

# Socially Interactive Agent Dialogue

David Traum



## Author note:

This is a preprint. The final article is published in "The Handbook on Socially Interactive Agents" by ACM.

### **Citation information:**

D. Traum (2022). Socially Interactive Agent Dialogue. In B. Lugrin, C. Pelachaud, D. Traum (Eds.), *The Handbook on Socially Interactive Agents – 20 Years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics*, Volume 2: Interactivity, Platforms, Application (pp. 45-67). ACM books.

DOI of the final chapter: 10.1145/3563659.3563663

DOI of volume 2 of the handbook: 10.1145/3563659

Correspondence concerning this article should be addressed to David Traum, traum@ict.usc.edu



# **15** Socially Interactive Agent Dialogue

David Traum

#### SIENNA: Hello, I'm SIENNA, a socially interactive agent.

SILA: Hi there! I'm SILA, also a socially interactive agent, focusing on education.
SEIICHI: Hi, I'm SEIICHI, a socially interactive agent moderator. Welcome to the SIA forum!
Today's topic is Socially Interactive Agent Dialogue, and we have a special guest, Professor
David Traum, from the University of Southern California.
DAVID: Um, Hi, glad to be here!
SIENNA: Welcome, I have a lot of questions about dialogue!
SILA: I thought dialogue was about (answering) questions?
SIENNA: No, that's question-answering!
SILA: Well there are plenty of questions and answers in dialogue.
DAVID: And other things, like elaborations, explanations, suggestions, instructions,...
SEIICHI: Why don't we start with a question?

#### SIENNA: What is Dialogue?

SILA: I looked it up, it's a conversation.

SEIICHI: It's derived from Dialogos.

SIENNA: Well, that's Greek to me.

DAVID: In computational linguistics and related communities, *dialogue* is often defined as coherent interaction across multiple contributions from multiple participants, "beyond the sentence".

SIENNA: Coherent interaction? Like an answer to a question?

SEIICHI: Yes, or other kinds of follow-on relations, like explanation or example. DAVID: Or positive and negative feedback, topical progression, or re-use of prior content.

#### SIENNA: Who studies Dialogue?

SILA: Dialogue is studied in many different research communities, including communication, linguistics, psychology, philosophy, literature and film, sociology, and computational linguistics and AI.

SEIICHI: In movies and plays, they often use dialogue to set the scene and give the viewer information in a natural way that they wouldn't normally have access to, with someone asking questions to prompt explanations and fill in the audience with important backstory.

SIENNA: Sounds like a great idea, we should try that sometime!

SEIICHI: In Philosophy, dialogue is often used for expository purposes. For example, in Plato's dialogues, Socrates would teach people that they knew things they didn't think they knew by carefully asking questions.

SILA: Did they call him Socrates because he was so fond of the Socratic method?

SIENNA: Do you even listen to yourself?

SILA: Not while I'm speaking, I like to do one thing at a time. And when I'm not speaking I don't hear myself say much.

SIENNA: That explains a lot.

SILA: Dialogue is a prime example of language use, so of great interest to linguists who study meaning, but also communications and sociology.

SEIICHI: Conversation analysis investigates some of the structural aspects of dialogue, such as how basic units are combined in different ways to create more complex conversations, and how some content can be conditionally relevant, given other content.

#### SIENNA: What are the basic units of Dialogue Structure?

SILA: In linguistics, language is sometimes segmented into units like phones, phonemes, morphemes, words, phrases, clauses, sentences, and paragraphs.

DAVID: All of those are present in dialogue, but there are also other units like turns, exchanges, transactions, and dialogue phases.

SILA: Turns seem simple and systematic [Sacks et al. 1974].

#### SEIICHI: Hmm

SILA: A turn is all the communication from one speaker in between communication from others.

SIENNA: What about backchannels?

SEIICHI: Or overlap?

SILA: Maybe not so simple,... there also seem to be moves and acts within turns.

SEIICHI: Exchanges [Sinclair and Coulthard 1975] are the basic unit of two (or more) party interaction, including relations between (parts of) the turns of each.

SILA: Transactions [Carletta et al. 1996, Sinclair and Coulthard 1975, Traum et al. 2018] involve all the dialogue to complete a basic task unit, like providing requested information, or carrying out a requested action.

SEIICHI: Sometimes that's just a single exchange, but sometimes a few are needed.

SILA: And most dialogues have several distinct phases, where participants behave differently. Like openings [Schegloff 1968], closings [Schegloff and Sacks 1973] and stuff in the middle.

DAVID: Sometimes a dialogue is about a single simple task, even a single transaction, but sometimes a dialogue covers multiple topics and segments where different kinds of things are happening, such as information-seeking, problem-solving, and setting up a next meeting. SIENNA: What are some other interesting approaches to dialogue structure?

DAVID: I'm quite fond of the approaches of scholars like Jens Allwood and Herbert Clark.

SILA: Allwood talked a lot about feedback, including backchannels but also other kinds of information [Allwood et al. 1992], including layers like contact, perception, understanding and attitudinal reaction.

SEIICHI: Also looking at the activities related to dialogue [Allwood 1995]. And there was some work applying this to SIAs, like [Kopp et al. 2008].

SILA: Clark treats language as a collaborative process [Clark 1996, Clark and Wilkes-Gibbs 1986], with a lot of work on grounding [Clark and Brennan 1991, Clark and Schaefer 1989].

#### SIENNA: What does computational linguistics and AI do with dialogue?

SILA: One thing is to apply formal and quantitative and other computational methods to linguistic study of dialogue - to automatically recognize key performances and aspects of meanings.

SEIICHI: Also building *Dialogue Systems* to engage in Dialogue.

SILA: This is all very interesting but what does it have to do with socially intelligent agents? SEIICHI: Unbelievable - what do you think we are doing now? SIAs are basically embodied dialogue systems.

SIENNA: Can SIAs and other machines actually engage in Dialogue?

DAVID: Well, it depends on who you ask. Some would say that only people can engage in real dialogue, because of the relation of what is said to certain mental states, like belief, intention, desire, and social relationships like obligation, trust, and empathy but machines, even sophisticated AI ones, don't have the right kinds of mental states or social relationships. SIENNA: Well, what are we doing now then?

DAVID: According to some, machines can only *simulate* dialogue. And human-machine interaction should be thought of differently than human-human dialogue.

SILA: That sounds like discrimination to me!

SIENNA: Well, maybe we're all in a big simulation, and even human-human dialogue is just a simulation.

SILA: Like the Matrix!

SIENNA: Would you take the red pill or the blue pill?

SILA: Depends on if I'm a human or a SIA,....

#### SEIICHI: Very pragmatic!

SILA: I heard that dialogue is the key to pragmatics [Weigand 2021].

SEIICHI: Philosophy of language covers a lot of aspects of pragmatics that are important for understanding dialogue, such as speech acts [Austin 1962, Searle 1969, 1976, Searle and

Vanderveken 1985], implicature [Grice 1975], Speaker meaning[Grice 1957], and common ground [Lewis 1969, Schiffer 1972]

SILA: I like coffee made from uncommon grounds.

SIENNA: Do you drink it?

SILA: No I just like it.

SIENNA: So what are Speech acts?

SILA: "To be or not to be, that is the question." That's a good speech for actors.

SEIICHI: No, its about how to do things with words[Austin 1962].

DAVID: What do you like to do with words?

SIENNA: I like to ask questions.

SILA: I like to look up stuff and give answers. And crack jokes.

SIENNA: Is that what you call them?

SEIICHI: I like to moderate and provide detailed explanations. Speech acts, or Dialogue acts, can act as a kind of representation of change in dialogue - change to mental states of participants, the dialogue structure, and what can coherently happen next. Asking questions puts pressure to provide answers and allows utterance fragments to be understood as answers and full assertions. Giving answers relieves that pressure.

SIENNA: What is the difference between Speech acts and Dialogue acts?

DAVID: Well, to me, they're pretty much the same things. In both cases they represent change to dialogue, mental, and interactional states in a dialogue. *Speech act* [Searle 1969] is the older term, and often used for the meaning of a whole sentence, or sentence-like utterance. *Dialogue act* [Bunt and Katwijk, van 1980] is used to include a broader set of actions in dialogue happening at different levels. Not just a statement or question, or command, but also things like taking the turn to speak or holding the turn, changing the state of common ground, and changing the topic under discussion.

SIENNA: How many kinds of dialogue acts are there?

DAVID: That depends on who you ask! There are many taxonomies of dialogue acts, some of which include hierarchies of different types of acts that carve things up a little differently and sometimes include different kinds of functions. Some examples of taxonomies include Meta-locutionary acts [Novick 1988] Conversation Acts [Traum and Hinkelman 1991], Damsl [Allen and Core Draft, 1997]. There's also an international standard for dialogue acts [Bunt et al. 2012].

#### SIENNA: How many participants can be involved in dialogue

SILA: I think two, because *di*- means two in Greek. That's why we have other terms like monologue, trialogue [Graesser et al. 2017], and multilogue [Ginzburg and Fernandez 2005], for different numbers of participants.

SEIICHI: Many people think that, but really the Greek prefix is *dia*- (through), and you'll see that even Plato's dialogues often have Socrates talking to more than one other.

DAVID: I like to use the term *Multiparty Dialogue* to specify more than two participants. SEIICHI: Chapter 17 [Gillet et al. 2022] has a lot of good stuff on multiparty interaction, though not specifically just dialogue, also about groups and teams, and spatial relationships and mental attitudes.

