CS 224S / LINGUIST 285
Spoken Language Processing

Andrew Maas
Stanford University
Spring 2017

Lecture 10: Dialogue System Introduction and Frame-Based Dialogue

Original slides by Dan Jurafsky

Dialog section

- May 3: Dialog introduction. Frame based systems
- May 8: Human conversation. Reinforcement learning for dialog
- May 10: Deep learning for dialog (Jiwei)
- May 31: Dialog in industry (Alex Lebrun, Founder of Wit.ai and Facebook M)

Outline

- Basic Conversational Agents
  - ASR
  - NLU
  - Generation
  - Dialogue Manager
- Dialogue Manager Design
  - Finite State
  - Frame-based
- Dialogue Design Considerations

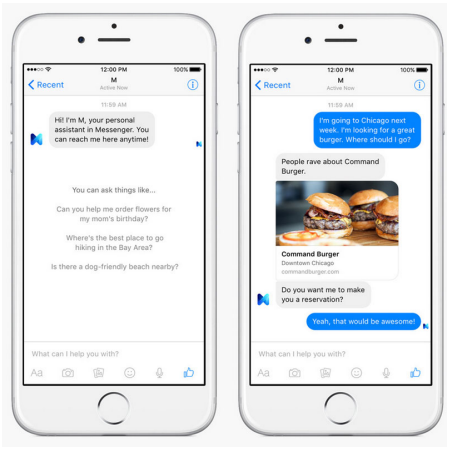
Conversational Agents

- 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

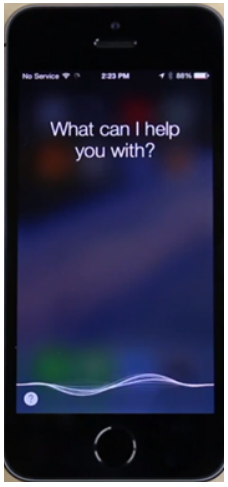
Conversational systems

Amazon Echo
2015

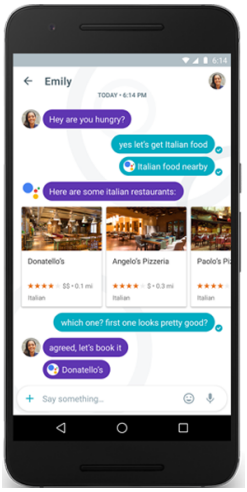
Google Home
2016



Facebook M
2015

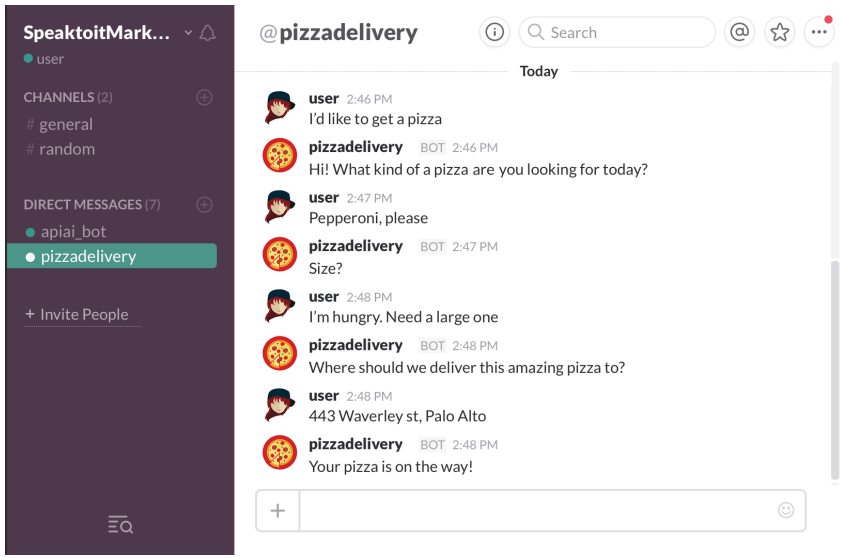


Apple
Siri
2011



Google
Assistant
2016

Microsoft
Cortana
2014



Slack Bot
2015

A travel dialog: Communicator Xu and Rudnicky (2000)

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*

Call routing: ATT HMIHY
Goren et al. (1997)

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.

