

Dialogue and Conversational Agents

Ling575
Spoken Dialog Systems
March 29, 2017

Roadmap

- Dialog and Dialog Systems
- Facets of Conversation:
 - Turn-taking
 - Speech Acts
 - Cooperativity
 - Grounding
- Spoken Dialogue Systems:
 - Pipeline Architecture
 - Finite-State & Frame-based Models
 - Evaluation

Dialogue is Different

The background of the slide features a series of soft, horizontal, wavy lines in various shades of blue and white, creating a sense of depth and movement. The lines are more pronounced in the lower half of the image, where they form a series of overlapping, undulating bands that resemble a stylized horizon or a series of gentle hills. The upper half of the image is a lighter, more uniform blue, providing a clean backdrop for the title text.

Dialogue is Different

- Two or more speakers
 - Primary focus on speech

Dialogue is Different

- Two or more speakers
 - Primary focus on speech
- Issues in multi-party spoken dialogue

Dialogue is Different

- Two or more speakers
 - Primary focus on speech
- Issues in multi-party spoken dialogue
 - Turn-taking – who speaks next, when?
 - Collaboration – clarification, feedback,...
 - Disfluencies
 - Adjacency pairs, dialogue acts

Conversations and Conversational Agents

- Conversation:
 - First and often most common form of language use
 - Context of language learning and use

Conversations and Conversational Agents

- Conversation:
 - First and often most common form of language use
 - Context of language learning and use
 - Goal:
 - Describe, characterize spoken interaction
 - Enable automatic recognition, understanding

Conversations and Conversational Agents

- Conversation:
 - First and often most common form of language use
 - Context of language learning and use
 - Goal:
 - Describe, characterize spoken interaction
 - Enable automatic recognition, understanding
- Conversational agents:
 - Spoken dialog systems, spoken language systems
 - Interact with users through speech

Why Spoken Dialog Systems?

Why Spoken Dialog Systems?

- Hands-free operation is needed
- Speech easier than typing
- Replace human agents

Why Spoken Dialog Systems?

- Hands-free operation is needed
 - In-car systems
 - In-field (in space!) assistants
 - Medical systems
 - Immersive training
- Speech easier than typing
 - Small form factor devices
 - Voice assistants (Siri, Alexa,..)
- Replace human agents
 - Call centers
 - Call routing

Why is it hard?

Why is it hard?

- Conversation is complicated
- Integrates multiple technologies

Commercial Systems: IVR

- Interactive voice response (IVR)
 - Designed to replace
 - human customer service agents
 - and improve DTMF phone trees: “Press or say ‘one’”
 - Available mid-90s onward
 - Many companies: Nuance, Tellme (MS), Aspect, etc
 - Multi-turn but inflexible interaction
 - Examples: Directory assistance, United Airlines, Verizon, Sears, etc

AT&T's How May I Help You?

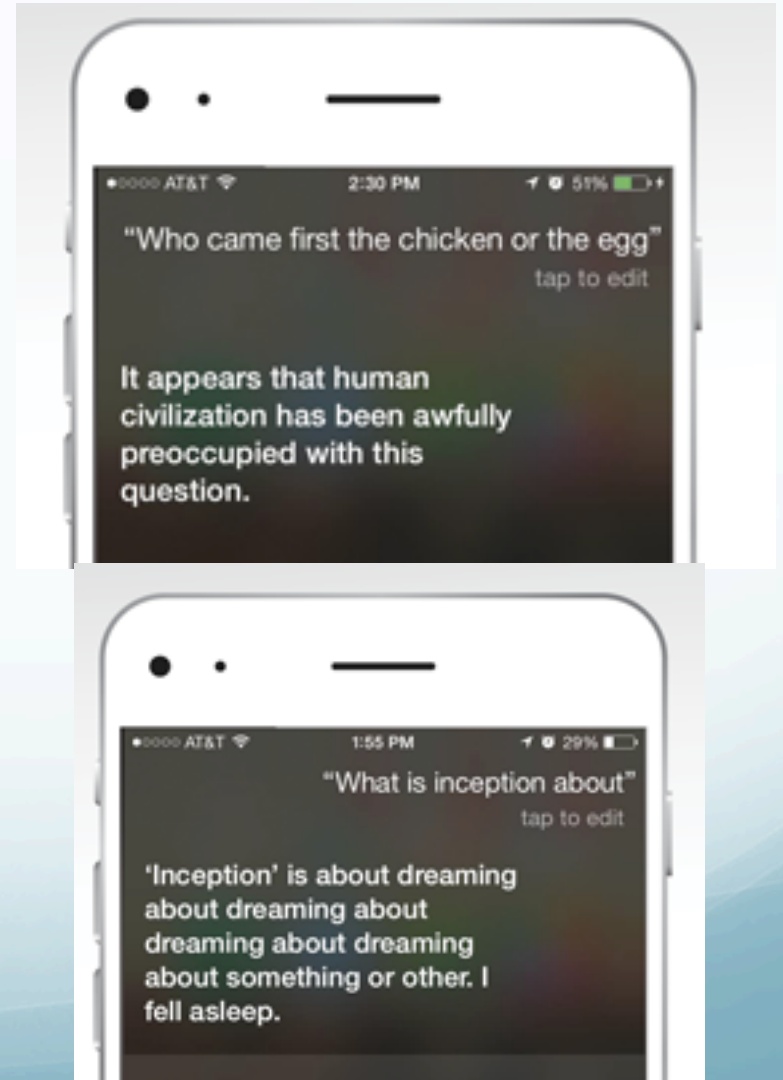
S: How may I help you?
U: *can you tell me how much it is to tokyo?*
S: You want to know the cost of a call?
U: *yes that's right*
S: Please hold on for rate information.