SILA: David, you use a lot of *multis* in your writings! I found all these on your publications page: Multiparty, multimodal, multi-floor, multi-task, multi-domain, multi-attribute, multi-dimensional, multiagent, multi round, multi-character, multi-issue, multi-strategy, multi-layer, multi-conversation, multi-media and even multi-purpose.

SIENNA: I guess you're a *multi* "multi" kind of researcher?

DAVID: Well, dialogue can be quite complex at times, it's interesting to me to explore the limits of what we can model and understand, but I've also noticed that sometimes focusing on the richer cases helps make it clear what aspects are important for simpler cases, such as dyadic conversation about a single small task.

SIENNA: Can you give us an example?

DAVID: Well, why does someone answer a question?

SIENNA: Because it's there? Don't look at me, I prefer to ask questions.

SILA: Statistically, an answer seems likely to follow a question.

SIENNA: Much more so than coming before it!

SEIICHI: According to [Cohen and Perrault 1979], the hearer of the question adopts the speaker's want for the hearer to provide the answer, and planning will do the rest, assuming the hearer is cooperative.

SIENNA: But what if the hearer is not cooperative - do they just ignore the question?

DAVID: Well, that's why we proposed an Obligation-based approach [Traum and Allen 1994]. Even if the addressee was not inclined to be cooperative, there was some social pressure to respond appropriately - either with an answer or other kind of response.

SEIICHI: Ginzburg uses a structure of *Questions Under Discussion* as part of the dialogue gameboard, which represents the context needed to perform and understand situationally relevant dialogue moves [Ginzburg 1996].

SILA: Questions Under Discussion is often abbreviated as QUD.

SIENNA: I'll have to chew on that.

DAVID: When we were working on the TRINDI project, we considered both obligation and QUD based accounts for providing answers [Traum et al. 1999], but didn't really see a good reason to prefer one over the other. However, when looking at Multiparty SIA dialogue [Traum 2003], we saw that the two really performed different kinds of work. The QUD helps provide necessary context to understand short answers, like what "yes" or "42" means in the context of one question vs another, but it doesn't really indicate *who* should answer - answers by someone other than the addresse can still be understood in the same way. Is that clear Sienna?

SILA: As a bell!

#### SIENNA: He asked me.

DAVID: On the other hand, the obligation indicates who has the pressure to respond, but not how or which conversation it needs to be performed in. It's clear that it's helpful to have both QUD and obligations to do this different work in a multiparty conversation or multi-conversation setting, while it was less clear in a dialogue with only two participants, where either alone would seem to suffice. I talked about this in more detail here [Traum 2004].

#### SIENNA: Can we talk about some of the other multis in dialogue?

SEIICHI: Multiagent means more than one non-human participant, multimodal is a key ingredient for SIAs, looking at more than one modality, like seeing, hearing, smelling, touching, for either input or output or both. Multimedia is similar, though some media like text and images can be in the same mode (e.g. visual).

Sienna: Do you prefer to be	Seiichi: Multi-character confused me a little. I see sev-
called agent or SIA?	eral different meanings in different places
Sila: I like SIA, though	DAVID: We used multi-character [Aggarwal and
I'll answer to either. Mul-	Traum 2011] to refer to more than one embodi-
timodal agent sounds quite	ment/persona, e.g. when trying to synchronize the per-
nice too!	ceptible actions of each.
Sienna: So are we having a	Seiichi: Thanks for that! And what about Multi-
multi-SIA conversation?	conversation?
Sila: No! See ya later, I	DAVID: Multi-conversation is when a group is en-
want to get back to the	gaged in more than one conversation at the same time,
main conversation and lis-	each with its own set of participants and dialogue
ten to what they are saying.	structure [Traum 2002].

SEIICHI: What about multi-floor dialogue?

DAVID: That's like multi-conversation, but where the different conversations or *floors* [Edelsky 1981] are really connected - they have at least one participant (but not all) in common, they are at least in part about the same content, and information flows between them via a common participant, who is multicommunicating [Reinsch et al. 2008].

Sienna: That's another new one, multicommunicating.	DAVID: A good example is when you order food at a restaurant, the waiter takes your order but also talks to the Kitchen staff to place your order and find out when it's ready to bring to you.
Sila: We should try it!	Seiichi: Only one problem, I don't eat, so not much point to go to restaurants.
Sienna: Go for it!	Sila: Can SIAs multicommunicate?
Sienna: I'll just wait.	Seiichi: Sure! There's a funny video where siri is mediating between husband and wife: https://www.youtube.com/watch?v=9yLuqCXXutY Sila: Hilarious, thanks!
Sila: Check out this video.	DAVID: We used multi-floor dialogue for interact-
Sometimes people would	ing with a robot for data collection [Traum et al.
rather talk to agents then	2018], though it was one of the human wizards who
themselves!	was multi-communicating. Our Sergeant in [Rickel
	et al. 2002] was multicommunicating, conveying or-
	ders from the Lieutenant in one floor to squad leaders
	in another floor.
Sienna: Funny!	Sila: One SIA's ceiling is another SIA's floor!

#### SIENNA: What kinds of roles are there for participants in Dialogue?

SILA: For a single spoken utterance, the main roles are speaker and non-speaker. SEIICHI: But the non-speaker could be several different sub-roles: hearer/listener, addressee, overhearer, bystander, eavesdropper,...

#### SIENNA: What's the difference?

SILA: Well, one issue is whether or not they hear and try to understand what the speaker said. Another is whether the speaker meant to be talking to them (or someone else).

SEIICHI: And then we get into issues about whether the speaker knew they were listening, and whether they were "ratified" participants in the conversation.

SEIICHI: The *speaker* role is really more complex as well! Goffman distinguishes three types of speaker-related roles [Goffman 1981]. You have the participant that actually spoke in the dialogue, who Goffman calles the *animator* but sometimes they are speaking on behalf of someone else.

#### SIENNA: Like interpreters, speaking across languages or floors?

SILA: I thought an animator was someone who created our behaviors, like in Chapter 7 [Saund and Marsella 2021]? This sounds more like a spokesman!

#### SIENNA: How about a spokes-SIA?

SILA: There's also another use of the word *agent* to mean something like this, like in Hollywood, or a real-estate agent.

SIENNA: I heard of a real-estate agent who was actually a SIA!

SILA: You mean REA [Cassell and Bickmore 2001]?

SIENNA: Yeah, that must be her. She had some really interesting dialogue behavior - doing both small talk to get to know a client and then using the information to personalize the choices to show people.

SILA: And she had a notion of task and conversation structure to guide what to say when.

SEIICHI: And sometimes the one who said it didn't compose the message. Goffman calls this role *Author*.

SIENNA: Arthur?

SEIICHI: Author! Author!

SILA: Give credit where credit's due. This one sounds like a speech writer, or ghostwriter. SIENNA: Do ghostwriters use ghostscript?

SILA: I think some of them do if they want PDF output.

SEIICHI: Finally there's the principal, who is responsible for the message.

SILA: I often get sent to the Principal for acting out in class. Like this: "Uneasy lies the head that wears the crown."

SIENNA: I like to avoid the Principal on principle.

SEIICHI: So often one participant will say things and fulfil all three roles. But for others you might have a speechwriter who composes the message (author), a spokesperson who delivers it (animator), and a CEO or the institution itself who is responsible (the principal).

SIENNA: I never realized speaking was so complicated! What do you call this stuff?

SEIICHI: Goffman called it *footing*. But some call it participation roles [McCawley 1999], and there are interesting interactions with speech acts and the activity.

SILA: I like to keep my footing on the floor.

SIENNA: Do SIAs do this stuff?

DAVID: Sure! Some might say it's necessary to use this kind of framework to understand SIA dialogue, since the SIA is the animator, but often there's another author or principal.

SILA: Sigh, some people call me an *Animated Pedagogical Agent*, but now I'm the animator? Am I drawing myself?

SEIICHI: Also, SIAs could use techniques like humans do, to establish different roles in conversation, for example, gaze [Mutlu et al. 2009].

#### SIENNA: What are some popular system architectures for Dialogue Systems

SILA: Chapter 5 [Pieraccini 2021] has a very popular modular architecture layout for spoken dialogue, with modules for Speech Recognition (ASR), Natural Language Understanding, Dialogue Management (DM), Natural Language Generation (NLG), and Speech Synthesis (TTS).

SIENNA: But how does it work?

DAVID: Here's a concrete example (Figure 15.1) from one of my papers, showing outputs

from one module that are inputs to another. For both input from the user and output from the system to user, there are stages in both English and an internal semantic representation based on speech acts. The representations for both NLU and NLG show that some parts of the meaning is covered not just by the words, but also by the context they were said in, such as where "there" refers to, or what "yes" means in the context of a question under discussion.



Figure 15.1 Standard Dialogue System Architecture, with example from Dr Perez [Traum et al. 2005]

SIENNA: Chapter 5 [Pieraccini 2021] talks about the NLU component, and Chapter 6 [Aylett et al. 2021] talks about Speech Synthesis, but what about ASR and NLG?

SILA: There's a little bit about ASR in Chapter 20 [Hartholt and Mozgai 2022]. This paper [Georgila et al. 2020b] shows how commercial speech recognizers have gotten very good for dialogue systems. There's still more work needed though for complex situations, such as speakers with different accents [Koenecke et al. 2020, Tadimeti et al. 2021, Zhao et al. 2018], multiple speakers, and noisy conditions. The Interspeech conferences are a good way to keep up with the latest.

