

CS447: Natural Language Processing

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

# Lecture 27:

# Generation and Dialog

Julia Hockenmaier

*juliahmr@illinois.edu*

3324 Siebel Center

# Final exam: Sat, Dec 16, 7:00pm

**Rooms:** DCL 1310 and Siebel 0216

(we'll get back to you on how we'll split up into two rooms)

**Conflict exam:**

Monday, Dec 18 (room/time TBD)

Email me asap if you need to take the conflict exam, and explain why.

# What is Generation?

Automatic production of natural language text,  
usually from underlying semantic representation

- As “natural-language front ends” used to present information in databases etc.:  
weather forecasts, train systems,  
(personalized) museum/restaurant/shopping guides,...
- In dialog systems
- In summarization systems
- In authoring aids to help people create routine documents:  
customer support, job ads, etc...

# Today's lecture

Very high-level overview of (classical) NLG research, with a bit about dialog at the end.

Today's materials are mainly taken from:  
Ehud Reiter and Robert Dale *“Building applied natural language generation systems”*, Natural Language Engineering, 1997

Many questions in NLG research are very domain and task specific

# Example: Rail travel information system

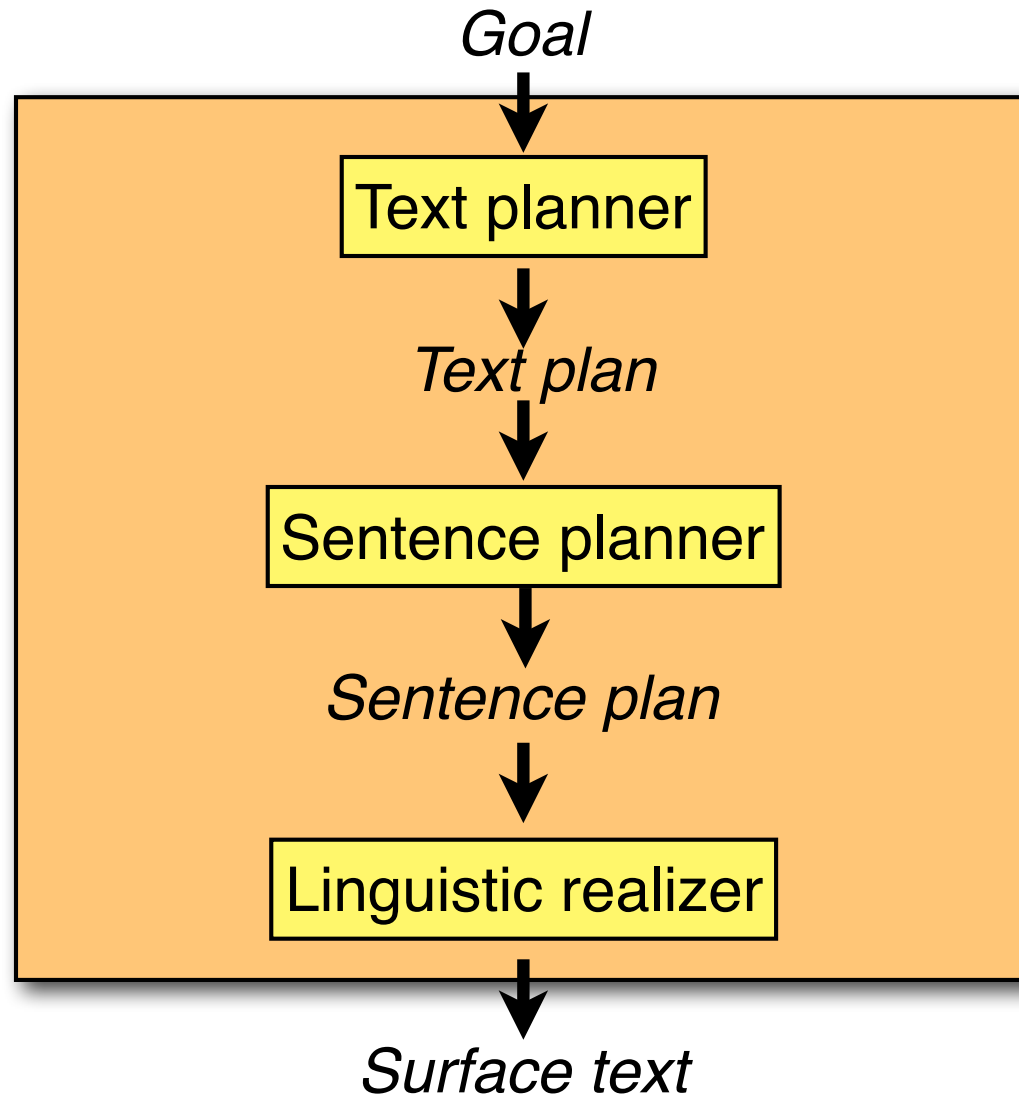
**-Domain knowledge:** Train schedules

**-User Input:** from a graphical user interface, or in natural language: *“How can I get from Aberdeen to Glasgow?”*

**-Desired output:**

*There are 20 trains each day from Aberdeen to Glasgow. The next train is the Caledonian Express; it leaves Aberdeen at 10am. It is due to arrive in Glasgow at 1pm, but arrival may be slightly delayed because of snow on the track near Stirling.*

