CS447: Natural Language Processing

http://courses.engr.illinois.edu/cs447

Lecture 27
Dialogue and Conversational Agents

Julia Hockenmaier

juliahmr@illinois.edu

3324 Siebel Center

Final exam

Wednesday, Dec 11 in class
Only materials after midterm
Same format as midterm

Review session this Friday!

Today's lecture

Dialogue
What happens when two or more people are having a conversation?

Dialogue Systems/Conversational Agents
How can we design systems to have a conversation with a human user?

- **Chatbots**
  Mostly chitchat, although also some use in therapy
- **Task-based Dialogue Systems**
  Help human user to accomplish a task
  (e.g. book a ticket, get customer service, etc.)

Dialogue

Recap: Discourse and Discourse Models

Discourse: any multi-sentence linguistic unit.
Speakers describe "some situation or state of the real or some hypothetical world" (Webber, 1983)
Speakers attempt to get the **listener** to construct a similar **model of the situation** they describe.

A **Discourse Model** is an explicit representation of:
— the **events and entities** that a discourse talks about
— the **relations** between them (and to the real world).

Dialogue

Dialogue: a conversation between two speakers

(multiparty dialogue: a conversation among more than two speakers)

Each dialogue consists of a sequence of **turns** (an **utterance** by one of the two speakers)

Turn-taking requires the ability to detect when the other speaker has finished

Speech/Dialogue Acts

Utterances correspond to actions by the speaker, e.g.

- **Constative** (answer, claim, confirm, deny, disagree, state)
  Speaker commits to something being the case
- **Directive** (advise, ask, forbid, invite, order, request)
  Speaker attempts to get listener to do something
- **Commissive** (promise, plan, bet, oppose)
  Speaker commits to a future course of action
- **Acknowledgment** (apologize, greet, thank, accept apology)
  S. expresses attitude re. listener wrt. some social action

In practice, much more fine-grained labels are often used, e.g:

Yes-No Questions, Wh-Questions, Rhetorical Questions,
Greetings, Thanks, …
Yes-Answers, No-Answers, Agreements, Disagreements, …
Statements, Opinions, Hedges, …

Dialogues have structure

Dialogues have (hierarchical) structure:

"Adjacency pairs": Some acts (first pair part) typically followed by (set up expectation for) another (second pair part):

Question → Answer, Proposal → Acceptance/Rejection, etc.

Sometimes, a **subdialogue** is required
(e.g. for clarification questions):

A: I want to book a ticket for tomorrow
B: Sorry, I didn't catch where you want to go?
A: To Chicago
B: And where do you want to leave from?
…
B: Okay, I've got the following options: …

Grounding in Dialogue

For communication to be successful, both parties have to know that they understand each other (or where they misunderstand each other)

— Both parties maintain (and communicate) ***their own beliefs about the state of affairs*** that they're talking about.
— Both parties also maintain ***beliefs about the other party's beliefs about the state of affairs*.**
— Both parties also maintain ***beliefs about the other party's beliefs about their own beliefs***,… etc.

Common ground: The set of **mutually agreed beliefs** among the parties in a dialogue

Grounding in Dialogue

John: *Dragons are scary!*

Common ground: {*John thinks dragons exist,*
Mary knows that John thinks dragons exist,
John finds dragons scary
Mary knows that John finds dragons scary, ….}

If Mary replies: *What dragons?*

—> **Additions to Common ground:**
{"*Mary doesn't think dragons exist*",
"*John knows that Mary doesn't think dragons exist*", …}

If Mary replies instead: *No, dragons are cute!*

—> **Additions to Common ground:**
{"*Mary and John both think dragons exist*",
"*Mary finds dragons cute.*"
"*John knows that Mary finds dragons cute*",
"*Mary disagrees with John that dragons are scary*",…}

Clark and Schaefer: Grounding

Grounding in dialog can be done by the following mechanisms:

- **Continued attention:** B continues attending to A
- **Relevant next contribution:** B starts in on next relevant contribution
- **Acknowledgement:** B nods or says continuer like uh-huh, yeah, assessment (great!)
- **Demonstration:** B demonstrates understanding A by paraphrasing or reformulating A's contribution, or by collaboratively completing A's utterance
- **Display:** B displays verbatim all or part of A's presentation

Initiative

Who controls the conversation?
— Who asks questions?
— Who introduces new topics?