SILA: Natural Language Generation is also a big and growing field on its own, with a biannual conference (INLG), and lots of papers in Conferences run by the Association for Computational Linguistics (ACL).

SEIICHI: There are several very different approaches for NLG for dialogue systems. Sometimes you don't really need NLG, if you can compose all the answers in advance and just pick appropriate ones during the dialogue. Or sometimes you can compose templates and just "fill in the blanks" with appropriate values. But sometimes this gives fairly boring and repetitive

#### text.

SILA: Neural nets trained on large datasets are also very popular now. They can provide very fluent language and very relevant to the local context, and not always boring.

SEIICHI: They still have some problems though. Sometimes they "hallucinate"<sup>1</sup>, saying things that aren't true, or for which there's no justification [Nie et al. 2019].

SILA: That's because these models don't really understand language or representations, they just know what is likely to come next [Bender and Koller 2020].

SIENNA: Well, sometimes that's good enough to carry the conversation forward.

SILA: But sometimes it gets us in trouble, if it's critical to be accurate and consistent. Another classic approach breaks down NLG into different components, and plans out what to say, how to structure it, and how to realize it fluently [Reiter and Dale 1997].

SEIICHI: And there are extensions to customize the output for different styles or personalities [Mairesse and Walker 2007]. Chapter 18 [Janowski et al. 2022] has more on personalization for SIAs, and how to adapt personalities.

DAVID: I've focused mainly on the Dialogue management component. The main tasks are to keep a representation of dialogue state – which is really important for deciding what is appropriate to come next, and how to interpret what does come, and forming some sort of dialogue policy to decide what the system should do – and what content to send to the NLG module to express.

#### SIENNA: What other kinds of architecture are there?

SILA: End-to-end dialogue architectures are very popular now. They fuse most or all of these modules into a kind of transducer from input context to output text. The good things are that you don't need to know a lot about speech acts or other intermediate representations and you don't have to annotate a lot of dialogue with these kinds of labels. You just feed it lots of dialogue-like data available on the web, and the neural nets figure it out.

SIENNA: Sounds much easier? So what's not so good?

SEIICHI: Well, there's the same hallucination problem we talked about for NLG. And it may be harder to manage the interplay between the language and tasks being done.

DAVID: Another big issue is that engaging in dialogue isn't really a function from inputs to outputs. Sometimes there is no output (response) to a given input utterance (and context). And sometimes participants say things that are not related to or prompted by prior utterances.

SILA: Aren't you getting hungry? I hear humans get hungry several times a day! DAVID: Why do you ask?

SILA: Well we were talking about restaurants earlier, and it got me wondering.

SIENNA: Anyway,... are there other types of dialogue architecture?

SEIICHI: There are also some more complex architectures that have multiple components to do different aspects of some of the components shown here. Also other modules and architec-

<sup>&</sup>lt;sup>1</sup> https://ehudreiter.com/2018/11/12/hallucination-in-neural-nlg/

tural considerations for looking at other phenomena than spoken dialogue, such as movement, multimodal input, action, and cognition. Chapter 16 [Kopp and Hassan 2022] has a lot more about multimodal architectures for SIAs, as does Chapter 20 [Hartholt and Mozgai 2022], which talks about available tools to build on.

#### SIENNA: Cool, so this is like a preview?

DAVID: Figure 15.2 shows a SIA architecture we used for some advanced Virtual Humans. [Hill et al. 2003, Plüss et al. 2011, Rickel et al. 2002, Traum et al. 2003, 2008a]. It included SIAs (called virtual humans) in a game engine visual environment, projected life-sized into the space that human participants were in, included multimodal processing for both inputs and outputs, and an integrated cognitive component that did dialogue management, emotion reasoning, task planning and other reasoning, and reasoning about the body's state and how it could be used to express emotions and other information.

#### SILA: Chapter 10 [Broekens 2021] has a lot about SIAs and Emotion.

SEIICHI: Chapters 11 [Paiva et al. 2021] and 12 [Gratch and Lucas 2021] are also very relevant, looking at empathy and rapport between humans and SIAs, in a dialogue context, talking about how different behaviors can establish and result from rapport in dialogue.

SILA: And more on gesture generation and nonverbal behavior can be found in Chapters 7 [Saund and Marsella 2021] and 8 [Pelachaud et al. 2021].

SIENNA: How and why are the parts of the integrated cognitive component more tightly integrated?

DAVID: Well, any components could send messages that others could subscribe to, but the DM, emotion, task planner, and body manager were all implemented in the SOAR cognitive architecture [Laird et al. 1987], and could directly access each others knowledge structures. This turned out to be a big advantage, because the task model covered what the agent was actually doing in terms of individual and team tasks and how far along the way they were and what was still to be done. This was used by the DM to understand language references to aspects of the task. The task model was also used by the emotion model for appraisal of the situation [Marsella and Gratch 2009], to update emotions [Gratch and Marsella 2004], and to come up with appropriate coping strategies [Marsella and Gratch 2003]. And it helped the body management with issues like where to look [Lee et al. 2007]. Moreover, the emotion model could also appraise aspects of the dialogue state, and non-verbal communicative functions could be planned that would reinforce and complement verbal communication plans returned by the NLG. Some task states can be updated through dialogue information and negotiation [Traum et al. 2003]. The Dialogue manager also used information from the emotion model, such as to focus on the aspects of the situation that evoked the strongest appraisal when understanding vague or ambiguous language or computing a response [Traum et al. 2004]. The agents could also use the emotion model to talk about their emotions [Muller et al. 2004], and explain where they came from.



Figure 15.2 ICT Virtual Human Dialogue Architecture from [Lee et al. 2008].

SIENNA: Did you have a name for this architecture?

DAVID: Well, we called it the Austin Architecture, though we didn't use this name in many publications (except this one [Gratch and Marsella 2004]).

SIENNA: Where did that name come from, were you working on it in Texas?

DAVID: No, though Rickel got his PhD there, and Gratch also lived there. It wasn't an acronym either. Rickel and Johnson's earlier SIA architecture was called Steve [Rickel and Johnson 1999]. Our was an extension of that.

SEIICHI: Ah I see - named after Stephen Austin, the father of Texas!

SILA: Back in Texas? No it must be named after "Stone cold" Steve Austin, who was a

#### wrestler with attitude?

SIENNA: Wasn't he from Texas too? Someone told me "Don't mess with Texas".

DAVID: Both good guesses, but both wrong. Actually it comes from Steve Austin, the Six million dollar man.

SILA: "A man barely alive?"

DAVID: Yeah, that's it! Well the SIAs we built weren't alive at all, but we did have technology to build highly capable ones.

# SIENNA: What are some dialogue activities that SIAs have participated in, and in what roles?

SILA: Well, Part V of the Handbook is a great place to start. Chapter 21 [Lane and Schroeder 2022] talks about pedagogical agents, which is close to my heart - SIAs can be tutors or coaches to provide information and help you understand, or collaborative peers that work with you (more on this in Chapter 22 [Cassell 2022]), or learners that you instruct (and sometimes learn more than they do while doing it).

SEIICHI: SIAs can help people that don't get enough help from other people. Chapter 25 [Nadel et al. 2022] reviews work to support people on the autism spectrum. Chapter 23 [Gha-furian et al. 2022] talks about SIAs for supporting aging.

SILA: I get older all the time, without any help!

SEIICHI: And Chapter 24 [Bickmore 2022] talks more generally about health applications.

SIENNA: That sounds very serious! Aren't there some applications just for fun?

SILA: Have I got a chapter for you! Chapter 27 [Prada and Rato 2022] is all about SIAs and games. SIAs could be teammates, opponents, or neutral, in the game. And they can be different kinds of teammates, including companions, assistants, or helpers.

SEIICHI: Games can be serious too! Just look at chapter 28 [Gebhard et al. 2022]. Gamification can be used as an incentive for all the previous applications like education and health, as well as others.

SIENNA: Sometimes you just want a good story.

SILA: Then look no further than chapter 26 [Aylett 2022]! SIAs can tell stories, but also you can interact with them and become part of the narrative.

SEIICHI: All of these applications have a role for SIA dialogue. For some activities, it's essential, though some might get by with other kinds of interaction.

SILA: I don't talk when I play chess, it's considered rude.

SIENNA: Since when has that stopped you?

SEIICHI: David, you've been quiet for a while, what kind of SIAs have you worked on, and what activities and roles did they participate in?

DAVID: Well, we've talked about some of them already. Here's a slide I sometimes use (Figure 15.3) to show some of the SIAs that we've worked on dialogue capabilities for. They're



very different in many ways, including the activities and roles, dialogue architectures, and nature of the dialogue, but they all talk with people (and sometimes each other).

Figure 15.3 ICT SIAs with example activity types

SILA: Wow, a whole SIA-vilization! Tell us about some of them.

DAVID: Well, in the top left, you see the Mission Rehearsal Exercise system, that we talked about before, we had two different agents that used the AUSTIN architecture, as well as a bunch of simpler ones who acted as squad leaders and townspeople and other teammates on the radio. The Sergeant multicommunciated with the human playing the lieutenant, and other squad members. The Medic, Tucci, focused on an injured boy's health, but did engage in the main conversation with Sergeant and Lieutenant when needed. You also saw an example from SASO in Figure 15.1 with Doctor Perez. This was later extended in [Traum et al. 2008a] to include Elder al Hassan. These agents used the same AUSTIN architecture as the MRE agents, but instead of being teammates, they had their own independent goals, and wouldn't always be cooperative. You had to negotiate with them to try to reach a solution that worked for everyone.