# NLG architecture



# NLG architectures

There are many dependencies between these tasks.  
The standard NLG system architecture consists of:

## **Text planning:**

Content determination and discourse planning

## **Sentence planning:**

Sentence aggregation, lexicalization and referring expression generation

## **Linguistic realization:**

Syntactic, morphological and orthographic processing.

# NLG tasks

Text planning	<b>1. Content determination</b> What information (what ‘messages’) should be communicated?
	<b>2. Discourse planning</b> How should the messages be structured/ordered?
Sentence planning	<b>3. Sentence aggregation</b> Which messages should be combined into individual sentences?
	<b>4. Lexicalization</b> In which words/phrases should domain concepts/relations be expressed?
	<b>5. Referring expression generation</b> How should entities be referred to?
	<b>6. Linguistic realization</b> Generate a grammatical and orthographically well-formed text



# Content determination

**Input:** user input and background knowledge (database)

**Output:** a set of 'messages' to be communicated  
(here shown with gloss)

a.  $\left[ \begin{array}{l} \text{message-id: msg01} \\ \text{relation: IDENTITY} \\ \text{arguments: } \left[ \begin{array}{l} \text{arg1: NEXT-TRAIN} \\ \text{arg2: CALEDONIAN-EXPRESS} \end{array} \right] \end{array} \right]$

b. The next train is the Caledonian Express

a.  $\left[ \begin{array}{l} \text{message-id: msg02} \\ \text{relation: DEPARTURE} \\ \text{arguments: } \left[ \begin{array}{l} \text{departing-entity: CALEDONIAN-EXPRESS} \\ \text{departure-location: ABERDEEN} \\ \text{departure-time: 1000} \end{array} \right] \end{array} \right]$

