Spoken Dialogue Systems: Human and Machine

Julia Hirschberg

CS 4706

Today

- Conversational Agents
  - ASR
  - NLU
  - Generation
- Cooperative Question-Answering
- Dialogue Representations

Conversational Agents

- AKA:
  - Interactive Voice Response Systems
  - Dialogue Systems
  - Spoken Dialogue Systems
- Applications:
  - Travel arrangements (Amtrak, United airlines)
  - Telephone call routing
  - Tutoring
  - Communicating with robots
  - Anything with limited screen/keyboard

A Travel Dialogue: 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 (?).

Conversational Structure

- Telephone conversations
  - Stage 1: Enter a conversation
  - Stage 2: Identification
  - Stage 3: Establish joint willingness to converse

Stage	Speaker & Utterance
1	A_1: (rings B's telephone)
1,2	B_1: Benjamin Holloway
2	A_1: this is Professor Dwight's secretary, from Polymania College
2,3	B_1: ooh yes –
4	A_1: uh:m . about the: lexicology *seminar*
4	B_1: *yes*

Why is this customer confused?

- Customer: (rings)
- Operator: Directory Enquiries, for which town please?
- Customer: Could you give me the phone number of um: Mrs. um: Smithson?
- Operator: Yes, which town is this at please?
- Customer: Huddleston.
- Operator: Yes. And the name again?
- Customer: Mrs. Smithson

Why is *this* customer confused?

- A: **And, what day in May did you want to travel?**
- C**: OK, uh, I need to be there for a meeting that's from the 12th to the 15th.**
- Note that client did not answer question.
- Meaning of client's sentence:
  - Meeting
    - Start-of-meeting: 12th
    - End-of-meeting: 15th
  - Doesn't say anything about flying!!!!!
- How does agent infer client is informing him/her of travel dates?

Will this client be confused?

A: … there's 3 non-stops today.

- True if, in fact, there are ***7*** non-stops today.
- But agent means: 3 and only 3.
- How can client infer that agent means:
  - *only 3*

Grice: Conversational Implicature

- Implicature: a particular class of licensed inferences
- Grice (1975): how do conversational participants draw inferences beyond what is actually asserted?
- Cooperative Principle
  - *"Make your contribution such as it is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged."*
  - A tacit agreement by speakers and listeners to cooperate in communication

4 Gricean Maxims

- Relevance: Be relevant
- Quantity: Do not make your contribution more or less informative than required
- Quality: Try to make your contribution one that is true (don’t say things that are false or for which you lack adequate evidence)
- Manner: Avoid ambiguity and obscurity; be brief and orderly

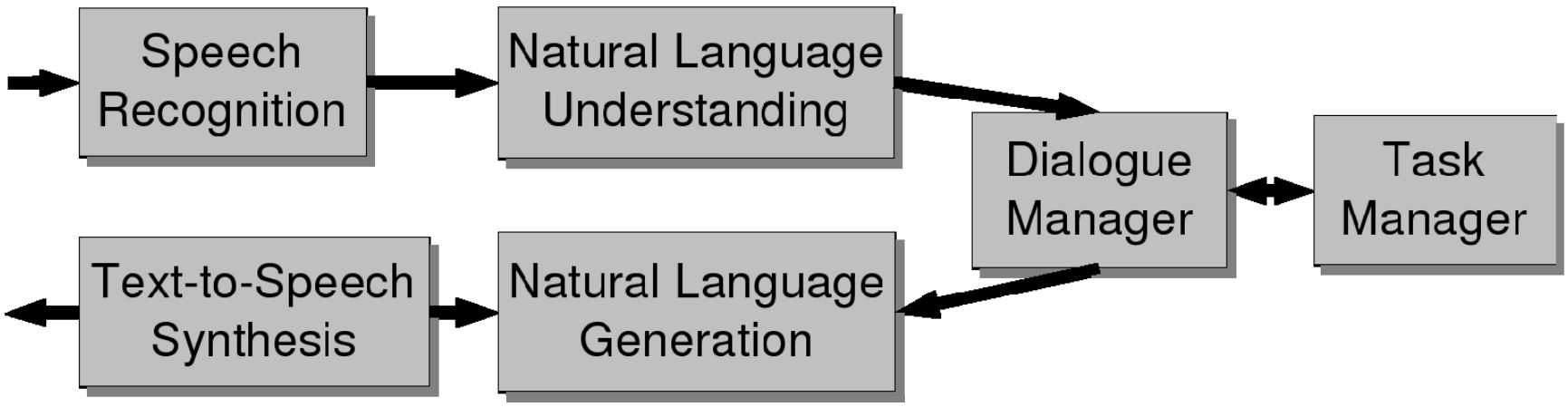
Maxim of Relevance

- A: Is Regina here?
- B: Her car is outside.
- Implicature: yes
  - Hearer thinks: *Why would B mention the car? It must be relevant. How could it be relevant? It could since if her car is here she is probably here.*
- Client: I need to be there for a meeting that's from the 12th to the 15th
  - Hearer thinks: *Speaker is following maxims, would only have mentioned meeting if it was relevant. How could meeting be relevant? If client meant me to understand that he had to depart in time to attend the meeting.*

Maxim of Quantity

- A:How much money do you have on you?
- B: I have 5 dollars
  - Implicature: not 6 dollars
- Similarly, 3 non-stops means ‘only 3’
  - Hearer thinks: *If speaker meant 7 non-stops she would have said 7 non-stops*
- A: Did you do the reading for today’s class?
- B: I intended to
  - Implicature: No
  - B’s answer would be true if B intended to do the reading AND in fact did do the reading, but would then violate Maxim of Quantity

Dialogue System Architecture



Speech Recognition

- Input: acoustic waveform
- Output: string of words
  - Basic components:
    - A **recognizer for phones**, small sound units like [k] or [ae].
    - A **pronunciation dictionary** like cat = [k ae t]
    - **A grammar** telling us what words are likely to follow what words
    - A **search algorithm** to find the best string of words

Natural Language Understanding

- Or “NLU”
- Or “Computational Semantics”
- Many ways to represent the meaning of sentences
- For speech dialogue systems, most common is Frame and Slot Semantics

A Sample Frame

- Show me morning flights from Boston to SF on Tuesday.

