Dialogue and Conversational Agents (part 1)

Ing. R. Tedesco. PhD, AA 20-21

(mostly from: Speech and Language Processing - Jurafsky and Martin)

Outline

A) The Linguistics of Conversation
B) Basic Conversational Agents
- ASR
- NLU
- Generation
- Dialogue Manager

C) Dialogue Manager Design
- Finite State
- Frame-based
- Initiative: User, System, Mixed

Conversational Agents

- 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).

A) Linguistics of Human Conversation

I. Turn-taking
II. Speech Acts
III. Grounding
IV. Implicature

Premise: communication

- Communication between two parties
- Required, but not sufficient conditions, for successful communication

- Common language
  - Symbols, syntax, semantics
- Channel
- Common "background" information

I. Turn-taking

- Dialogue is characterized by turn-taking.
  - A:
  - B:
  - A:
  - B:
  - …
- How do speakers know when to take the floor?
  - Total amount of overlap, in human conversations, relatively small (5% - Levinson 1983)
  - Don't pause either
  - Must be a way to know who should talk and when.

Turn-taking rules

- TRP: Transition-Relevance Place
  - Places where the structure of the language allows speaker shift to occur
  - TRP usually at utterance boundaries
- At each TRP of each turn:
  a) If during this turn the current speaker has selected B as the next speaker then B must speak next.
  b) If the current speaker does not select the next speaker, any other speaker may take the next turn.
  c) If no one else takes the next turn, the current speaker may take the next turn.

Implications of subrule a)

- For some utterances the current speaker implicitly selects the next speaker
  - Adjacency pairs
    - Question/answer
    - Greeting/greeting
    - Compliment/downplayer
    - Request/grant
- Silence between 2 parts of adjacency pair is *significant silence* (refusal to respond or dispreferred response)
  - A: *Is there something bothering you or not?*
  - (1.0 s)
  - A: *Yes or no?*
  - (1.5 s)
  - A: *Eh?*
  - B: *No.*

II. Speech Acts

- Austin (1962): An utterance is a kind of action
- Clear case: performative verbs
  - *I [hereby]* ***sentence*** *you to five years in prison*
  - *I* ***name*** *this ship the Titanic*
- Performative verbs (sentence, name, …)
- Austin's idea: not just these verbs

Each utterance is 3 acts

- Locutionary act: the utterance of a sentence with a particular meaning
- Illocutionary act: the act of *asking*, *answering*, *promising*, etc., in uttering a sentence
- Perlocutionary act: the (often intentional) production of certain effects upon the thoughts, feelings, or actions of addressee in uttering a sentence

Example of 3 acts

You can't do that!

- Locutionary force:
  - Exclamation
- Illocutionary force:
  - Protesting
- Perlocutionary force:
  - Intent to annoy addressee
  - Intent to stop addressee from doing something

The 3 levels of act revisited

	Locutionary Force	**Illocutionary Force**	**Perlocutionary Force**
Can I have the rest of your sandwich?	Question	Request	Intent: You give me sandwich
I want the rest of your sandwich	Declarative	Request	Intent: You give me sandwich
Give me your sandwich!	Imperative	Request	Intent: You give me sandwich

Illocutionary Acts

- What are they?
- A theory: Searle's Speech Acts

5 classes of speech acts: Searle (1975)

- **Assertives**: committing the speaker to something's being the case
  - *suggesting, putting forward, swearing, boasting, concluding*
- **Directives**: attempts by the speaker to get the addressee to do something
  - *asking, ordering, requesting, inviting, advising, begging*
- **Commissives:** committing the speaker to some future course of action
  - *promising, planning, vowing, betting, opposing*
- **Expressives**: expressing the psychological state of the speaker about a state of affairs
  - *thanking, apologizing, welcoming, deploring*
- **Declarations**: bringing about a different state of the world via the utterance
  - *I resign; You're fired*

III. Grounding

- Dialogue is a collective act performed by speaker and hearer
- Common ground: set of things mutually believed by both speaker and hearer
- Need to achieve common ground, so hearer must ground or acknowledge speaker's utterance.
- Clark (1996):
  - **Principle of closure.** Agents performing an action require evidence, sufficient for current purposes, that they have succeeded in performing it
  - Need to know whether an action succeeded *or failed*
  - I.e., need acknowledgement by the hearer
- How to maintain grounding…

- NB: Grounding ≠ shared, background knowledge!