b. The Caledonian Express leaves Aberdeen at 10am

# Content determination

**Input:** user input and background knowledge (database)

**Output:** a set of ‘messages’ to be communicated

**User model:** User’s task, user’s level of expertise, previous interactions with system (esp. in dialog)

Need to filter, summarize and process input data

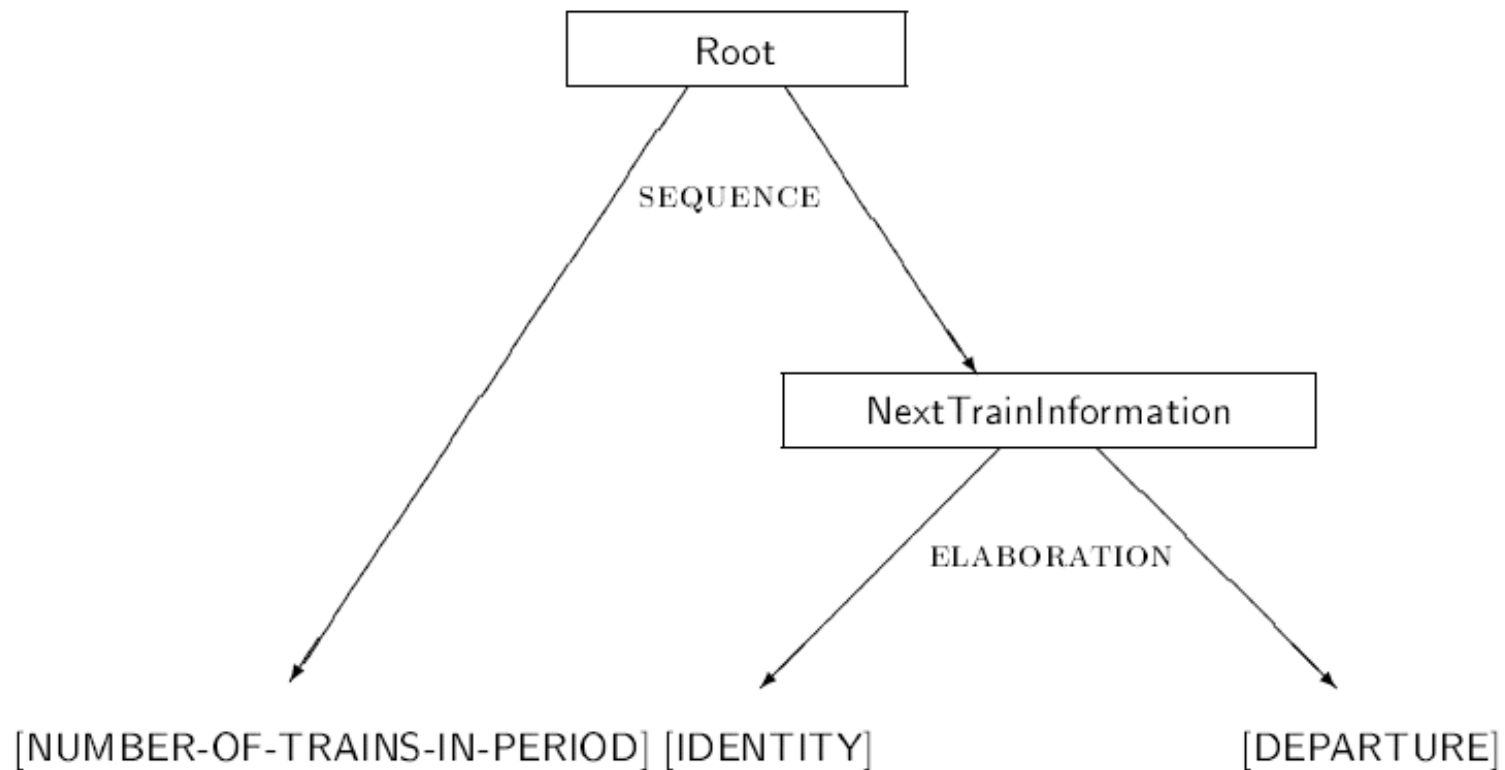
Relies often on (system-specific) heuristics (looking at corpus helps!)

# Discourse planning

How should the messages be ordered?

What are the discourse relations that hold between them?

Often represented as a tree:



# Sentence aggregation

Which messages should be conveyed in a single sentence?

*The next train leaves at 10am. It is the Caledonian Express.*

*The next train, which leaves at 10am, is the Caledonian Express.*

**Linguistic means to combine messages (=clauses):**

- **Relative clauses:** *The next train, which leaves at 10 am, is the Caledonian Express*
- **Coordination:** *The Caledonian Express leaves at 10am, and is the next train*
- **Subordination:** *The Caledonian Express is the next train, although it leaves only at 10am.*
- **Lists:** *There are trains at 10am, at 11:30am and at 1:00pm.*

# Lexicalization and referring expressions

**Lexicalization:** Which words and phrases should be used to express domain concepts:

- does the train 'leave' or 'depart'?
- a 'statistical error' is not the same as a 'statistical mistake'

NLG systems need a domain lexicon

**Referring expression generation:**

When do you use a pronoun/a definite NP/an indefinite NP to refer to an entity?

Needs a discourse model

# Linguistic realization

Generate a grammatically and orthographically correct English utterance:

```
[message-id: msg03
relation: NUMBER-OF-TRAINS-IN-PERIOD
arguments: [source: ABERDEEN
            destination: GLASGOW
            number: 20
            period: DAILY]]
```

***There are 20 trains each day from Glasgow to Edinburgh.***

# NLG evaluation

Many areas of NLP have shared task evaluations that allow comparisons of different algorithms/systems on the same data.

But most NLG systems are very domain/application specific.