Human-human dialogue is typically **mixed initiative** where both parties take initiative at different points
(But it is difficult to design mixed initiative dialogue systems)

Systems often assume a **user-initiative strategy**
(User asks questions, System responds)
or a **system-initiative strategy**
(System-initiative systems can be very frustrating to use)

Inference and implicature

A_1: And, what day in May did you want to travel?
C_2: OK uh I need to be there for a meeting that's from the 12th to the 15th.

The customer (speaker) doesn't answer the question directly, but assumes the provided information allows the agent (hearer) to infer the requested information,

=> customer needs to travel on 11th/12th—15th/16th.

Hearers can draw these inferences ("conversational implicatures") because they assume speakers are being cooperative

Gricean maxims: rules that govern (cooperative) human communication. Here: "Be relevant!"

Chatbots vs Dialogue Systems

Chatbots: Chitchat, often used for entertainment, originally as testbed for clinical therapy

Dialogue Systems: Typically to perform specific tasks (e.g. customer service, reservations, etc., smart devices, cars, etc.)

Chatbots

A computer therapist?

Computer: TELL ME MORE ABOUT YOUR FAMILY
Human: My mother takes care of me.
Computer: WHO ELSE IN YOUR FAMILY TAKES CARE OF YOU
Human: My father.
Computer: YOUR FATHER
Human: You are like my father in some ways.
Computer: WHAT RESEMBLANCE DO YOU SEE
Human: You are not very aggressive but I think you don't want me to notice that.
Computer: WHAT MAKES YOU THINK I AM NOT VERY AGGRESSIVE
Human: You don't argue with me.
Computer: WHY DO YOU THINK I DON'T ARGUE WITH YOU
Human: You are afraid of me.

Weizenbaum (1966), ELIZA.

ELIZA as a FST cascade

Human: *You don't argue with me.*
Computer: *WHY DO YOU THINK I DON'T ARGUE WITH YOU*

1. Replace **you** with ***I*** and ***me*** with ***you:***
   I don't argue with you.
2. Replace ***<...>*** with ***Why do you think <...>:***
   Why do you think I don't argue with you.

What about other NLP tasks?

Could we write an FST for machine translation?

Current Chatbots

IR-based approaches: mine lots of human-human dialogues

Neural approaches: seq2seq models, again trained on lots of human-human dialogues

Dialogue Systems

Dialogue systems

Systems that are capable of performing a task-driven dialogue with a human user.

AKA:

- Spoken Language Systems
- Dialogue Systems
- Speech Dialogue Systems

Applications:

- Travel arrangements (Amtrak, United airlines)
- Telephone call routing
- Tutoring
- Communicating with robots
- Anything with limited screen/keyboard

A travel dialog: Communicator

S1:	Hello. You've reached the [Communicator]. Tell me your full name
U2:	*hi i'd like to fly to seattle tuesday morning*
S3:	Traveling to Seattle on Tuesday, August eleventh in the morning. Your full name?
U4:	*[John Doe]*
S5:	Hello, [John], I'm glad to hear from you again. You can now ask me for help.
S6:	OK. Let's see I have a United flight departing Pittsburgh at ten forty five a.m. arriving into Seattle at one fifty seven p.m. Is that OK?
U7:	*Yes*
S8:	Will you return to Pittsburgh from Seattle?
U9:	*no i will go to Los Angeles*
S10:	Traveling to Los Angeles. on what date
U11:	*Sorry can we go back to talk about the previous leg again*
S12:	For the first leg, you are on a United flight 1446 departing Pittsburgh at ten forty five A.M. tomorrow, arriving in Seattle at one fifty seven P.M. Do you want to take that?
U13:	*actually i want to leave on wednesday*
S14:	a flight on wednesday.
S15:	I have a flight departing Pittsburgh at ten forty five a.m., arrives Seattle at one fifty seven p.m. Is that OK?
U16:	*Yes*

Figure 19.1 The travel domain: a fragment from a successful conversation between a user (U) and the Communicator system (S) of Xu and Rudnicky (2000).

Call routing: ATT HMIHY

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).

A tutorial dialogue: ITSPOKE

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 (?).

GUS: Hello. My name is GUS. I can help you plan a simple trip by air. Where do you want to go ?

Client: I want to go to San Diego on May 28

GUS: What time do you want to leave?

Client: I must be in San Diego before 10 am

GUS: Would you like P.S.A. flight 102 that arrives at 9.15 am ?