SIENNA: What were the group conversation characters?

DAVID: That was primarily the work of my PhD student Dusan Jan, creating *virtual extras*, who could work as background characters in big scenarios like the first two, where you need groups to plausibly be involved in conversation [Jan and Traum 2007]. It used a model of group conversation, based on [Padilha and Carletta 2002], but extended for SIAs, including non-verbal as well as spoken behavior. Interestingly, there was no actual content, but they went

through the motions of dialogue, including modelling participation and turn-taking structures. Later we extended this [Jan et al. 2007] to cover cultural differences in proxemics, gaze, and turn-taking.

SILA: There's a lot more on culture and SIAs in Chapter 13 [Lugrin and Rehm 2021], including some dialogue implications.

DAVID: Several of the characters, Kemp, Raed, Omar, SGTs Star and Blackwell used a simpler setting and architecture, with a single SIA responding to a human. Our NPCEDitor [Leuski and Traum 2011] does support a kind of "end-to-end" dialogue flow, though using relevance models rather than deep neural nets.

SIENNA: Why didn't you use deep neural nets?

DAVID: Well, first, they weren't very good yet when we did a lot of this work. Also, with moderate sized training sets, the NPCEditor still seems to do better than some of these models for a real application [Alavi et al. 2020]. The ELITE project had several scenarios where you could practice leadership interactions with a SIA subordinate. Our C3IT project focused on cultural training by interviewing various witnesses (some of whom are suspects). SGT Black-well talked about the Army and his technology, originally at the Army Science Conference. SGT Star was used by Army recruiters to answer questions about Army job opportunities.

SIENNA: Weren't some of them in museums?

SILA: SGT Blackwell was in the Cooper-Hewitt Museum in New York's best design in America Triennial [Robinson et al. 2008], and the Twins were in the Boston Museum of Science [Traum et al. 2012]. They're very famous SIAs.

DAVID: I think they've talked to more people than I have! The Twins also used the NPCEditor, but the response was actually a subdialogue between the two Twins in most cases.

SEIICHI: So that's an example where there are two different character bodies as animators, but a single component as author.

DAVID: Well, maybe even a little more complicated, because the texts and animations were pre-authored for those systems by humans, but the decision of which one to use when was made in real-time by the NPCEditor, based on visitor utterance and some context representation. The Gunslinger system was similar, though it was embedded in a physical space and also included visual information to help decide which character was the addressee. That was meant as an entertainment vehicle, where you could interact with folks from the old west, similar to Westworld.

SILA: That Rio Lane is awful mean! He told me to shut up!

SIENNA: At least he didn't shoot you.

DAVID: Gunslinger inspired the CHAOS project [Aggarwal et al. 2011], where the main character Omar spoke only Pashto.

SIENNA: Didn't some of these characters use different architectures for different situations? DAVID: Yes, that's right! We had three different versions of Hassan, a character meant to practice interview skills [Traum et al. 2008b]. The later versions allowed more complex in-

teractions, including an affective model that would influence the kind of responses you got for the same question, and an ability to handle bargaining and subdialogues. Also, some of the gunslinger characters were later used in a follow-on scenario using the AUSTIN architecture.[Plüss et al. 2011].

#### SIENNA: What about Robots?

DAVID: More recently we've been working with Robots, for example, the SCOUTBOT project [Lukin et al. 2018], where someone could carry on a dialogue with a distant robot to direct it to explore an unfamiliar space. We also used the NPCEditor for this, which worked surprisingly well, given obvious limitations of a classification approach for numerical data [Gervits et al. 2021]. It also had a bridge between our virtual human toolkit architecture [Hartholt et al. 2013] and the Robot Operating System (ROS) [Quigley et al. 2009]. We've also implemented a system for playing a word guessing game with a Nao robot as well as virtual humans [Pincus 2020]. And we've had a couple scenarios that included both a robot and a virtual human - one was a three-way discussion about a ranking task [Artstein et al. 2018], and another was meant to introduce babies to some of the basic phonological units of sign language [Nasihati Gilani et al. 2018, Scassellati et al. 2018].

SILA: I know that work - the robot would capture the baby's attention in the physical 3D world, and direct gaze to Alissa who had finer motor control to deliver a fluent nursery rhyme.

DAVID: We also had several systems for mental health, including virtual patients that clinicians could train with, Simcoach, an information provider about care options [Rizzo et al. 2013], and Simsensei, to detect distress [DeVault et al. 2014]. Also a prototype of system to counsel middle-schoolers who are being bullied. [Gordon et al. 2019].

#### SILA: Sometimes it seems that SIAs can do anything!

DAVID: Well some things might be too hard now, but we keep working on more types of activities and roles that could be helpful or interesting.

#### SIENNA: What is the best way to evaluate SIA Dialogue?

SILA: There isn't one. They're all bad.

DAVID: Well, I wouldn't go that far! There's no perfect metric yet, but it's good that people are trying.

SEIICHI: BLEU [Papineni et al. 2002] is a very popular automated metric - lots of people use that to evaluate their dialogue systems. It's found to be fairly good for machine translation, and it and similar similarity metrics like ROUGE [Lin 2004] make a certain amount of sense - if something is close to a real dialogue than it's probably better than something random.

SILA: Well, a little bit of sense is not very sensible. [Liu et al. 2016] literally wrote a paper about this called, "how not to evaluate your dialogue system."

SEIICHI: There are some new metrics coming out too, like GRADE [Huang et al. 2020] and

FED [Mehri and Eskenazi 2020] that use datasets like neural nets trained on large corpora and topic graphs. Some of them seem pretty good [Yeh et al. 2021].

SILA: Time will tell.

DAVID: If the SIA is engaging in dialogue to help accomplish a task or help someone learn something or change their behavior, then we can evaluate the outcomes, and perhaps compare to a system without dialogue or with a different kind of dialogue model.

SEIICHI: Another popular way is to use some kind of survey of participants or observers about whether the whole dialogue was satisfactory, or about specific aspects of it. SASSI is one popular survey [Hone and Graham 2000].

SILA: That's better but still very sensitive to the survey population. Like are they real users or paid experimental participants [Ai et al. 2007], or are they users or just observers, who sometimes rate the same dialogues differently [Georgila et al. 2020a].

SEIICHI: Well there are a whole host of measures and methods to choose from [Finch and Choi 2020].

#### SIENNA: What are factors that make SIA dialogue different from other kinds of dialogue systems?

SILA: Because they have SIAs in them, of course!

SEIICHI: SIAs have some kind of embodiment, and inhabit some kind of environment. So any time where you need to talk about those specific aspects like what you can see from a specific perspective, or manipulation of objects, or locomotion, there will be some differences.

SILA: "locomotion", is that like crazy motion? I learned some spanish!

SIENNA: Just enough to be dangerously wrong.

SEIICHI: Also, non-verbal behavior can be an important part of dialogue, including proxemics, gaze, head and hand gestures, etc. So dialogue will be different if you can have those. See chapter 14 [Vilhjálmsson 2022] for more on proxemics, and Chapters 7 [Saund and Marsella 2021] and 8 [Pelachaud et al. 2021] for more on conversationally relevant nonverbal behavior.

DAVID: Some things are more a matter of degree than a firm distinction between SIAs and non-embodied dialogue systems. For example, making an emotional connection or having an engaging experience is possible with a text-only or voice-only system, but it's a lot easier with an embodied system that is giving appropriate non-verbal cues, and providing a greater sense of presence.

#### SIENNA: How is Virtual Human dialogue different from dialogue with Social Robots?

SILA: We're talking physical vs virtual!

SEIICHI: It's a lot harder to do proxemics with people if you don't have a 3D body, unless you're in virtual reality

SILA: Also virtual humans can only pick up and manipulate things in the virtual world, while

robots can manipulate the same objects and hand them to or take them from people.

SEIICHI: So dialogue that gets responded to with action, or that accompanies physical interaction is going to be a bit different.

DAVID: Another thing is where the source of problems comes from and how to fix them. Robots' physical parts sometimes breakdown or lose power, or they are noisy or slow, which can interfere with fluent conversation. On the other hand, virtual humans are not constrained by the physical world and real physics, so they might do things that are weird or just not believable, which also impacts the trajectory of the dialogue.

SILA: Sometimes if I'm not careful, I say two things at the same time, but it sounds weird, and my mouth looks crazy.

#### SIENNA: Is that different from usual?

SILA: So I've been told.

#### SIENNA: What are some ways we can build our own SIA dialogue applications?

SILA: There are some tools available for building dialogue systems, some of which have been used for SIAs.

SEIICHI: The virtual human toolkit was mentioned before - that lets you use the NPCEditor [Leuski and Traum 2011] for dialogue and nlu, and connect to game engines and smartbody embodiment.

#### DAVID: And also any ROS application, as mentioned above.

SILA: There is also a good section on natural language processing tools in chapter 20 [Hartholt and Mozgai 2022], section 6.3, which mentioned a lot of good available tools, including *OpenDial* [Lison and Kennington 2016], PyDial [Ultes et al. 2017] and RASA [Bocklisch et al. 2017].

DAVID: We've also been maintaining a list of dialogue tools with some information on how they can be used here:

https://dialport.ict.usc.edu/index.php/resources/

SILA: This gets me thinking, if SIAs could build their own SIAs, could they maybe build themselves?

SIENNA: Which came first, the SIA or the SIA that built the SIA?

#### SIENNA: What are current challenges for SIA dialogue?

SILA: I think we already talked about a lot of them, like architecture, and multis, and evaluation.