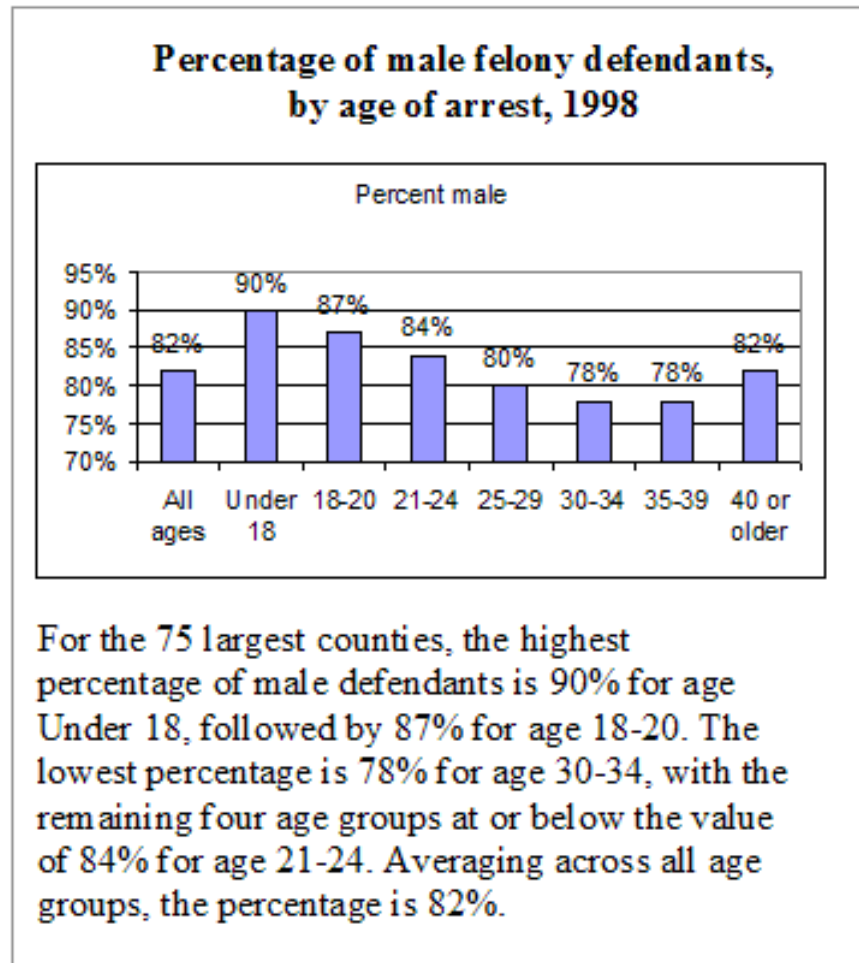
- Every system starts from its own input representation
- Not a single gold standard data set
- Can we evaluate subtasks (e.g. referring expression generation)?
- How can we compare system outputs against each other/ against human produced text?  
(metrics such as BLEU/ROUGE may not correlate highly enough with human judgments)

# Some NLG systems



# Cogentex's chart explainer

<http://www.cogentex.com/products/chartex/faq/bjs-sample.png>



# Cogentex's Camera system

## User Input

I want a camera for:

☐ Basic snapshots ☒ Business use ☐ Pro photography

I want to view the pictures as:

☐ On-screen images ☒ 4x6 prints ☐ 8x10 blowups

I don't want to pay more than:

☐ \$250 ☐ \$500 ☒ \$1000

## Generated Response

I've found three cameras that match your photographic needs:

Model	Max. Resolution	Weight (oz.)	Price
Phentax DigiLens 110	1024 x 768	11.3	\$199
Kodiak Zoomer 450	1024 x 768	9.9	\$249
Nikad FlashBack Q20	2048 x 1360	13.1	\$649

All of these models are easy to use, which makes them solid, predictable performers in a business context. And they all have 1024 x 768 pixels or more of resolution, which will give satisfactory print quality in smaller sizes.

The most affordable choice is the Phentax DigiLens 110. The Kodiak Zoomer 450 is more compact and portable, but it costs \$50 more than the DigiLens. Given the price difference, my choice would be the DigiLens, unless you value the Zoomer's compactness.

The DigiLens is in stock for immediate shipment. The Zoomer is currently out of stock, but should ship by next Sunday, August 24. As usual, a carrying case is yours for free if you order before next Wednesday, August 27.