Clark and Schaefer: Maintaining Grounding

- 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

Grounding examples

- Display:
  - C: *I need to travel in May*
  - A: *And, what day in May did you want to travel?*
- Acknowledgement
  - C: *He wants to fly from Boston*
  - A: *mm-hmm*
  - C: *to Baltimore Washington International*

Mm-hmm (usually transcribed "uh-huh") is called **backchannel**, **continuer**, or **acknowledgement** token

Grounding Examples (2)

- Acknowledgement + next relevant contribution
  - *And, what day in May did you want to travel?*
  - *And you're flying into what city?*
  - *And what time would you like to leave?*
- The and indicates to the client that agent has successfully understood answer to the last question

A human-human conversation

C_1:	. . . I need to travel in May.
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.
A_2:	And you're flying into what city?
C_3:	Seattle.
A_3:	And what time would you like to leave Pittsburgh?
C_4:	Uh hmm I don't think there's many options for non-stop.
A_4:	Right. There's three non-stops today.
C_5:	What are they?
A_5:	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.
C_6:	OK I'll take the 5ish flight on the night before on the 11th.
A_6:	On the 11th? OK. Departing at 5:55pm arrives Seattle at 8pm, U.S. Air flight 115.
C_7:	OK.

Grounding negative responses

From Cohen et al. (2004)

- System: *Did you want to review some more of your personal profile?*
- Caller: *No.*
- System: *Okay, what's next?*

Good!

- System: *Did you want to review some more of your personal profile?*
- Caller: *No.*
- System: *What's next?*

Bad!

Grounding and Dialogue Systems

- Grounding is not just a tidbit about humans
- Is key to design of conversational agent
- Why?
  - HCI researchers find users of speech-based interfaces are confused when system doesn't give them an explicit acknowledgement signal
  - Stifelman et al. (1993), Yankelovich et al. (1995)

IV. Conversational Implicature

- Agent: *And, what day in May did you want to travel?*
- Client: *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!!!!!
- What does permit the agent to infer travel dates from information provided by the client?

Grice: conversational implicature

- Implicature means a particular class of licensed inferences.
- Cooperative Principle
  - This is a tacit agreement by speakers and listeners to cooperate in communication
- Grice (1975) proposed that what enables hearers to draw correct inferences is expressed by 4 "rules"

4 Gricean Maxims

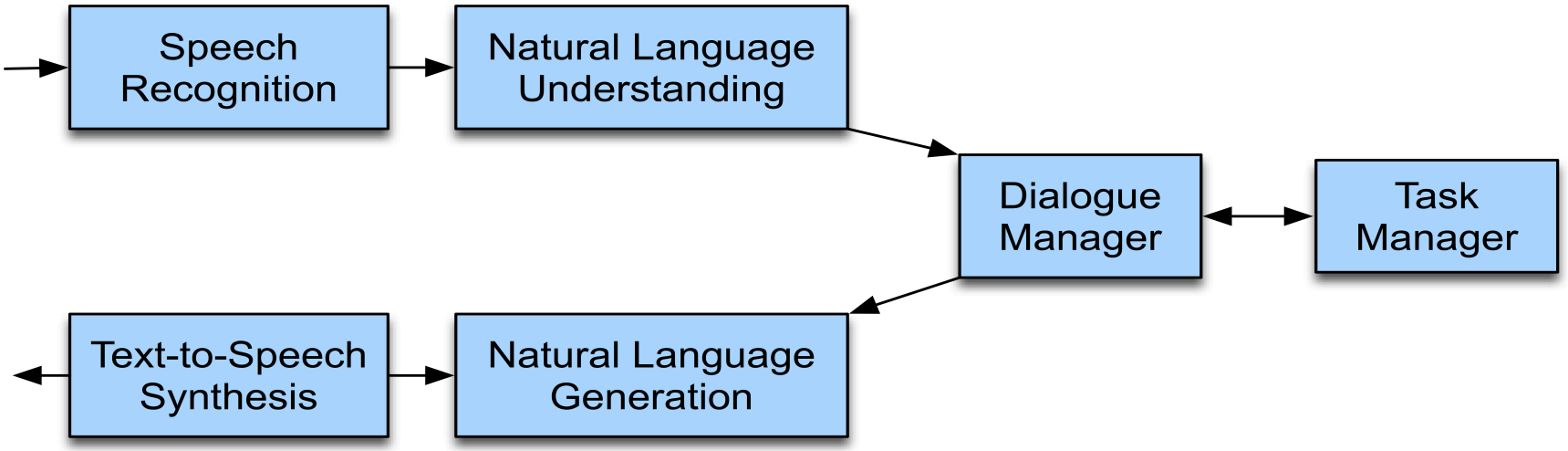
- Relevance: Be relevant
- Quantity: Do not make your contribution more or less informative than required
- Quality: try to make your contribution one that is true (don't say things that are false or for which you lack adequate evidence)
- Manner: Avoid ambiguity and obscurity; be brief and orderly