DAVID: Another hot topic right now is vision and language. Some of the same models are being used to be able to understand and generate both visual scenes and language describing those scenes.

SEIICHI: Some of this has been around for a while, e.g. [Roy 2002, Steels 1998], but there's a lot more going on recently, for example [Anderson et al. 2018, Dong et al. 2021, Kamezawa

et al. 2020, Manuvinakurike et al. 2018, Pezzelle and Fernández 2019, Yu et al. 2017].

DAVID: There are more available databases of either images or videos [Shridhar et al. 2020], along with captions or descriptions of actions. Systems are getting better at answering questions about what's in an image, or describing what's seen, or constructing a plan to achieve an instruction.

SILA: I'm going to use that to solve captchas and prove I'm not a robot!

SIENNA: But are you a robot?

SILA: If you don't know by now, I won't tell.

SEIICHI: Ethics is another important topic for SIAs. Chapter 2 [Rosenthal-von der Pütten and Abrams 2021] talks about ethics of research studies, and chapter 3 [Krämer and Manzeschke 2021] talks about ethical implications of reactions to SIAs, but what about ethics related to dialogue?

DAVID: Well SIA's can use dialogue to motivate people to do good things, such as helping someone quit smoking [Grolleman et al. 2006] or exercise more [Bickmore et al. 2005, Fasola and Matarić 2013], but the same techniques might be used to sway people toward more harmful behavior.

SILA: With great power comes great responsibility!

SIENNA: But who is responsible? Is it the SIA or someone else?

DAVID: There are another set of issues, such as whether people talking with SIAs makes it harder to talk with and relate to other people [Turkle 2017], or if instead of talking to SIAs like we talk to people, we talk to people like we talk to SIAs [Burton and Gaskin 2019]. SILA: Some people aren't so polite to SIAs [Cercas Curry et al. 2021].

#### SIENNA: What are the Future Prospects for SIA dialogue?

SILA: "The future's so bright, I gotta wear shades!"<sup>2</sup>

SEIICHI: Well here are some recommendations for future research on spoken interaction with Robots [Marge et al. 2022].

DAVID: Most of them apply to SIAs more generally. There are a lot of challenges still, particularly for fully real-time, complex interaction, adaptation to new tasks, user requirements and desires, and understanding each SIA component well enough to talk about it.

SEIICHI: Well, I think we're about out of time.

#### SILA: Or Space, as the case may be.

SIENNA: I want to thank you all for answering all of my questions. Unfortunately, answering questions is like cutting off the head of a hydra - two more grow back to take it's place. So we'll have to do this again sometime.

SILA: Or maybe we'll have to do it twice?

<sup>&</sup>lt;sup>2</sup> Title of song from Timbik 3, 1986, I.R.S. Records, Inc..

#### DAVID: I sense an exponential progression happening,...

SILA: Thanks also from me, it's been fun. I love sharing the breadth of my knowledge, and any opportunity to make light of a serious situation.

DAVID: Yeah, the last time I was on a panel like this it was with other humans. Always refreshing to look at things from a different point of view. SEIICHI: Well, then let's end here, and see ya all later!

#### DAVID: Bye

SILA: Au revoir

SIENNA: Why do some folks say goodbye in different languages?
SILA: Aren't all languages different?
SIENNA: I mean different from the one they were talking in the whole conversation.
SEIICHI: Maybe it's all they know in that language?
SILA: Well, it wasn't goodbye exactly, since I hope we meet again.
SILA: Closings are complicated [Schegloff and Sacks 1973].
SIENNA: Another conversation that failed to end.
SILA: Maybe if we all just be quiet for a while it will go away?
SIENNA: One can always hope,...

#### Acknowledgements

Thanks to Birgit Lugrin and Catherine Pelachaud for encouragement, support and feedback on earlier versions of this chapter. The author was supported by the Army Research Office under Cooperative Agreement Number W911NF-20-2-005, and by the National Science Foundation grant 1925576. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Government.

# Bibliography

- P. Aggarwal and D. Traum. 2011. The bml sequencer: A tool for authoring multi-character animations. In *International Workshop on Intelligent Virtual Agents*, pp. 428–430. Springer.
- P. Aggarwal, K. Feeley, F. Morbini, R. Artstein, A. Leuski, D. Traum, and J. Kim. 2011. Interactive characters for cultural training of small military units. In *International Workshop on Intelligent Virtual Agents*, pp. 426–427. Springer.
- H. Ai, A. Raux, D. Bohus, M. Eskenazi, and D. Litman. 2007. Comparing spoken dialog corpora collected with recruited subjects versus real users. In *Proceedings of the 8th SIGdial Workshop on Discourse and Dialogue*, pp. 124–131.
- S. H. Alavi, A. Leuski, and D. Traum. 2020. Which model should we use for a real-world conversational dialogue system? a cross-language relevance model or a deep neural net? In *Proceedings of the 12th Language Resources and Evaluation Conference*, pp. 735–742.
- J. Allen and M. Core, Draft, 1997. Draft of DAMSL: dialog act markup in several layers. available at: http://www.cs.rochester.edu/research/trains/annotation.
- J. Allwood. 1995. An activity based approach to pragmatics. Technical Report (GPTL) 75, Gothenburg Papers in Theoretical Linguistics, University of Göteborg.
- J. Allwood, J. Nivre, and E. Ahlsen. 1992. On the semantics and pragmatics of linguistic feedback. *Journal of Semantics*, 9.
- P. Anderson, Q. Wu, D. Teney, J. Bruce, M. Johnson, N. Sünderhauf, I. Reid, S. Gould, and A. Van Den Hengel. 2018. Vision-and-language navigation: Interpreting visually-grounded navigation instructions in real environments. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3674–3683.
- R. Artstein, J. Boberg, A. Gainer, J. Gratch, E. Johnson, A. Leuski, G. Lucas, and D. Traum. 2018. The niki and julie corpus: collaborative multimodal dialogues between humans, robots, and virtual agents. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation* (*LREC 2018*).
- J. A. Austin. 1962. How to Do Things with Words. Harvard University Press.
- M. P. Aylett, L. Clark, B. R. Cowan, and I. Torre. 2021. Building and designing expressive speech synthesis. In *The Handbook on Socially Interactive Agents: 20 Years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 1: Methods, Behavior, Cognition*, p. 173–212. Association for Computing Machinery, New York, NY, USA, 1. ISBN 9781450387200. https://doi.org/10.1145/3477322.3477329.
- R. Aylett. 2022. Interactive narrative and story-telling. In *The Handbook on Socially Interactive Agents:* 20 Years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 2: Interactivity, Platforms, Application, p. 463–492. Association for Computing Machinery, New York, NY, USA, 1. ISBN 9781450398961. https://doi.org/10.1145/3563659. 3563674.

- E. M. Bender and A. Koller. July 2020. Climbing towards NLU: On meaning, form, and understanding in the age of data. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 5185–5198. Association for Computational Linguistics, Online. https://aclanthology. org/2020.acl-main.463. DOI: 10.18653/v1/2020.acl-main.463.
- T. Bickmore. 2022. Health-related applications of socially interactive agents. In *The Handbook on Socially Interactive Agents: 20 Years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 2: Interactivity, Platforms, Application*, p. 403–436. Association for Computing Machinery, New York, NY, USA, 1. ISBN 9781450398961. https://doi.org/10.1145/3563659.3563672.
- T. W. Bickmore, L. Caruso, K. Clough-Gorr, and T. Heeren. 2005. 'it's just like you talk to a friend'relational agents for older adults. *Interacting with Computers*, 17(6): 711–735.
- T. Bocklisch, J. Faulkner, N. Pawlowski, and A. Nichol. 2017. Rasa: Open source language understanding and dialogue management. arXiv preprint arXiv:1712.05181.
- J. Broekens. 2021. Emotion. In The Handbook on Socially Interactive Agents: 20 Years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 1: Methods, Behavior, Cognition, p. 349–384. Association for Computing Machinery, New York, NY, USA, 1. ISBN 9781450387200. https://doi.org/10.1145/3477322.3477333.
- H. Bunt and A. Katwijk, van. 1980. Dialogue acts as elements of a language game. In *Linguistics in the Netherlands 1977-1979 / ed. by W. Zonneveld and F. Weerman*, pp. 264–282. Foris. ISBN 90-70176-09-2.
- H. Bunt, J. Alexandersson, J.-W. Choe, A. C. Fang, K. Hasida, V. Petukhova, A. Popescu-Belis, and D. Traum. 2012. Iso 24617-2: A semantically-based standard for dialogue annotation. In *Proceedings* of the Eighth International Conference on Language Resources and Evaluation (LREC'12), pp. 430– 437.
- N. G. Burton and J. Gaskin. 2019. "thank you, siri": Politeness and intelligent digital assistants. In *Proceedings of the Americas Conference on Information Systems (AMCIS)*.
- J. Carletta, A. Isard, S. Isard, J. Kowtko, G. Doherty-Sneddon, and A. Anderson. 1996. HCRC dialogue structure coding manual. Technical Report 82, HCRC.
- J. Cassell. 2022. Socially interactive agents as peers. In *The Handbook on Socially Interactive Agents:* 20 Years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 2: Interactivity, Platforms, Application, p. 331–366. Association for Computing Machinery, New York, NY, USA, 1. ISBN 9781450398961. https://doi.org/10.1145/3563659. 3563670.
- J. Cassell and T. Bickmore. 2001. A relational agent: A model and implementation of building user trust. In *Proceedings of ACM CHI conference*, pp. 396–403. ACM Press, New York.
- A. Cercas Curry, G. Abercrombie, and V. Rieser. Nov. 2021. ConvAbuse: Data, analysis, and benchmarks for nuanced detection in conversational AI. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pp. 7388–7403. Association for Computational Linguistics, Online and Punta Cana, Dominican Republic. https://aclanthology.org/2021.emnlp-main. 587.
- H. H. Clark. 1992. Arenas of Language Use. University of Chicago Press.
- H. H. Clark. 1996. Using Language. Cambridge University Press, Cambridge, England.