# Edinburgh's ILEX and M-PIRO

ILEX: a web-based virtual museum gallery and a phone-based system for an actual gallery

M-PIRO: adds an authoring tool for curators

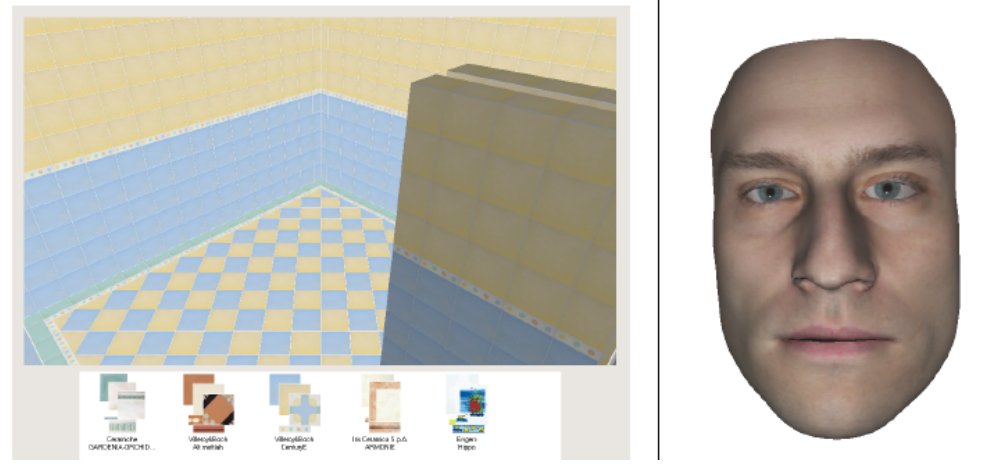
## *What is that?*



This exhibit is a lekythos, created during the archaic period. It dates from circa 500 BC. It was painted by Amasis with the red figure technique and it originates from Attica.

# The COMIC system

**C**onversational **M**ultimodal **I**nteraction with **C**omputers  
Dialog system for bathroom design applications



---

User Tell me about this design *[click on Alt Mettlach]*

COMIC *[Look at screen]*

THIS DESIGN is in the CLASSIC style.  
*[circle tiles]*

As you can see, the colours are DARK RED and OFF WHITE.  
*[point at tiles]*

The tiles are from the ALT METTLACH collection by VILLEROY AND BOCH.  
*[point at design name]*

---

# Summarization

“The process of distilling the most important information from a text to produce an abridged version for a particular task and user”

- Abstract or extract?
- Generic (no specific task/user) or query-focused?
- Single-document or multi-document?

## Output:

- Abstracts (of scientific papers)
- Headlines (or newspaper articles)
- Snippets (for webpages)
- Answers to complex questions (from multiple sources)

# Extracts from a single document

Goal: Produce a paragraph that summarizes a document

## 1. Content selection:

- Find 'important' (key) sentences

- Extract key facts/phrases

## 2. Information ordering:

- What order should these key facts be presented in?

## 3. Sentence realization:

- Produce a coherent paragraph from the list of key facts

# Centroid-based content selection

Which sentences are most **central** in a document?

Binary classification task: *sentence*  $\rightarrow$  {*include*, *don't include*}

Method A: Central sentences = **salient/informative** sentences:

- a **sentence is salient** if it contains many salient words.
  - a **word is salient** (=informative) in a document  
if it occurs significantly more often than expected (if  $-2 \log \lambda(w) > 10$ )
- Likelihood ratio  $\lambda(w)$ :  $P_{\text{doc}}(w)/P_{\text{English}}(w)$

Method B: Central sentences = most **similar** to other s's in doc.

- compute sentence-based TF/IDF for the words in a document  
(sentence=TF/IDF's document, document= TF/IDF's collection)
- distance between sentences: cosine of TD/IDF vectors of all words
- centrality of sentence i: average distance to all other sentences in document

# RST-based summarization

Use a **discourse parser** to identify rhetorical relations between sentences/clauses in a document.

This gives a **discourse tree** with hierarchical **nucleus-satellite** relations between clauses

This discourse tree defines a **salience ranking**: the **highest nuclei** in the tree are the most salient



# Information ordering and Sentence Realization

In which order should the key phrases be presented?

Simplest case: order in which they appear in document

Finding the optimal solution is NP-complete, but we can approximate

Use centering theory to measure **coherence**

Use coreference resolution and parsing to produce an ‘entity grid’  
(which entities occur in which sentence, and in which role),  
then find good sequences of transitions

Sentence realization may require some rephrasing:

Use longer descriptions to introduce entities, shorter ones to refer back

“**Bush** met Putin today. **George W. Bush** said...”

