Conversational UX Design

Robert J. Moore
IBM Research-Almaden
San Jose, CA 95120, USA
rjmoore@us.ibm.com

Margaret H. Szymanski
IBM Research-Almaden
San Jose, CA 95120, USA
Margaret.Szymanski@ibm.com

Raphael Arar
IBM Research-Almaden
San Jose, CA 95120, USA
rarar@us.ibm.com

Guang-Jie Ren
IBM Research-Almaden
San Jose, CA 95120, USA
gren@us.ibm.com

CHI'17 Extended Abstracts, May 06-11, 2017, Denver, CO, USA
ACM 978-1-4503-4656-6/17/05.
http://dx.doi.org/10.1145/3027063.3027077

Abstract
From Siri to Alexa to Cortana, conversational interfaces are hitting the mainstream and becoming ubiquitous in our daily lives. However, user experiences with such applications remain disappointing. Although it is easy to get a system to produce words, none of the current agents or bots display general conversational competence. Modeling natural conversation is still a hard problem. But in order to tackle it, conversational UX designers must possess a technical understanding of the structures of natural conversation. For this, UX designers can turn to the literature in Conversation Analysis, which provides natural patterns of how people talk. In applying Conversation Analysis to Conversational UX Design, we outline some design principles and guidelines.

Author Keywords
Conversational systems; virtual agents; chat bots; UX design; interaction design; conversation analysis; ethnomethodology

ACM Classification Keywords
H.5.2. Information interfaces and presentation (e.g., HCI): User Interfaces.

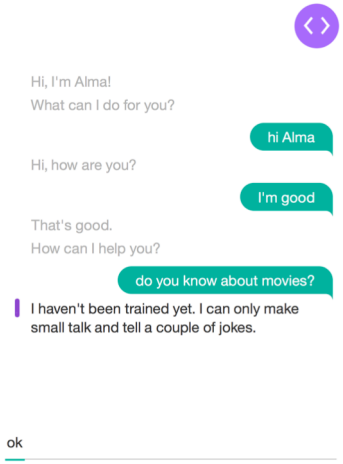

Figure 1: Example of a simple conversational interface

Background

While virtual agents and chat bots have been around for decades [1], there has been a recent resurgence of interest in them as major computer companies have released their own. Apple's Siri, Google's Assistant, Microsoft's Cortana, Amazon's Alexa, Facebook's M and IBM's Watson are just a few examples, not counting conversational agents by startups.

With persistent Internet connections and statistical algorithms, virtual agents are much smarter today than they were 20 years ago. While most of these systems accept voice input from users, a growing number accept text input, sometimes from standard applications like SMS and Instant Messaging. But although today's virtual agents are often touted as "easy to build," interactions with them are still awkward, confusing, limited and fraught with troubles in mutual understanding.

Conversational interfaces are very different from graphical user interfaces. In conversational interfaces, the graphical elements are generally minimal, for example, a chat history and text box or microphone button (Figure 1) or nothing at all. User interaction is conducted primarily through the words: typed or spoken. The interaction metaphor for these interfaces is the natural, human conversation.

Although natural language processing has given us powerful, automated tools for analyzing the spoken and the written word alike, it does not provide a model of how bits of language are sequenced by multiple parties into an interaction that is recognizable as a "conversation" [3]. Natural human conversation is a complex system [9,3], which Harvey Sacks called a "machinery" in its own right [6]. How to create a user interface that mimics features of such a machinery is nontrivial. But rather than avoiding this complexity by producing simplistic interactions, we should embrace the complexity of conversational systems because it "mirrors the complexity of the world," while at the same time avoiding any complexity that is due instead to "poor design" [4].

The time is ripe for developing Conversational UX Design as a distinct discipline. Just as graphical user interfaces improved dramatically as visual artists became involved in development (1990s-2000s), so will conversational interfaces when conversation experts get involved. Rather than a background in the visual arts, conversation experts possess a background in the study natural conversation, for example, in the fields of sociology, communication, linguistics, psychology, etc. Conversation Analysis, in particular, offers over 50 years worth of rich, empirical studies of naturally occurring talk-in-interaction in a wide range of settings and languages, which offer formal, qualitative models of how natural conversation is structured. While the proposal to apply these findings to the design of dialogue interfaces may not be entirely new [1,5], it has become especially timely as new conversational technology platforms are becoming ubiquitous.

Conversation experts are keen observers of natural conversation and can articulate the mechanics of human conversation, which most others know only tacitly. For example, a conversation expert may describe the function of the word "oh" to mark speakers' realizations [2] or how the phrase, "to the what?," in response to "I'm going to the workshop," elegantly elicits a repeat of a single word "workshop"

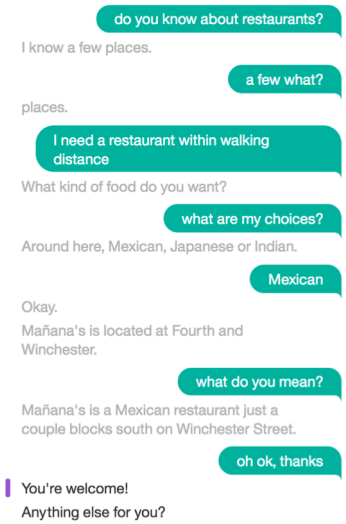

Figure 2: Base pair with pre-, insert and post expansions.