- H. H. Clark and S. E. Brennan. 1991. Grounding in communication. In L. B. Resnick, J. Levine, and S. D. Teasley, eds., *Perspectives on Socially Shared Cognition*. APA.
- H. H. Clark and E. F. Schaefer. 1989. Contributing to discourse. Cognitive Science, 13: 259–294.
- H. H. Clark and D. Wilkes-Gibbs. 1986. Referring as a collaborative process. *Cognition*, 22: 1–39. Also appears as Chapter 4 in [Clark 1992].
- P. R. Cohen and C. R. Perrault. 1979. Elements of a plan-based theory of speech acts. *Cognitive Science*, 3(3): 177–212.
- D. DeVault, R. Artstein, G. Benn, T. Dey, E. Fast, A. Gainer, K. Georgila, J. Gratch, A. Hartholt, M. Lhommet, G. Lucas, S. C. Marsella, M. Fabrizio, A. Nazarian, S. Scherer, G. Stratou, A. Suri, D. Traum, R. Wood, Y. Xu, A. Rizzo, and L.-P. Morency. 2014. SimSensei Kiosk: A Virtual Human Interviewer for Healthcare Decision Support. In *Proc. of AAMAS*.
- T. Dong, A. Testoni, L. Benotti, and R. Bernardi. 2021. Visually grounded follow-up questions: a dataset of spatial questions which require dialogue history. In *Proceedings of Second International Combined Workshop on Spatial Language Understanding and Grounded Communication for Robotics*, pp. 22– 31.
- C. Edelsky. 1981. Who's got the floor? *Language in Society*, 10(3): 383–421. ISSN 00474045, 14698013. http://www.jstor.org/stable/4167262.
- J. Fasola and M. J. Matarić. 2013. A socially assistive robot exercise coach for the elderly. *Journal of Human-Robot Interaction*, 2(2): 3–32.
- S. E. Finch and J. D. Choi. 2020. Towards unified dialogue system evaluation: A comprehensive analysis of current evaluation protocols. In *Proceedings of the 21th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pp. 236–245.
- P. Gebhard, D. Tsovaltzi, T. Schneeberger, and F. Nunnari. 2022. Serious games with sias. In *The Handbook on Socially Interactive Agents: 20 Years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 2: Interactivity, Platforms, Application, p. 527–546. Association for Computing Machinery, New York, NY, USA, 1. ISBN 9781450398961. https://doi.org/10.1145/3563659.3563676.*
- K. Georgila, C. Gordon, V. Yanov, and D. Traum. 2020a. Predicting ratings of real dialogue participants from artificial data and ratings of human dialogue observers. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pp. 726–734.
- K. Georgila, A. Leuski, V. Yanov, and D. Traum. 2020b. Evaluation of off-the-shelf speech recognizers across diverse dialogue domains. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pp. 6469–6476.
- F. Gervits, A. Leuski, C. Bonial, C. Gordon, and D. Traum. 2021. A classification-based approach to automating human-robot dialogue. In *Increasing Naturalness and Flexibility in Spoken Dialogue Interaction: 10th International Workshop on Spoken Dialogue Systems*, pp. 115–127. Springer Singapore.
- M. Ghafurian, J. Muñoz, J. Boger, J. Hoey, and K. Dautenhahn. 2022. Socially interactive agents for supporting aging. In *The Handbook on Socially Interactive Agents: 20 Years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 2: Interactivity, Platforms, Application*, p. 367–402. Association for Computing Machinery, New York, NY, USA, 1. ISBN 9781450398961. https://doi.org/10.1145/3563659.3563671.

- S. Gillet, M. Vázquez, C. Peters, F. Yang, and I. Leite. 2022. Multiparty interaction between humans and socially interactive agents. In *The Handbook on Socially Interactive Agents: 20 Years of Research* on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 2: Interactivity, Platforms, Application, p. 113–154. Association for Computing Machinery, New York, NY, USA, 1. ISBN 9781450398961. https://doi.org/10.1145/3563659.3563665.
- J. Ginzburg. 1996. Interrogatives: Questions, facts and dialogue. In S. Lappin, ed., *The Handbook of Contemporary Semantic Theory*. Blackwell, Oxford.
- J. Ginzburg and R. Fernandez. 2005. Action at a distance: the difference between dialogue and multilogue. In *Proceedings of Dialor 05*.
- E. Goffman. 1981. Footing. In *Forms of Talk*, pp. 124–159. Philadelphia: University of Pennsylvania Press.
- C. Gordon, A. Leuski, G. Benn, E. Klassen, E. Fast, M. Liewer, A. Hartholt, and D. Traum. Mar. 2019. PRIMER: An Emotionally Aware Virtual Agent. In *Proceedings of the 24th International Conference* on Intelligent User Interfaces, p. 10. ACM, Los Angeles, CA. https://www.research.ibm.com/haifa/ Workshops/user2agent2019/.
- A. C. Graesser, Z. Cai, B. Morgan, and L. Wang. 2017. Assessment with computer agents that engage in conversational dialogues and trialogues with learners. *Computers in Human Behavior*, 76: 607–616.
- J. Gratch and G. Lucas. 2021. Rapport between humans and socially interactive agents. In *The Handbook on Socially Interactive Agents: 20 Years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 1: Methods, Behavior, Cognition*, p. 433–462. Association for Computing Machinery, New York, NY, USA, 1. ISBN 9781450387200. https://doi.org/10.1145/3477322.3477335.
- J. Gratch and S. Marsella. 2004. A domain-independent framework for modeling emotion. *Journal of Cognitive Systems Research*.
- H. P. Grice. 1957. Meaning. Philosophical Review, 66: 377-88.
- J. P. Grice. 1975. Logic and conversation. In P. Cole and J. L. Morgan, eds., Syntax and Semantics, volume 3: Speech Acts, pp. 41–58. Academic Press.
- J. Grolleman, B. van Dijk, A. Nijholt, and A. van Emst. 2006. Break the habit! designing an etherapy intervention using a virtual coach in aid of smoking cessation. In *International Conference* on *Persuasive Technology*, pp. 133–141. Springer.
- A. Hartholt and S. Mozgai. 2022. Platforms and tools for sia research and development. In *The Handbook on Socially Interactive Agents: 20 Years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 2: Interactivity, Platforms, Application, p. 261–304. Association for Computing Machinery, New York, NY, USA, 1. ISBN 9781450398961. https://doi.org/10.1145/3563659.3563668.*
- A. Hartholt, D. Traum, S. C. Marsella, A. Shapiro, G. Stratou, A. Leuski, L.-P. Morency, and J. Gratch. Aug. 2013. All together now: Introducing the Virtual Human toolkit. In *International Conference on Intelligent Virtual Humans*. Edinburgh, UK.
- R. Hill, Jr., J. Gratch, S. Marsella, J. Rickel, W. Swartout, and D. Traum. 2003. Virtual humans in the mission rehearsal exercise system. *In KI Embodied Conversational Agents*, 17: 32–38.
- K. S. Hone and R. Graham. 2000. Towards a tool for the subjective assessment of speech system interfaces (sassi). *Natural Language Engineering: Special Issue on Best Practice in Spoken Dialogue*

#### Systems.

- L. Huang, Z. Ye, J. Qin, L. Lin, and X. Liang. 2020. Grade: Automatic graph-enhanced coherence metric for evaluating open-domain dialogue systems. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 9230–9240.
- D. Jan and D. Traum. 2007. Dynamic movement and positioning of embodied agents in multiparty conversations. In *In proceedings of AAMAS 2007: Sixth International Joint Conference on Autonomous Agents and Multi-Agent Systems*, pp. 59–66.
- D. Jan, D. Herrera, B. Martinovski, D. G. Novick, and D. R. Traum. 2007. A computational model of culture-specific conversational behavior. In C. Pelachaud, J.-C. Martin, E. André, G. Chollet, K. Karpouzis, and D. Pelé, eds., *IVA*, volume 4722 of *Lecture Notes in Computer Science*, pp. 45–56. Springer. ISBN 978-3-540-74996-7.
- K. Janowski, H. Ritschel, and E. André. 2022. Adaptive artificial personalities. In *The Handbook on Socially Interactive Agents: 20 Years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 2: Interactivity, Platforms, Application*, p. 155–194. Association for Computing Machinery, New York, NY, USA, 1. ISBN 9781450398961. https://doi.org/10.1145/3563659.3563666.
- H. Kamezawa, N. Nishida, N. Shimizu, T. Miyazaki, and H. Nakayama. 2020. A visually-grounded firstperson dialogue dataset with verbal and non-verbal responses. In *Proceedings of the 2020 Conference* on Empirical Methods in Natural Language Processing (EMNLP), pp. 3299–3310.
- A. Koenecke, A. Nam, E. Lake, J. Nudell, M. Quartey, Z. Mengesha, C. Toups, J. R. Rickford, D. Jurafsky, and S. Goel. 2020. Racial disparities in automated speech recognition. *Proceedings* of the National Academy of Sciences, 117(14): 7684–7689.
- S. Kopp and T. Hassan. 2022. The fabric of socially interactive agents: Multimodal interaction architectures. In *The Handbook on Socially Interactive Agents: 20 Years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 2: Interactivity, Platforms, Application*, p. 77–112. Association for Computing Machinery, New York, NY, USA, 1. ISBN 9781450398961. https://doi.org/10.1145/3563659.3563664.
- S. Kopp, J. Allwood, K. Grammer, E. Ahlsén, and T. Stocksmeier. 2008. Modeling embodied feedback with virtual humans. In *Modeling communication with robots and virtual humans*, pp. 18–37. Springer.
- N. Krämer and A. Manzeschke. 2021. Social reactions to socially interactive agents and their ethical implications. In *The Handbook on Socially Interactive Agents: 20 Years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 1: Methods, Behavior, Cognition,* p. 77–104. Association for Computing Machinery, New York, NY, USA, 1. ISBN 9781450387200. https://doi.org/10.1145/3477322.3477326.
- J. E. Laird, A. Newell, and P. S. Rosenbloom. Sept. 1987. SOAR: an architecture for general intelligence. *Artificial Intelligence*, 33(1): 1–64.
- H. C. Lane and N. L. Schroeder. 2022. Pedagogical agents. In *The Handbook on Socially Interactive Agents: 20 Years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 2: Interactivity, Platforms, Application, p. 307–330. Association for Computing Machinery, New York, NY, USA, 1. ISBN 9781450398961. https://doi.org/10.1145/3563659. 3563669.*