=> “**George W. Bush** met Putin today. **Bush** said...”

# Conversational Agents (Chapter 24)

# Conversational Agents

Systems that are capable of performing a task-driven dialog 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?

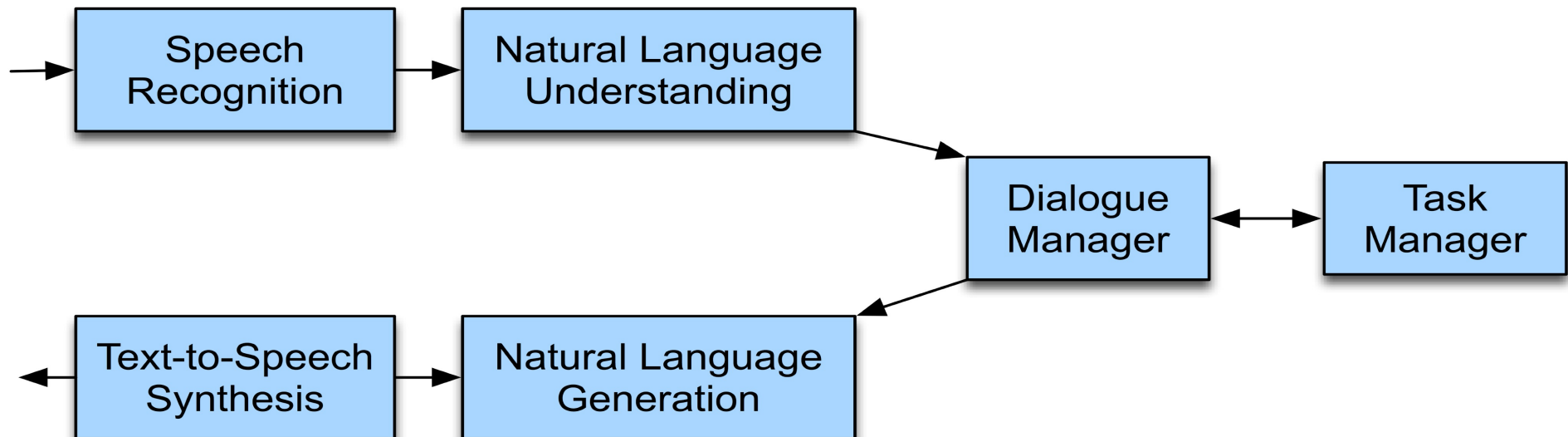
U: *it will be moving*

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 (?).

# Dialogue System Architecture



# NLU with frame/slot semantics

There are many ways to represent the meaning of sentences. For speech dialogue systems, most common is “Frame slot semantics”:

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

SHOW:

FLIGHTS:

ORIGIN:

CITY: Boston

DATE: Tuesday

TIME: morning

DEST:

CITY: San Francisco



# Task-driven dialog as slot filling

If the purpose of the dialog is to complete a specific task (e.g. to 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 the necessary slots are filled.