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

How do we Generate this Semantics?

- Many methods
  - Simplest: “semantic grammars”
  - E.g. a simple Context Free Grammar
- LHS of rules is a semantic category:
  - LIST -> show me | I want | can I see|…
  - DEPARTTIME -> (after|around|before) HOUR
    | morning | afternoon | evening
  - HOUR -> one|two|three…|twelve (am|pm)
  - FLIGHTS -> (a) flight|flights
  - ORIGIN -> from CITY
  - DESTINATION -> to CITY
  - CITY -> Boston | San Francisco | Denver | Washington

Semantics for a sentence

LIST FLIGHTS ORIGIN

Show me flights from Boston

DESTINATION DEPARTDATE

to San Francisco on Tuesday

DEPARTTIME

morning

Generation and TTS

- Generation component
  - Chooses concepts to express to user
  - Plans how to express these concepts in words
  - Assigns appropriate prosody to the words
- TTS component
  - Takes words and prosodic annotations
  - Synthesizes a waveform

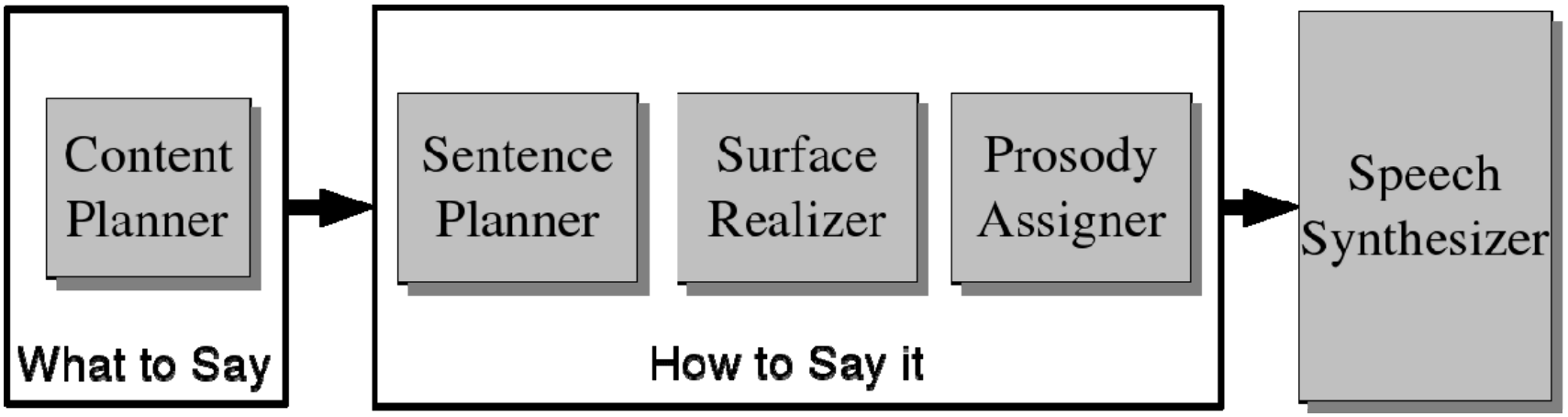
Generation Component

- Content Planner
  - Decides what information to express to user
    - (ask a question, present an answer, etc.)
  - Often merged with dialogue manager
- Language Generation
  - Chooses syntactic structures and words to express meaning.
  - Simplest method
    - Pre-specify basic sentences and fill in the blanks as appropriate
    - Template-based generation
    - Can have variables:
      - What time do you want to leave CITY-ORIG?
      - Will you return to CITY-ORIG from CITY-DEST?

More sophisticated language generation component

- Natural Language Generation
- Approach:
  - Dialogue manager builds representation of meaning of utterance to be expressed
  - Passes this to a “generator”
  - Generators have three components
    - Sentence planner
    - Surface realizer
    - Prosody assigner

Architecture of a generator for a dialogue system
(after Walker and Rambow 2002)



HCI constraints on generation for dialogue: Coherence

- Discourse markers and pronouns (Coherence”):

Please say the date.…
Please say the start time.…
Please say the duration…
Please say the subject…

Bad!

First, tell me the date.…
Next, I’ll need the time it starts.…
Thanks. <pause> Now, how long is it supposed to last?…
Last of all, I just need a brief description

Good!

HCI Constraints on Dialogue Generation Coherence: Tapered Prompts

- Prompts which get incrementally shorter:

  System: Now, what's the first company to add to your watch list?

  Caller: Cisco

  System: What's the next company name? (Or, you can say, "Finished")

  Caller: IBM

  System: Tell me the next company name, or say, "Finished."

  Caller: Intel

  System: Next one?

  Caller: America Online.

  System: Next?

  Caller: …

Issues in Building SDS

- Which architecture to use?
- How much discourse history to collect?
- What type of dialogue strategy to follow?
- What type of confirmation strategy to use?
- What kind of help to provide?
- How to recover from errors?

Project 3

- Connecting Project 1 (TTS) and Project 2 (ASR)

Next Class

- SDS Architecture: the Let's Go system from CMU