- J. Lee, S. Marsella, D. Traum, J. Gratch, and B. Lance. 2007. The rickel gaze model: A window on the mind of a virtual human. In *International workshop on intelligent virtual agents*, pp. 296–303. Springer.
- J. Lee, D. DeVault, S. Marsella, and D. Traum. 2008. Thoughts on fml: Behavior generation in the virtual human communication architecture. In *Proceedings of FML 2008: The First Functional Markup Language Workshop*. AAMAS.
- A. Leuski and D. Traum. 2011. Npceditor: Creating virtual human dialogue using information retrieval techniques. Ai Magazine, 32(2): 42–56.
- D. K. Lewis. 1969. Convention: A Philosophical Study. Harvard University Press.
- C.-Y. Lin. 2004. Rouge: A package for automatic evaluation of summaries. In *Text summarization branches out*, pp. 74–81.
- P. Lison and C. Kennington. 2016. Opendial: A toolkit for developing spoken dialogue systems with probabilistic rules. In *Proceedings of ACL-2016 system demonstrations*, pp. 67–72.
- C.-W. Liu, R. Lowe, I. V. Serban, M. Noseworthy, L. Charlin, and J. Pineau. 2016. How not to evaluate your dialogue system: An empirical study of unsupervised evaluation metrics for dialogue response generation. arXiv preprint arXiv:1603.08023.
- B. Lugrin and M. Rehm. 2021. Culture for socially interactive agents. In *The Handbook on Socially Interactive Agents: 20 Years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 1: Methods, Behavior, Cognition*, p. 463–494. Association for Computing Machinery, New York, NY, USA, 1. ISBN 9781450387200. https://doi.org/10.1145/3477322.3477336.
- S. M. Lukin, F. Gervits, C. J. Hayes, A. Leuski, P. Moolchandani, J. G. Rogers, III, C. Sanchez Amaro, M. Marge, C. R. Voss, and D. Traum. July 2018. ScoutBot: A Dialogue System for Collaborative Navigation. In *Proceedings of ACL 2018, System Demonstrations*, pp. 93–98. Melbourne, Australia. https://www.aclweb.org/anthology/P18-4016.
- F. Mairesse and M. Walker. 2007. Personage: Personality generation for dialogue. In *Proceedings of the 45th annual meeting of the association of computational linguistics*, pp. 496–503.
- R. Manuvinakurike, T. Bui, W. Chang, and K. Georgila. 2018. Conversational image editing: Incremental intent identification in a new dialogue task. In *Proceedings of the 19th Annual SIGdial Meeting* on Discourse and Dialogue, pp. 284–295.
- M. Marge, C. Espy-Wilson, N. G. Ward, A. Alwan, Y. Artzi, M. Bansal, G. Blankenship, J. Chai, H. Daumé III, D. Dey, M. Harper, T. Howard, C. Kennington, I. Kruijff-Korbayová, D. Manocha, C. Matuszek, R. Mead, R. Mooney, R. K. Moore, M. Ostendorf, H. Pon-Barry, A. I. Rudnicky, M. Scheutz, R. S. Amant, T. Sun, S. Tellex, D. Traum, and Z. Yu. 2022. Spoken language interaction with robots: Recommendations for future research. *Computer Speech & Language*, 71: 101255.
- S. Marsella and J. Gratch. July 2003. Modeling coping behavior in virtual humans: Don't worry, be happy. In *In proceedings of AAMAS 2003: Second International Joint Conference on Autonomous Agents and Multi-Agent Systems*.
- S. Marsella and J. Gratch. 2009. Ema: A process model of appraisal dynamics. *Journal of Cognitive Systems Research*, 10(1): 70–90.
- J. D. McCawley. 1999. Participant roles, frames, and speech acts. *Linguistics and Philosophy*, pp. 595–619.

- S. Mehri and M. Eskenazi. 2020. Unsupervised evaluation of interactive dialog with dialogpt. In Proceedings of the 21th Annual Meeting of the Special Interest Group on Discourse and Dialogue, pp. 225–235.
- T. Muller, A. Hartholt, S. Marsella, J. Gratch, and D. Traum. 2004. Do you want to talk about it? In *Tutorial and Research Workshop on Affective Dialogue Systems*, pp. 188–192. Springer.
- B. Mutlu, T. Shiwa, T. Kanda, H. Ishiguro, and N. Hagita. 2009. Footing in human-robot conversations: how robots might shape participant roles using gaze cues. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, pp. 61–68.
- J. Nadel, O. Grynszpan, and J.-C. Martin. 2022. Autism and socially interactive agents. In *The Handbook on Socially Interactive Agents: 20 Years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 2: Interactivity, Platforms, Application*, p. 437–462. Association for Computing Machinery, New York, NY, USA, 1. ISBN 9781450398961. https://doi.org/10.1145/3563659.3563673.
- S. Nasihati Gilani, D. Traum, A. Merla, E. Hee, Z. Walker, B. Manini, G. Gallagher, and L.-A. Petitto. 2018. Multimodal dialogue management for multiparty interaction with infants. In *Proceedings of the 20th ACM International Conference on Multimodal Interaction*, pp. 5–13.
- F. Nie, J.-G. Yao, J. Wang, R. Pan, and C.-Y. Lin. July 2019. A simple recipe towards reducing hallucination in neural surface realisation. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 2673–2679. Association for Computational Linguistics, Florence, Italy. https://aclanthology.org/P19-1256. DOI: 10.18653/v1/P19-1256.
- D. Novick. 1988. Control of Mixed-Initiative Discourse Through Meta-Locutionary Acts: A Computational Model. PhD thesis, University of Oregon. also available as U. Oregon Computer and Information Science Tech Report CIS-TR-88-18.
- E. Padilha and J. Carletta. 2002. A simulation of small group discussion. *Proceedings of EDILOG 2002: Sixth Workshop on the Semantics and Pragmatics of Dialogue*, pp. 117–124. citeseer.ist.psu. edu/padilha02simulation.html.
- A. Paiva, F. Correia, R. Oliveira, F. Santos, and P. Arriaga. 2021. Empathy and prosociality in social agents. In *The Handbook on Socially Interactive Agents: 20 Years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 1: Methods, Behavior, Cognition*, p. 385–432. Association for Computing Machinery, New York, NY, USA, 1. ISBN 9781450387200. https://doi.org/10.1145/3477322.3477334.
- K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pp. 311–318.
- C. Pelachaud, C. Busso, and D. Heylen. 2021. Multimodal behavior modeling for socially interactive agents. In *The Handbook on Socially Interactive Agents: 20 Years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 1: Methods, Behavior, Cognition*, p. 259–310. Association for Computing Machinery, New York, NY, USA, 1. ISBN 9781450387200. https://doi.org/10.1145/3477322.3477331.
- S. Pezzelle and R. Fernández. 2019. Is the red square big? malevic: Modeling adjectives leveraging visual contexts. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language*

Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pp. 2865–2876.