# 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

# Four architectures for dialogue management

Finite State

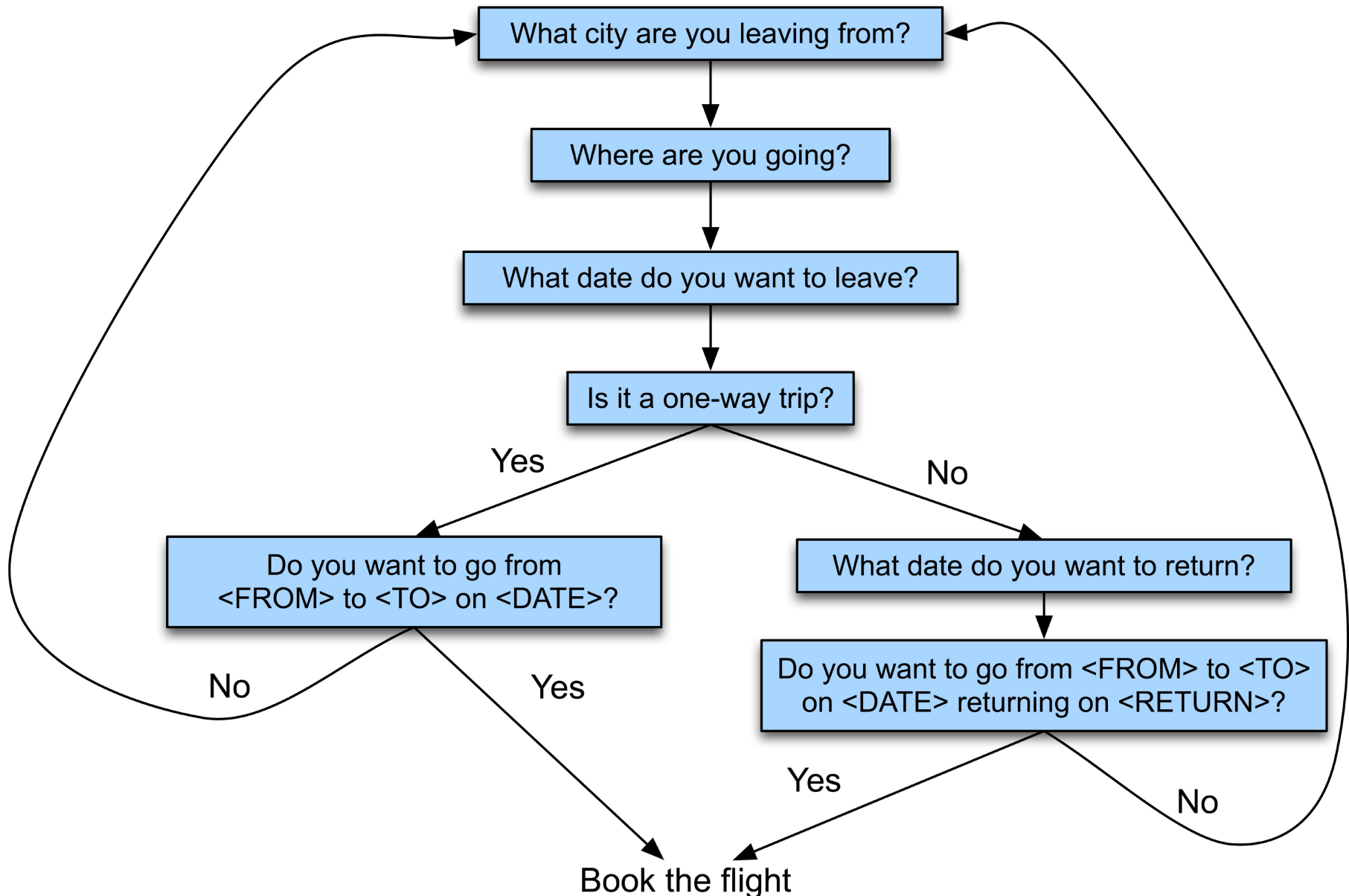
Frame-based

Information State

Markov Decision Processes

AI Planning

# Finite State Dialogue Manager



# Finite-state dialogue managers

System completely controls the conversation with the user:

- It asks the user a series of questions
- It may ignore (or misinterpret) anything the user says that is not a direct answer to the system's questions

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.

# 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 they want 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

ORIGIN

DEST

DEPT DATE

DEPT TIME

AIRLINE

## Question

What city are you leaving from?

Where are you going?

What day would you like to leave?

What time would you like to leave?

What is your preferred airline?

# True Mixed Initiative

C<sub>1</sub>: ...I need to travel in May.

A<sub>1</sub>: And, what day in May did you want to travel?

C<sub>2</sub>: OK uh I need to be there for a meeting that's from the 12th to the 15th.

A<sub>2</sub>: And you're flying into what city?

C<sub>3</sub>: Seattle.

A<sub>3</sub>: And what time would you like to leave Pittsburgh?

C<sub>4</sub>: Uh hmm I don't think there's many options for non-stop.

A<sub>4</sub>: Right. There's three non-stops today.

C<sub>5</sub>: What are they?

A<sub>5</sub>: 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<sub>6</sub>: OK I'll take the 5ish flight on the night before on the 11th.

A<sub>6</sub>: On the 11th? OK. Departing at 5:55pm arrives Seattle at 8pm, U.S. Air  
flight 115.

C<sub>7</sub>: OK.



# 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

# Grounding in Dialog

Dialogue is a **collective act** performed by speaker and hearer

**Common ground:** set of things mutually believed by both speaker and hearer.

To achieve common ground, hearer must *ground* or *acknowledge* speakers 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

# 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