Example: Relevance

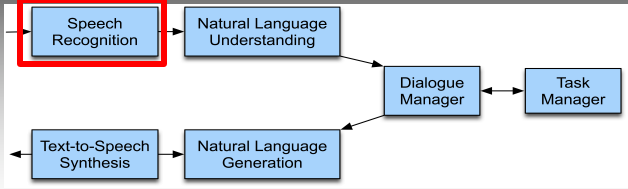
- A: ***Is Regina here?***
- B: ***Her car is outside.***
- Implication: yes
  - Hearer thinks: **why would he mention the car? It must be relevant. How could it be relevant? It could since if her car is here she is probably here.**

- Back to the example on travel…
- 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 for the meeting.**

B)Dialogue System Architecture

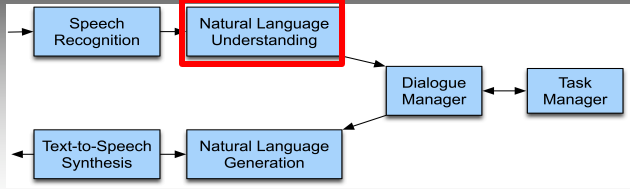


Speech recognition



- Or ASR (Automatic Speech Recognition)
  - Speech to words
- Input: acoustic waveform
- Output: string of words
  - Basic components:
    - a **recognizer for phones**, small sound units like [k] or [ae].
    - a **pronunciation dictionary**: cat = [k ae t]
    - **a language model** telling us what words are likely to follow what words
    - A **search algorithm** to find the best string of words
  - More on that in a dedicated lecture…

Natural Language Understanding



- Or "NLU"
- Or "Computational semantics"
- There are many ways to represent the meaning of sentences
- For speech dialogue systems, most common is "Frame and slot semantics"

An example of a 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 to generate this semantics?

- Many methods,
- Simplest: "semantic grammars"
- CFG in which the 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