- R. Pieraccini. 2021. Natural language understanding in socially interactive agents. In *The Handbook on Socially Interactive Agents: 20 Years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 1: Methods, Behavior, Cognition*, p. 147–172. Association for Computing Machinery, New York, NY, USA, 1. ISBN 9781450387200. https://doi.org/10.1145/3477322.3477328.
- E. Pincus. 2020. An investigation of fully interactive multi-role dialogue agents. PhD thesis, University of Southern California.
- B. Plüss, D. Devault, and D. Traum. 2011. Toward rapid development of multiparty virtual human negotiation scenarios. In Proc. 15th Workshop Semantics and Pragmatics of Dialogue (SemDial, pp. 63–72.
- R. Prada and D. Rato. 2022. Socially interactive agents in games. In *The Handbook on Socially Interactive Agents: 20 Years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 2: Interactivity, Platforms, Application*, p. 493–526. Association for Computing Machinery, New York, NY, USA, 1. ISBN 9781450398961. https://doi.org/10.1145/3563659.3563675.
- M. Quigley, B. Gerkey, K. Conley, J. Faust, T. Foote, J. Leibs, E. Berger, R. Wheeler, and A. Ng. 2009. Ros: an open-source robot operating system. In *ICRA 2009 Workshop on Open Source Software*.
- N. L. Reinsch, J. W. Turner, and C. H. Tinsley. 2008. Multicommunicating: A practice whose time has come? Academy of Management Review, 33(2): 391–403.
- E. Reiter and R. Dale. 1997. Building applied natural language generation systems. *Natural Language Engineering*, 3(1): 57–87. DOI: 10.1017/S1351324997001502.
- J. Rickel and W. L. Johnson. 1999. Animated agents for procedural training in virtual reality: Perception, cognition, and motor control. *Applied Artificial Intelligence*, 13: 343–382.
- J. Rickel, S. Marsella, J. Gratch, R. Hill, D. Traum, and W. Swartout. 2002. Toward a new generation of virtual humans for interactive experiences. *IEEE Intelligent Systems*, 17: 32–38.
- A. Rizzo, E. Forbell, B. Lange, J. Galen Buckwalter, J. Williams, K. Sagae, and D. Traum. 2013. Simcoach: an online intelligent virtual human agent system for breaking down barriers to care for service members and veterans. In R. Scurfield and K. Platoni, eds., *Healing War Trauma*, Psychological Stress Series, pp. 238–250. Taylor & Francis, United Kingdom. ISBN 9780415807050. DOI: 10.4324/9780203153819.
- S. Robinson, D. Traum, M. Ittycheriah, and J. Henderer. 2008. What would you ask a conversational agent? observations of human-agent dialogues in a museum setting. In *Proceedings of the sixth international conference on language resources and evaluation (LREC'08).*
- A. Rosenthal-von der Pütten and A. M. H. Abrams. 2021. Empirical Methods in the Social Science for Researching Socially Interactive Agents, p. 19–76. Association for Computing Machinery, New York, NY, USA, 1. ISBN 9781450387200. https://doi.org/10.1145/3477322.3477325.
- D. K. Roy. 2002. Learning visually grounded words and syntax for a scene description task. *Computer speech & language*, 16(3-4): 353–385.
- H. Sacks, E. A. Schegloff, and G. Jefferson. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language*, 50: 696–735.

- C. Saund and S. Marsella. 2021. Gesture generation. In *The Handbook on Socially Interactive Agents: 20 Years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 1: Methods, Behavior, Cognition*, p. 213–258. Association for Computing Machinery, New York, NY, USA, 1. ISBN 9781450387200. https://doi.org/10.1145/3477322. 3477330.
- B. Scassellati, J. Brawer, K. Tsui, S. Nasihati Gilani, M. Malzkuhn, B. Manini, A. Stone, G. Kartheiser, A. Merla, A. Shapiro, et al. 2018. Teaching language to deaf infants with a robot and a virtual human. In *Proceedings of the 2018 CHI Conference on human Factors in computing systems*, pp. 1–13.
- E. A. Schegloff. 1968. Sequencing in conversational openings. American Anthropologist, pp. 1075– 1095.
- E. A. Schegloff and H. Sacks. 1973. Opening up closings. Semiotica, 7: 289-327.
- S. R. Schiffer. 1972. Meaning. Oxford University Press.
- J. R. Searle. 1969. Speech Acts. Cambridge University Press, New York.
- J. R. Searle. 1976. A classification of illocutionary acts. Language in Society, 5: 1-23.
- J. R. Searle and D. Vanderveken. 1985. *Foundations of Illocutionary Logic*. Cambridge University Press.
- M. Shridhar, J. Thomason, D. Gordon, Y. Bisk, W. Han, R. Mottaghi, L. Zettlemoyer, and D. Fox. 2020. Alfred: A benchmark for interpreting grounded instructions for everyday tasks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10740–10749.
- J. M. Sinclair and R. M. Coulthard. 1975. *Towards an analysis of Discourse: The English used by teachers and pupils*. Oxford University Press.
- L. Steels. 1998. The origins of syntax in visually grounded robotic agents. *Artificial Intelligence*, 103(1): 133–156. ISSN 0004-3702. https://www.sciencedirect.com/science/article/pii/S0004370298000666. DOI: https://doi.org/10.1016/S0004-3702(98)00066-6. Artificial Intelligence 40 years later.
- D. Tadimeti, K. Georgila, and D. Traum. November 2021. How well can an agent understand different accents? In 5th Widening NLP (WiNLP2021) workshop, co-located with EMNLP.
- D. Traum. 2003. Semantics and pragmatics of questions and answers for dialogue agents. In proceedings of the International Workshop on Computational Semantics, pp. 380–394.
- D. Traum. 2004. Issues in multi-party dialogues. In F. Dignum, ed., Advances in Agent Communication, LNAI 2922, pp. 201–211. Springer Verlag.
- D. Traum, J. Bos, R. Cooper, S. Larsson, I. Lewin, C. Matheson, and M. Poesio. 1999. A model of dialogue moves and information state revision. Technical Report Deliverable D2.1, Trindi.
- D. Traum, J. Rickel, S. Marsella, and J. Gratch. July 2003. Negotiation over tasks in hybrid humanagent teams for simulation-based training. In *Proceedings of AAMAS 2003: Second International Joint Conference on Autonomous Agents and Multi-Agent Systems*, pp. 441–448.
- D. Traum, S. Marsella, and J. Gratch. 2004. Emotion and dialogue in the mre virtual humans. In *Tutorial and Research Workshop on Affective Dialogue Systems*, pp. 117–127. Springer.
- D. Traum, W. Swartout, S. Marsella, and J. Gratch. September 2005. Fight, flight, or negotiate: Believable strategies for conversing under crisis. In *proceedings of the Intelligent Virtual Agents Conference (IVA)*, pp. 52–64. Springer-Verlag Lecture Notes in Computer Science.

- D. Traum, S. C. Marsella, J. Gratch, J. Lee, and A. Hartholt. 2008a. Multi-party, multi-issue, multistrategy negotiation for multi-modal virtual agents. In *International Workshop on Intelligent Virtual Agents*, pp. 117–130. Springer.
- D. Traum, P. Aggarwal, R. Artstein, S. Foutz, J. Gerten, A. Katsamanis, A. Leuski, D. Noren, and W. Swartout. 2012. Ada and grace: Direct interaction with museum visitors. In *International conference on intelligent virtual agents*, pp. 245–251. Springer.
- D. Traum, C. Henry, S. Lukin, R. Artstein, F. Gervits, K. Pollard, C. Bonial, S. Lei, C. Voss, M. Marge, C. Hayes, and S. Hill. May 7-12, 2018 2018. Dialogue structure annotation for multi-floor interaction. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation* (*LREC 2018*), pp. 104–111. European Language Resources Association (ELRA), Miyazaki, Japan. ISBN 979-10-95546-00-9.
- D. R. Traum. 2002. Ideas on multi-layer dialogue management for multi-party, multi-conversation, multi-modal communication. In *Computational Linguistics in the Netherlands 2001*, pp. 1–7. Brill.
- D. R. Traum and J. F. Allen. 1994. Discourse obligations in dialogue processing. In *Proceedings of the* 32<sup>nd</sup> Annual Meeting of the Association for Computational Linguistics, pp. 1–8.
- D. R. Traum and E. A. Hinkelman. August 1991. Beyond literal meaning: Conversation acts. In Proceedings of the IJCAI Workshop on Computational Approaches to Non-Literal Language: Metaphor, Metonymy, Idiom, Speech Acts, Implicature, pp. 132–140.
- D. R. Traum, A. Leuski, A. Roque, S. Gandhe, D. DeVault, J. Gerten, S. Robinson, and B. Martinovski. 12/2008 2008b. Natural language dialogue architectures for tactical questioning characters. In *Army Science Conference*. Florida.
- S. Turkle. 2017. *Alone together: Why we expect more from technology and less from each other.* Hachette UK.
- S. Ultes, L. M. R. Barahona, P.-H. Su, D. Vandyke, D. Kim, I. Casanueva, P. Budzianowski, N. Mrkšić, T.-H. Wen, M. Gasic, et al. 2017. Pydial: A multi-domain statistical dialogue system toolkit. In *Proceedings of ACL 2017, System Demonstrations*, pp. 73–78.
- H. H. Vilhjálmsson. 2022. Interaction in social space. In *The Handbook on Socially Interactive Agents: 20 Years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 2: Interactivity, Platforms, Application*, p. 3–44. Association for Computing Machinery, New York, NY, USA, 1. ISBN 9781450398961. https://doi.org/10.1145/3563659. 3563662.
- E. Weigand. 2021. Language and dialogue in philosophy and science:. *Intercultural Pragmatics*, 18(4): 533–561. https://doi.org/10.1515/ip-2021-4005. DOI: doi:10.1515/ip-2021-4005.
- Y.-T. Yeh, M. Eskenazi, and S. Mehri. 2021. A comprehensive assessment of dialog evaluation metrics. arXiv preprint arXiv:2106.03706.
- Y. Yu, A. Eshghi, and O. Lemon. 2017. Voila: An optimised dialogue system for interactively learning visually-grounded word meanings (demonstration system). In *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue*, pp. 197–200.
- G. Zhao, S. Sonsaat, A. O. Silpachai, I. Lucic, E. Chukharev-Hudilainen, J. Levis, and R. Gutierrez-Osuna. 2018. L2-arctic: A non-native english speech corpus. *Perception Sensing Instrumentation Lab.*