S: How may I help you?
U: *what's the area code for chicago?*
S: You want an area code.
U: *(silence)*
S: The area code for Chicago is 312.

Figure 19.2 The call routing/telephone information domain: two actual conversations between a user (U) and the AT&T How May I Help You (HMIHY) system (S) (Gorin et al., 1997).

Commercial Systems: Virtual Assistants

- Leading example: Siri™
- Spun out of DARPA project
- Ties to specific applications
 - Reminders,
- Backs off to search
- Lots of “Easter eggs”



Commercial Systems: Virtual Assistants (II)

- Rapidly growing area:
 - Google Now™
 - Microsoft's Cortana™
 - Some mobile and desktop (Windows/iOS)
 - Amazon's Echo™ / Alexa™
 - Extensible “skill” sets
 - Plays music, answers questions, interfaces to IoT
 - Far field microphone recognition
 - Google Home™ / Assistant™

Dialog Example

C₁: ...I need to travel in May.

A₁: And, what day in May did you want to travel?

C₂: OK uh I need to be there for a meeting that's from the 12th to the 15th.

A₂: And you're flying into what city?

C₃: Seattle.

A₃: And what time would you like to leave Pittsburgh?

C₄: Uh hmm I don't think there's many options for non-stop.

A₄: Right. There's three non-stops today.

C₅: What are they?

A₅: The first one departs PGH at 10:00am arrives Seattle at 12:05 their time.
The second flight departs PGH at 5:55pm, arrives Seattle at 8pm. And the
last flight departs PGH at 8:15pm arrives Seattle at 10:28pm.

C₆: OK I'll take the 5ish flight on the night before on the 11th.

A₆: On the 11th? OK. Departing at 5:55pm arrives Seattle at 8pm, U.S. Air
flight 115.

C₇: OK.

Research Systems

- Aim to push state-of-the-art in component technologies, dialog and interaction strategies
- Wide range of tasks
 - Training, tutoring, language teaching
 - Planning and negotiation
 - Museum guides, etc
- Novel components
 - Novel formal models: reinforcement learning, discourse
 - Less-restrictive speech recognition
 - Richer conversational models:
 - turn-taking, back-channels, affect, uncertainty

Travel Planning: Let's Go!

S: Thank you for calling the CMU Let's Go! Bus Information System. This conversation will be recorded for research purposes. If you do not want to be recorded, please hang up now. What can I do for you?

U: I would like to go to the airport tomorrow morning.

S: To the airport. When do you want to arrive?

U: I'd like to arrive at 10:30.

S: Arriving at around 10:30 AM. Where do you want to leave from?

U: I'd like to leave from Carnegie Mellon.

S: From Carnegie Mellon. There is a 28X leaving Forbes Avenue at Morewood at 9:20 AM. It will arrive at Greater Pittsburgh Airport at Departure Stops at 10:15 AM.