Client: What is the next flight ?

GUS: Air California flight 310 that leaves at 8.30am and arrives at 10.00am

Client: I'll take the first one

GUS: In what name should I make a reservation, please?

Client: Dan Bobrow

GUS: I have confirmed the following flight: P.S.A. flight 102 on Wednesday May 28 from San Jose to San Diego that leaves at 7.30 am and arrives at 9.15 am.

GUS: What date do you want to return on ?

Client: On Friday in the evening.

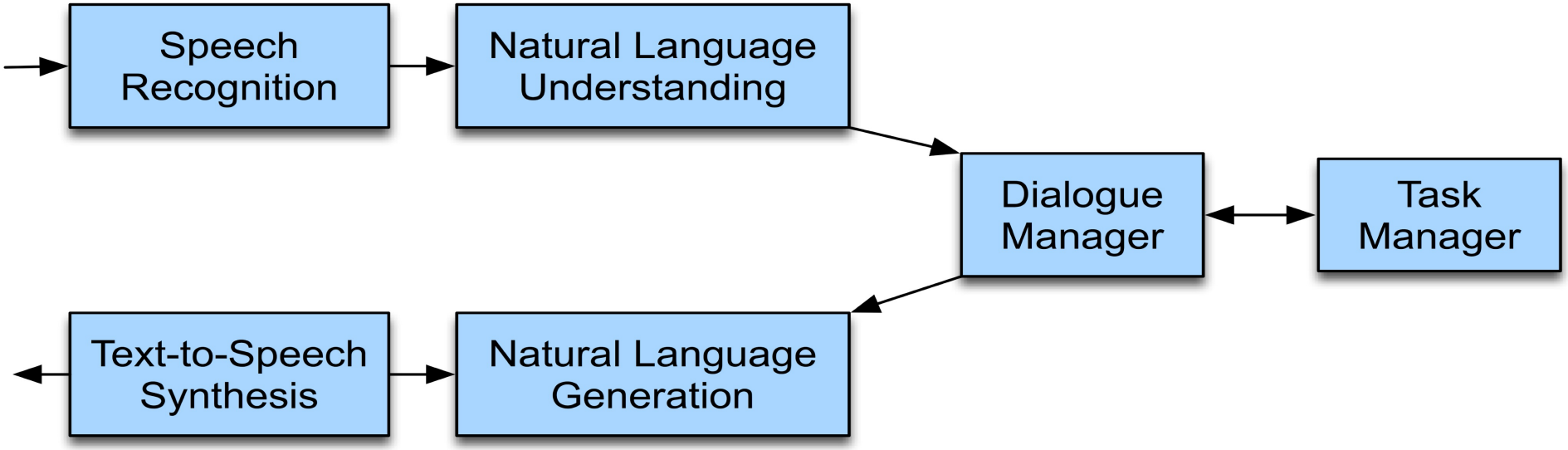
GUS: Would you like the flight that leaves at 7.45 pm ?

Client: That's fine.

GUS: I have confirmed the following flight: P.S.A. flight 307 on Friday May 30 from San Diego to San Jose that leaves at 7.45 pm and arrives at 9.30 pm Thank you for calling. Goodbye

The state of the art in 1977 !!!!

Dialogue System Architecture



Dialogue Manager

Controls the architecture and structure of dialogue

- Takes input from ASR (speech recognizer) & NLU components
- Maintains some sort of internal state
- Interfaces with Task Manager
- Passes output to Natural Language Generation/ Text-to-speech modules

Task-driven dialog as slot filling

If the purpose of the dialog is to complete a specific **task** (e.g. book a plane ticket), that task can often be represented as a **frame** with a number of **slots** to fill.

The task is completed if all necessary slots are filled.