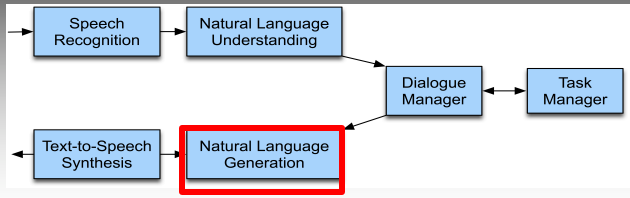
Semantics for a sentence

LIST FLIGHTS ORIGIN
Show me flights from Boston

DESTINATION DEPARTDATE
to San Francisco on Tuesday

DEPARTTIME
morning

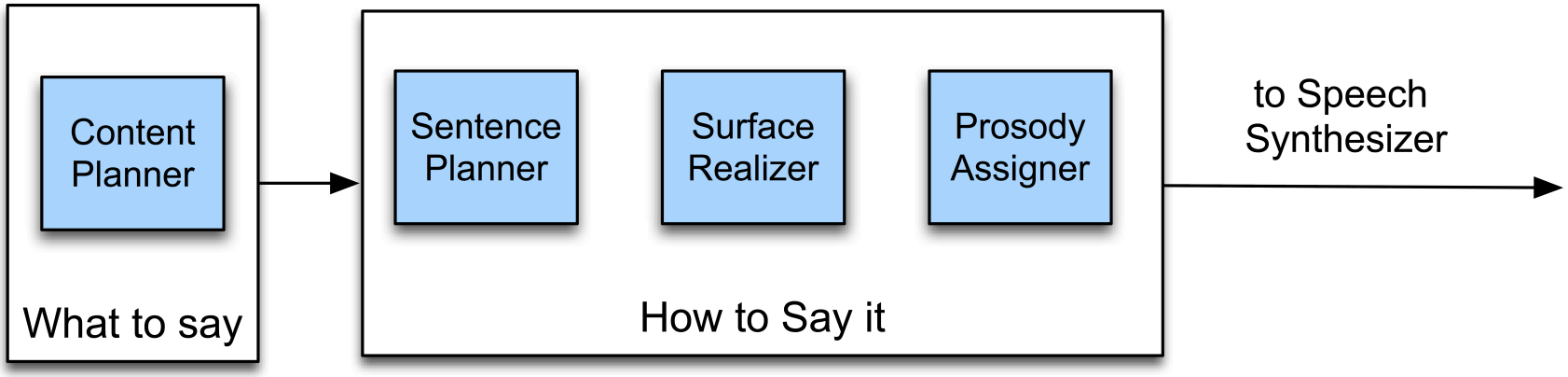
NL Generation



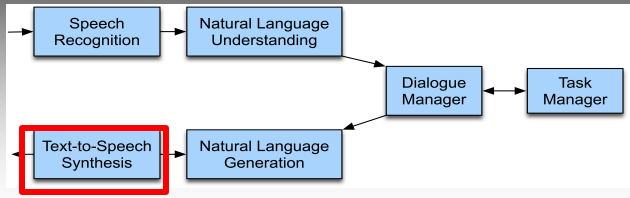
- Natural Language Generation
  - Chooses syntactic structures and words to express meaning.
- Simplest method
  - All words in sentence are prespecified!
  - "Template-based generation"
  - Can have variables:
    - What time do you want to leave CITY-ORIG?
    - Will you return to CITY-ORIG from CITY-DEST?

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

- Dialogue Manager (aka Content Planner) builds representation of meaning of utterance to be expressed
- Passes this to a "generator"
- Generators have three components
  - Sentence planner
  - Surface realizer
  - Prosody assigner



Text To Speech



- Speech synthesizer
- From a sentence and a tagging about prosody…
- … generates the waveform
- Details in a dedicated lecture…

HCI constraints on generation for dialogue: "Coherence"

- Discourse markers and pronouns ("Coherence"):

(1) *Please say the date.*
…
Please say the start time.
…
Please say the duration…
…
Please say the subject…

Bad!

(2) *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*

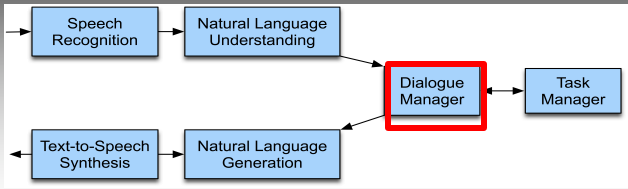
Good!

HCI constraints on generation for dialogue: coherence (II): tapered prompts

- Prompts which get incrementally shorter:

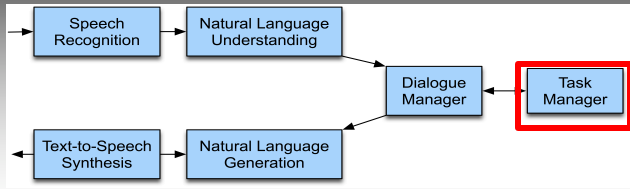
- System: Now, what's the first company to add to your watch list?
- Caller: Cisco
- System: What's the next company name? (Or, you can say, "Finished")
- Caller: IBM
- System: Tell me the next company name, or say, "Finished."
- Caller: Intel
- System: Next one?
- Caller: America Online.
- System: Next?
- Caller: …

Dialogue Manager



- Controls the architecture and structure of dialogue
  - Takes input from ASR/NLU components
  - Maintains some sort of state
  - Interfaces with Task Manager
  - Passes output to NLG/TTS modules

Task Manager



- The "behavior" of the Agent
- Depends on the goal of the Agent
  - Specific: flight booking, …
  - Less specific: synthetic psychologist
  - General: open conversation; way more difficult!
- Models
  - Logics
  - ML: reinforcement learning is quite popular