U: When's the next bus?

S: There is a 28X leaving Forbes Avenue at Morewood at 9:40 AM. It will arrive at Greater Pittsburgh Airport at Departure Stops at 10:35 AM.

U: I'd like to go from the airport back to CMU on Friday.

S: From the airport to CMU. When would you like to leave?

U: At 6:15 PM.

S: Leaving at around 6:15 PM. There's a 28X leaving Greater Pittsburgh Airport at Departure Stops at 6:26 PM. It will arrive at Forbes Avenue at Morewood at 7:20 PM. U: Thank you very much.

ItSpoke Tutoring System

S: I didn't understand. What force acts on all objects within close proximity to earth?

U: *uh the force of gravity*

S: Fine. Besides the Earth's gravitational force, are there any other forces acting on the person, keys, and elevator after the release?

U: *no*

S: Yep. Thus, keep in mind that after the release, the only force on the person, keys, and elevator is the force of gravity. If gravity is the only force acting on an object, will it be moving or staying still?

U: *it will be moving*

S: Fine. Not only are the person, keys, and elevator moving, they have only gravitational forces on them. When an object is falling and has only gravitational force on it, it is said to be in what?

U: *freefall*

Figure 19.3 The tutoring domain: part of an actual conversation between a student user (U) and the ITSPOKE system (S) of (Litman and Silliman, 2004), based on the Why2-Atlas text-based tutoring system (?).

Commercial vs Research

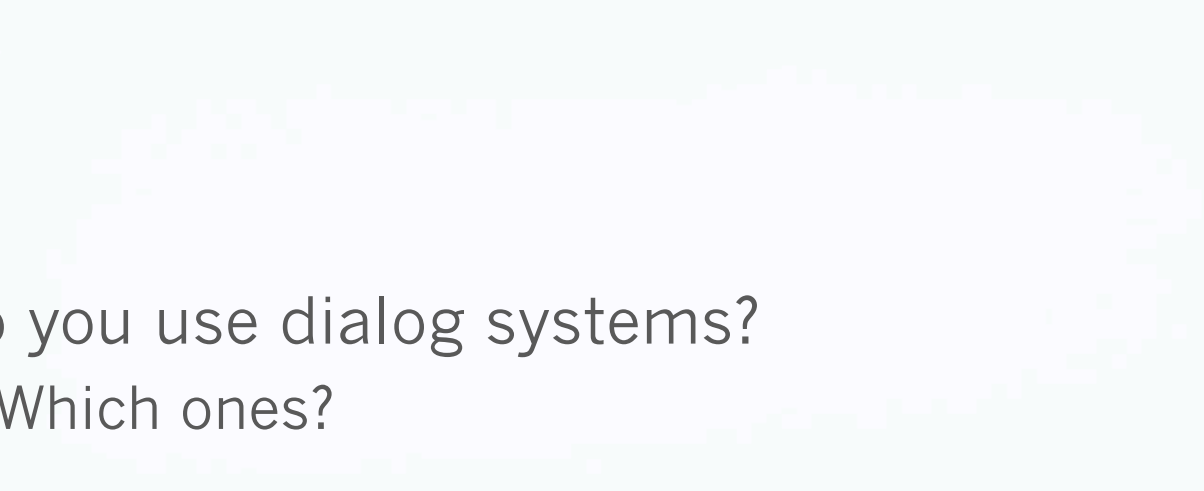
- Commercial systems:
 - Emphasize reliability
 - Typically implying more rigid design
 - > 95% accuracy for deployment
- Research systems:
 - Emphasize flexibility
 - Broader capabilities, flexible language, interaction
- Goal: research-style flexibility + high reliability

Types of Dialog Systems

- Command & control
 - In-car entertainment
 - Robotics
- Information access
 - Q&A, databases
 - Travel, flight booking
- IVR/ customer service
 - Call-routing, flexible menu structure
- Flexible assistants
 - Conversational partners, planning, etc

Aspects of Dialog Systems

- Modalities:
 - Voice only
 - Voice+GUI
 - ECA: robot, talking head
- Backend system
 - What data/API can it access?
- What does it know?
 - About the world (domain, open knowledge)
 - About the user (your Google/Amazon info)

- 
- Do you use dialog systems?
 - Which ones?
 - Why?
 - What do you like/dislike?