[10]. Conversational UX designers use such observations of the machinery of human conversation in building conversational machines.

This position paper explores the intersection of conversational UX design, of both text- or voice-based virtual agents, and the analysis of naturally occurring human conversation (e.g., the Conversation Analysis literature). In it we outline a set of design principles and guidelines for today's conversational platforms.

Principles

The design of conversational user experiences should reflect the design principles that speakers in natural conversation use themselves. Work in Conversation Analysis has revealed three of these:

Recipient Design

In natural conversation, speakers tailor their talk, spoken or written, to their particular recipients in multiple ways, such as, adapting to their perceived level of knowledge [7, 8]. For example, they choose different topics and levels of detail depending on what they believe the other person knows and will recognize. When designing for conversation, don't force the user down a single path of your conversation flow. Instead adapt to how *this* user proceeds and to what *this* user knows.

Minimization

Speakers in natural conversation design their talk to be efficient, using the least number of words or requiring the least amount of effort, on the part of the particular recipient, to understand [7]. Unnecessary details may make the point of the speaker's action harder to grasp. Reading and especially listening to a conversational agent's responses takes time and effort. Strive to minimize these for the user without sacrificing understandability.

Repair

Speakers tend to use just enough words to enable the recipient to understand, and then they wait to see if it works. If it doesn't, the speaker relaxes the concern for minimization to repair the misunderstanding, for example by paraphrasing or elaborating [7, 10]. If you build robust repair mechanisms, then your conversational agent does not always need to get it right on the first attempt.

This mechanic of recipient design, minimization and repair thereby provide for efficiency in natural conversation. Conversational UX design should incorporate this mechanic.

Building Blocks

The building blocks of most natural conversation are pairs of actions. Conversation analysts call these "adjacency pairs" [8, 9], and they include common pairs such as question-answer, greeting-greeting, request-grant/deny, offer-accept/reject, complaint-remedy/excuse, and more.

These action pairs are like an accordion: they can occur in a compact two turns, or they can be expanded to arbitrary lengths with additional dependent pairs, or "expansions." Figure 2 offers a request-grant sequence: "I need a restaurant within walking distance" and "Okay. Mañana's is located at Fourth and Winchester." Although this 2-turn sequence could occur alone, we can see examples of the kinds of expansions that can occur. These expansions are dependent on the base

pair and are critical resources for enabling speakers to do things like manage dependencies (checking the other's knowledge or role, eliciting required details), hearing troubles (repeat of "places"), understanding troubles and closings (elaboration of answer) [9].

With adjacency pair expansions, speakers can manage the conversation itself and achieve mutual understanding, to a practical degree.

Conversation Types

Different types of conversations have different characteristics. The following conversation types describe the ways we communicate with one another. Knowing the type of conversation you plan to have with your user will help you craft dialogue that lends to desirable experiences for your users.

Ordinary

The kind of conversation you have with family, friends and even strangers. Ordinary conversations consist of the broadest range of activities from delivering news to checking up to seeking help or advice to learning to small talk and much more. Sometimes the purpose of ordinary conversation is just simply to open a social connection with another person for its own sake. In conversation analytic theory, ordinary conversation is considered the least constrained type of conversation from which other types are optimized for particular purposes.

Service

The kind of conversation you have with a customer service or other organizational representative. The roles are fixed: One person, such as a customer, member or citizen, requests service; the other person, usually a stranger, provides services on behalf of an organization. Services may consist simply of answering inquiries or taking actions or guiding the other through troubleshooting. Service conversations have distinctive openings. Service providers typically do a greeting, self-identify, identify the organization and offer service, rushing through the transitions, so that the service seeker’s first turn can be their request. Other times such openings may include a series of questions for verifying the service seeker’s identity.

Teaching

The kind of conversation you have with in a classroom or with a tutor. One person (or more) seeks knowledge; the other presents knowledge and tests understanding. In teaching conversations, teachers routinely ask the student questions to which they already know the answers. They may withhold the answers in an attempt to elicit the correct answers from the student. Whereas correcting other people is discouraged in most other kinds of conversations for the sake of politeness [10], it is required in the teaching conversation.

Counseling

The kind of conversation you have with a therapist, counselor or advisor. One person seeks advice; the other listens and provides advice. The counselee may report a problem of a personal nature or a long-term goal and seek advice on how to manage it, rather than requesting that the other person manage it directly. The counselor tends to withhold judgment and let the counselee lead the conversation without interrupting or changing the topic.