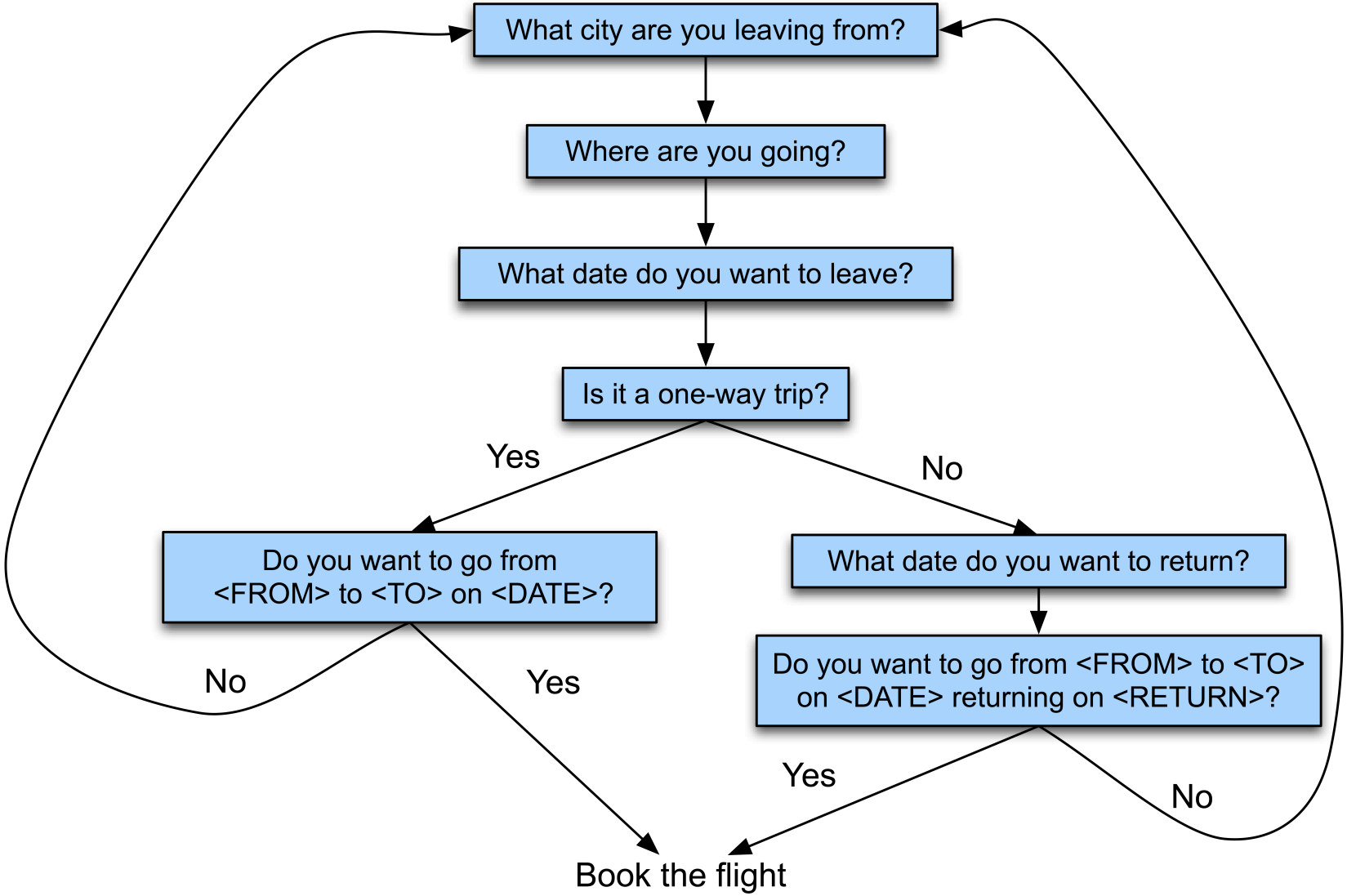
C) Architectures for dialogue management

1. Finite State
2. Frame-based
- Other architectures
  - Markov Decision Processes
  - AI Planning

1. Finite-State Dialogue Mgmt

- Consider a trivial airline travel system
  - Ask the user for a departure city
  - For a destination city
  - For a time
  - Whether the trip is round-trip or not

Example of State Dialogue Manager



Dialogue control

- In the previous slide, the system completely controls the conversation with the user
- It asks the user a series of questions
- Ignoring (or misinterpreting) anything the user says that is not a direct answer to the system's questions

Dialogue Initiative

- Systems that control conversation like this are system initiative or single initiative.
- "Initiative": who has control of conversation
- In normal human-human dialogue, initiative shifts back and forth between participants.

