBO9780511791208>
- Shareef, F. (2024). Enhancing conversational AI with LLMs for customer support automation. In *2024 2nd international conference on self sustainable artificial intelligence systems (ICSSAS)* (pp. 239–244). IEEE. <https://doi.org/10.1109/ICSSAS64001.2024.10760403>
- Sheeran, P., & Webb, T. L. (2016). The intention-behavior gap. *Social and Personality Psychology Compass*, 10(9), 503–518. <https://psycnet.apa.org/doi/10.1111/spc3.12265>
- Sikveland, R. O., Kevoe-Feldman, H., & Stokoe, E. (2022). *Crisis talk: Negotiating with individuals in crisis*. Routledge. <https://doi.org/10.4324/9780429354892>
- Silverman, D. (1998). *Harvey Sacks: Social science and conversation analysis*. Polity.
- Silverman, D. (2016). *Qualitative research* (4th ed). Sage.
- Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., De Ruiter, J. P., Yoon, K.-E., & Levinson, S. C. (2009). Universals and cultural variation in turn-taking in

- conversation. *Proceedings of the National Academy of Sciences of the United States of America*, 106(26), 10587–10592. <https://doi.org/10.1073/pnas.0903616106>
- Stokoe, E. (2013). The (in) authenticity of simulated talk: Comparing role-played and actual interaction and the implications for communication training. *Research on Language & Social Interaction*, 46(2), 165–185. <https://doi.org/10.1080/08351813.2013.780341>
- Stokoe, E. (2018). *Talk: The science of conversation*. Little, Brown.
- Stokoe, E. (2021). The sense of a conversational ending. *Nature*, 593(7859), 347–349. <https://doi.org/10.1038/d41586-021-01074-z>
- Stokoe, E., Albert, S., Buschmeier, H., & Stommel, W. (2024). Conversation analysis and conversational technologies: Finding the common ground between academia and industry. *Discourse & Communication*, 18(6), 837–847. <https://doi.org/10.1177/17504813241267118>
- Stokoe, E., Antaki, C., Chrisostomou, L., Henderson, E., & Stewart, S. (2026). The softness of hard data: Discursive psychology, conversation analysis, and psychological science. *Qualitative Psychology*. <https://doi.org/10.1037/qap0000358>
- Stokoe, E., & Richardson, E. (2023). Asking for help without asking for help: How victims request and police offer assistance in cases of domestic violence when perpetrators are potentially co-present. *Discourse Studies*, 25(3), 383–408. <https://doi.org/10.1177/14614456231157293>
- Stokoe, E., Sikveland, R. O., Albert, S., Hamann, M., & Housley, W. (2020). Can humans simulate talking like other humans? Comparing simulated clients to real customers in service inquiries. *Discourse Studies*, 22(1), 87–109. <https://doi.org/10.1177/1461445619887537>
- Stokoe, E., Sikveland, R. O., & Symonds, J. (2016). Calling the GP surgery: Patient burden, patient satisfaction, and implications for training. *British Journal of General Practice*, 66(652), e779–e785. <https://doi.org/10.3399/bjgp16x686653>
- Strietholt, S., & Petron, E. (2025). From pop psychology to L2 pedagogy: The ‘Let Them Theory’ as a reflective frame in applied linguistics. *Applied Linguistics*, 46(6), 958–963. <https://doi.org/10.1093/applin/amaf058>
- Thornberg, R., Halldin, K., Bolmsjö, N., & Petersson, A. (2013). Victimising of school bullying: A grounded theory. *Research Papers in Education*, 28(3), 309–329. <https://doi.org/10.1080/02671522.2011.641999>
- Van Edwards, V. (2025, September 15). *Don't ask 'How are you?' Here's how successful people get others to like and trust them: Psychology expert*. CNBC. <https://www.cnn.com/2025/09/15/dont-ask-how-are-you-how-successful-people-get-others-to-like-and-trust-them-psychology-expert.html>
- Wass, V., Roberts, C., Hooogenboom, R., Jones, R., & Van der Vleuten, C. (2003). Effect of ethnicity on performance in a final objective structured clinical examination: Qualitative and quantitative study. *British Medical Journal*, 326(7393), 800–803. <https://doi.org/10.1136/bmj.326.7393.800>
- Whitehead, K. A., Stokoe, E., & Raymond, G. (2025). Categories in social interaction: Unlocking the resources of conversation analysis and membership categorization for psychological science. *Annual Review of Psychology*, 76, 531–557. <https://doi.org/10.1146/annurev-psych-020124-023147>
- Wieder, D. (1988). From resource to topic: Some aims of conversation analysis. *Annals of the International Communication Association*, 11(1), 444–454. <https://doi.org/10.1080/23808985.1988.11678701>
- Woofttt, R. (2007). Communication and laboratory performance in parapsychology experiments: Demand characteristics and the social organization of interaction. *British Journal of Social Psychology*, 46(3), 477–498. <https://doi.org/10.1348/014466606x152667>