Other

There are, of course, many other recognizable types of conversations. When building a conversational user experience, identify what type of conversation you want to create and then think about, or examine, how people talk in other instances of that type. For example, if you're building a virtual agent that recommends travel destinations, model your conversation flow on how human travel agents or customer service representatives speak.

Best Practices

Create agents that converse, inform and delight!

Onboard Users

Conversational agents should always be able to talk about what they can do or what they know. They should be able to handle preliminary questions like, "Can you help me find a restaurant?," "Do you sell movie tickets?" or simply "What can you do?"

Progressive Disclosure

Progressive disclosure is an interaction design technique that breaks down information into sequences in order not to overwhelm end users. Provide next steps sequentially and break down a process into bite-sized chunks. This is especially important for voice-based agents where the conversation history must be memorized. For example, if a user asks, "What's a good restaurant to go to?," elicit additional details one at a time: "What kind of food are you in the mood for?," "Within walking distance or a short drive?"

History

Relaying the current state of the conversation to the user will differ depending on the interaction modality. In text-based interactions, consider dividing a chat history into breakpoints. Provide these as conversation landmarks, so that the user can successfully review the history of a conversation. In voice-based interactions, use repetition tactfully to not only provide feedback but also to mark location.

Artifacts

Whenever possible leverage the medium to facilitate the conversation. Just as in human conversation, an image or visual aide can sometimes complement the utterances. For example, an image of an object can establish a referent when the name is not known, or a map with an X can better relay complicated instructions than words.

Multimodal Feedback

Give the user feedback through the conversation and the visual user interface (if present) to illustrate whether or not a request was heard or an action took place. For example, a simple "One moment please" and a spinning wheel can show the user that the system is working on a request.

Fail Gracefully

Don't be afraid to let the conversational agent admit a lack of understanding. Sometimes humans don't understand each other either. Display what the agent does understand so the user can better diagnose and repair the trouble. For example, "Do you mean, can I recommend a restaurant?" or "Do you want attractions in Rome or flights to Rome?"

Personality

Construct the persona of your agent somewhat like you would for your user. How serious or professional do you

want your agent to be? Enable your agent to employ humor and emotion in ways that are consistent with its persona. For example, your agent might give a humorous response to "How are you doing?" or an emotional one to "You're not very smart."

Conclusion

We see Conversational UX Design as a distinctive, emerging discipline that combines interaction design with a formal understanding of human conversation. To seed this discipline, we have attempted to apply principles (e.g., recipient design) and models (e.g., adjacency pairs) from the Conversation Analysis literature to the design and building of chatflows. As this discipline matures, we expect to see the development of standards for conversation flow design, just as we see standards in the more mature discipline of automatic speech recognition (e.g., VoiceXML or Speech Synthesis Markup Language).

References

1. David M. Frohlich and Paul Luff. 1990. Applying the technology of conversation to the technology for conversation, in P. Luff, G.N. Gilbert and D.M. Frohlich (eds), *Computers and Conversation*. London: Academic Press, 187–220.
2. John Heritage. 1984. A change-of-state token and aspects of its sequential placement, in J.M. Atkinson and J.C. Heritage (eds), *Structures in Social Action: Studies in Conversation Analysis*. Cambridge: Cambridge University Press, 299–345.
3. Robert J. Moore, Rafah A. Hosn and Ashima Arora. 2016. The Machinery of Natural Conversation and the Design of Conversational Machines. *American Sociological Association annual meeting*. Seattle (August 2016).
4. Donald A. Norman. 2011. *Living with complexity*, Cambridge, MA: MIT Press.
5. Michael Norman and Peter Thomas. 1990. The Very Idea: Informing HCI Design from Conversation Analysis, in P. Luff, G.N. Gilbert and D.M. Frohlich (eds), *Computers and Conversation*. London: Academic Press, 51-66.
6. Harvey Sacks. 1984. Notes on methodology, in J.M. Atkinson and J.C. Heritage (eds), *Structures in Social Action: Studies in Conversation Analysis*. Cambridge: Cambridge University Press, 21–27.
7. Sacks, Harvey. and Schegloff, Emanuel. A. 1979. Two preferences in the organization of reference to persons in conversation and their interaction. In G. Psathas (ed.), *Everyday language: Studies in ethnomethodology*, 15–21. New York: Irvington.
8. Sacks, Harvey, Emanuel A. Schegloff and Gail Jefferson. 1974. A simplest systematics for the organization of turn-taking for conversation, *Language* 50: 696–735.
9. Emanuel A. Schegloff. 2007. *Sequence Organization in Interaction: A Primer in Conversation Analysis*, vol 1. Cambridge: Cambridge University Press.
10. Emanuel A. Schegloff, Gail Jefferson and Harvey Sacks. 1977. The preference for self-correction in the organization of repair in conversation, *Language*, 53(2): 361–382.