System Initiative

- Systems which completely control the conversation at all times are called system initiative.
- Advantages:
  - Simple to build
  - User always knows what they can say next
  - System always knows what user can say next
    - Known words: Better performance from ASR
    - Known topic: Better performance from NLU
  - Ok for VERY simple tasks (entering a credit card, or login name and password)
- Disadvantage:
  - Too limited

User Initiative

- User directs the system
- Generally, user asks a single question, system answers
- System can't ask questions back, engage in clarification dialogue, confirmation dialogue
- Used for simple database queries
- User asks question, system gives answer
- Web search is user initiative dialogue.

Problems with System Initiative

- Real dialogue involves give and take!
- In travel planning, users might want to say something that is not the direct answer to the question
- For example answering more than one question in a sentence:
  - *Hi, I'd like to fly from Seattle Tuesday morning*
  - *I want a flight from Milwaukee to Orlando one way leaving after 5 p.m. on Wednesday*

Single initiative + universals

- We can give users a little more flexibility by adding universal commands
- Universals: commands you can say anywhere
- As if we augmented every state of FSA with these
  - Help
  - Start over
  - Correct
- This describes many implemented systems
- But still doesn't allow user to say what he wants to say

Mixed Initiative

- Conversational initiative can shift between system and user
- Simplest kind of mixed initiative: use the structure of the frame itself to guide dialogue

Slot	Question
ORIGIN	What city are you leaving from?
DEST	Where are you going?
DEPT DATE	What day would you like to leave?
DEPT TIME	What time would you like to leave?
AIRLINE	What is your preferred airline?

2. Frames (mixed-initiative)

- User can answer multiple questions at once
- System asks questions of user, filling any slots that user specifies
- When frame is filled, do database query
- If user answers 3 questions at once, system has to fill slots and not ask these questions again!
- Anyhow, we avoid the strict constraints on order of the finite-state architecture

Defining Mixed Initiative

- Mixed Initiative could mean:
  - User can arbitrarily take or give up initiative in various ways
    - This is really only possible in very complex plan-based dialogue systems
    - Important research area
  - Something simpler and quite specific which we will define in the next few slides

True Mixed Initiative

C_1:	. . . I need to travel in May.
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.
A_2:	And you're flying into what city?
C_3:	Seattle.
A_3:	And what time would you like to leave Pittsburgh?
C_4:	Uh hmm I don't think there's many options for non-stop.
A_4:	Right. There's three non-stops today.
C_5:	What are they?
A_5:	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.
C_6:	OK I'll take the 5ish flight on the night before on the 11th.
A_6:	On the 11th? OK. Departing at 5:55pm arrives Seattle at 8pm, U.S. Air flight 115.
C_7:	OK.

How mixed initiative is usually defined

- First, we need to define two other factors
- Open prompts vs. directive prompts
- Restrictive versus non-restrictive grammar

Open vs. Directive Prompts

- Open prompt
  - System gives user very few constraints
  - User can respond how they please:
  - "*How may I help you?*" "*How may I direct your call?*"
- Directive prompt
  - Explicit instructs user how to respond
  - "*Say yes if you accept the call; otherwise, say no*"

Restrictive vs. Non-restrictive grammars

- Restrictive grammar
  - Language model which strongly constrains the ASR system, based on dialogue state
  - I.e., the Agent ASR is only able to understand specific words at specific conversation points
    + better ASR accuracy
    – restrictive
- Non-restrictive grammar
  - Open language model which is not restricted to a particular dialogue state

Definition of Mixed Initiative

Grammar	Open Prompt	Directive Prompt
Restrictive	Doesn't make sense	System Initiative
Non-restrictive	User Initiative	Mixed Initiative