This assumes a "**domain ontology**":

A knowledge structure representing possible user intentions for the given task

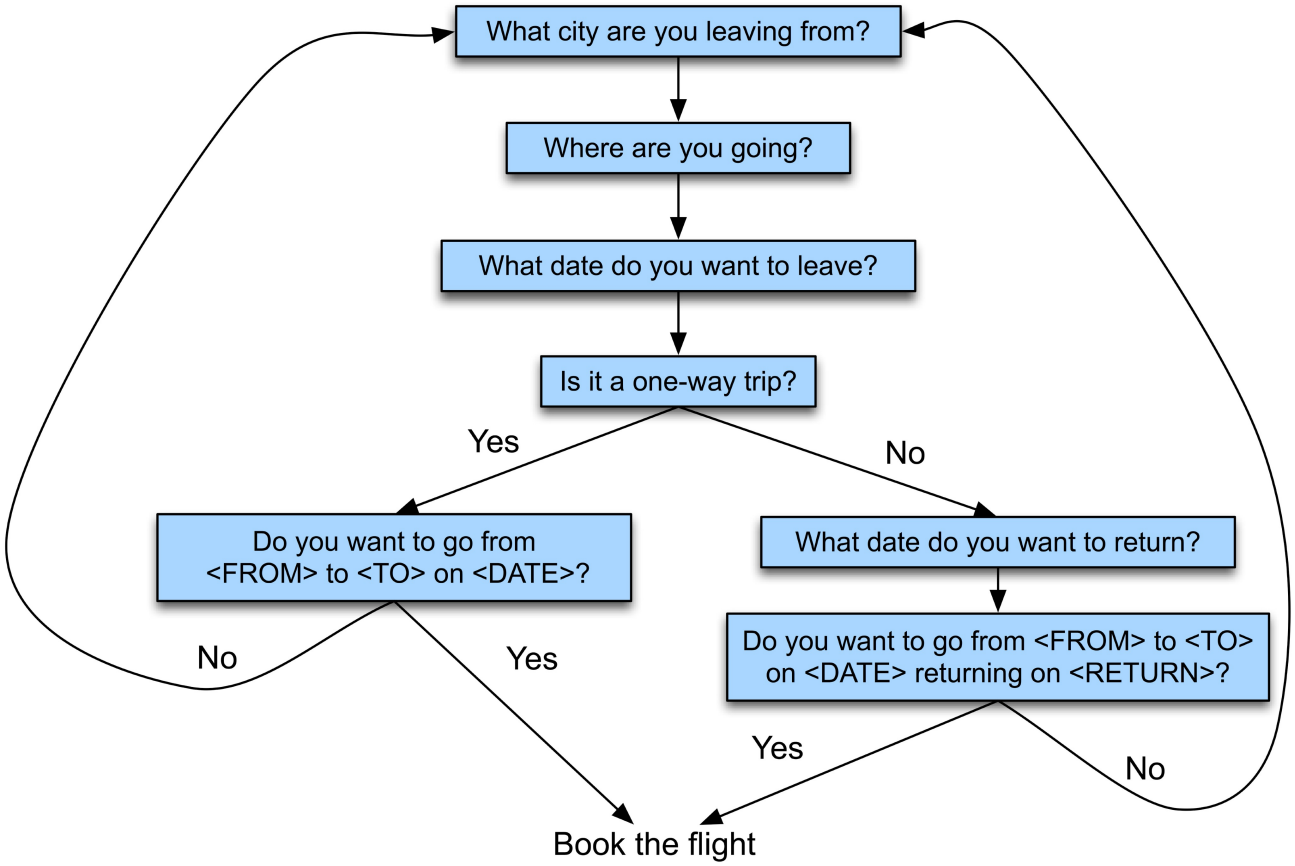
The Frame

A **frame** is set of **slots**, each to be
— filled with information of a given **type**, and
—associated with a **question** to the user

Slot	**Type**	**Question**
ORIGIN	city	*What city are you leaving from?*
DEST	city	*Where are you going?*
DEP-DATE	date	*What day would you like to leave?*
DEP-TIME	time	*What time would you like to leave?*
AIRLINE	line	*What is your preferred airline?*

Finite-state dialogue managers

Represent dialog structure as a finite state diagram



Purely sytem initiative

NLU with frame/slot semantics

But if we map user utterances to frames, we can detect which slots are filled or remain to be filled:

Show me morning flights from Boston to SF on Tuesday.

The system needs to identify the flight frame and fill in the correct slots:

```
SHOW:   FLIGHTS: ORIGIN: CITY: Boston
                         DATE: Tuesday
                         TIME: morning
                 DEST:   CITY: San Francisco
```

This allows for mixed-initiative dialogue systems.

Information-State and Dialogue Acts

If we want a dialogue system to be more than just form-filling, it needs to be able to:

- Decide when the user has asked a question, made a proposal, rejected a suggestion
- Ground a user's utterance, ask clarification questions, suggestion plans

This suggests that:

- Conversational agent needs sophisticated models of interpretation and generation
  - In terms of speech acts and grounding
  - Needs more sophisticated representation of dialogue context than just a list of slots

Grounded dialogue

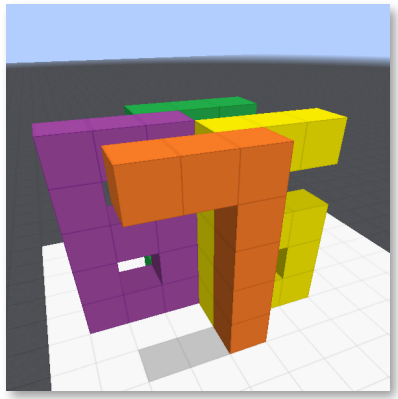
"Grounding" may also mean that utterances are mapped to/interpreted in a world

— human-robot communication: physical world
— computer games: simulated world
— talking about images/videos: world=images/videos

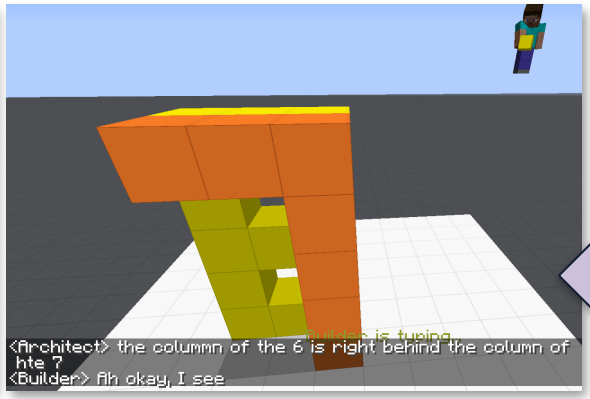
Increasingly important for communication with smart devices, (self-driving) cars, etc.

Minecraft Collaborative Building Task

ARCHITECT

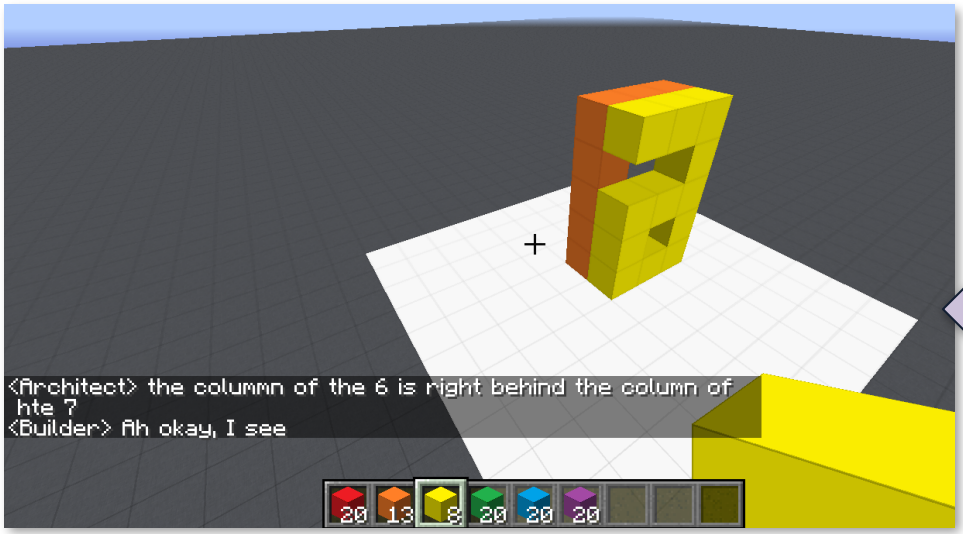


Target Structure



Build Region

BUILDER



CHAT INTERFACE

Architect: in about the middle build a column five tall

(Builder puts down five orange blocks)

Architect: then two more to the left of the top to make a 7

(Builder puts down two orange blocks)

Architect: now a yellow 6

Architect: the long edge of the 6 aligns with the stem of the 7 and faces right

Builder: Where does the 6 start?

Architect: behind the 7 from your perspective

Builder: Is it directly adjacent?

Architect: yes directly behind it. touches it

(Builder puts down twelve yellow blocks, in the shape of a 6)

Architect: too much overlap unfortunately

Architect: **the columnn of the 6 is right behind the column of hte 7**