A tutorial dialogue: ITSPOKE Litman and Silliman (2004)

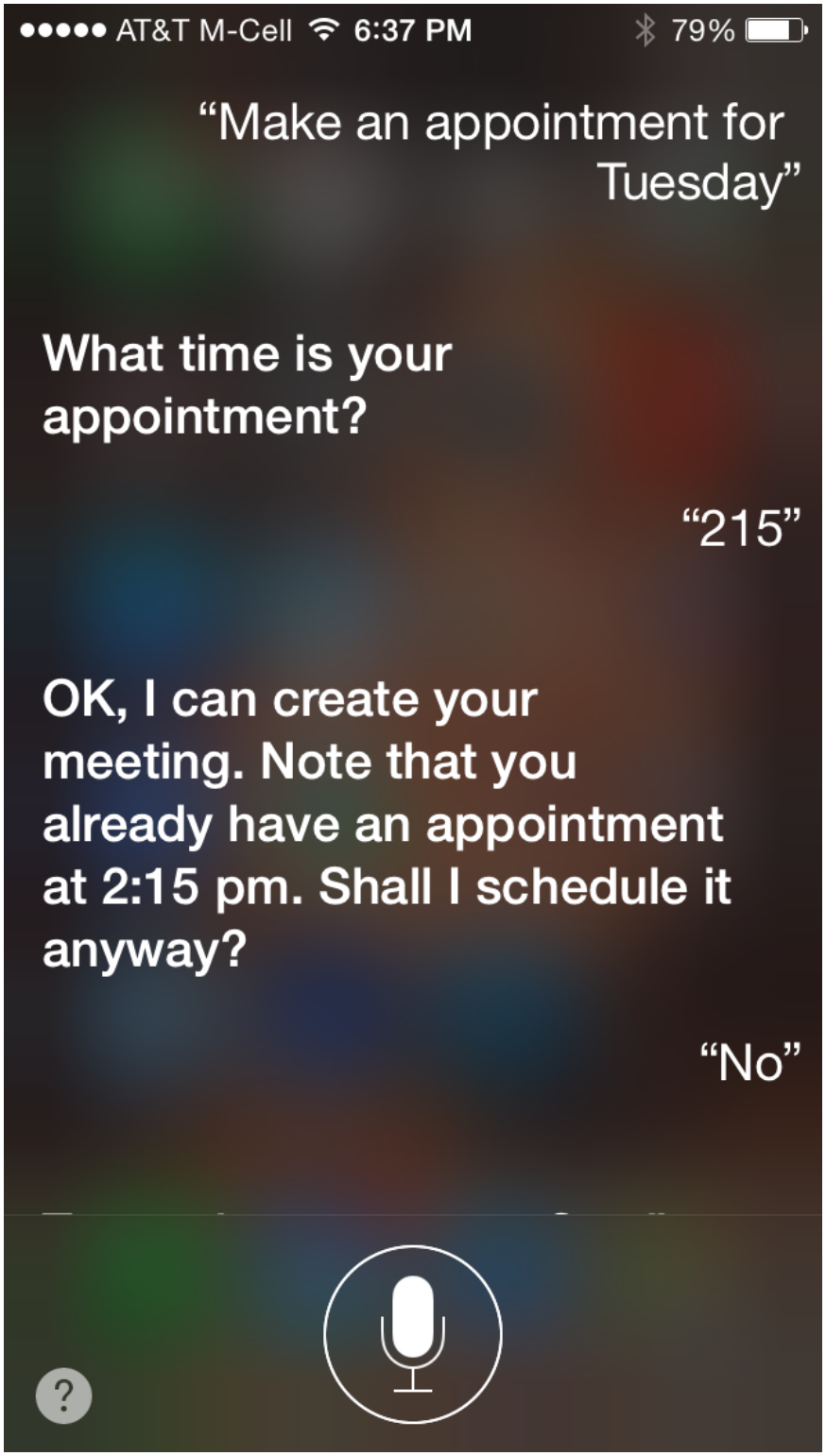
S: I didn't understand. What force acts on all objects within close proximity to earth?
U: *uh the force of gravity*
S: Fine. Besides the Earth's gravitational force, are there any other forces acting on the person, keys, and elevator after the release?
U: *no*
S: Yep. Thus, keep in mind that after the 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*

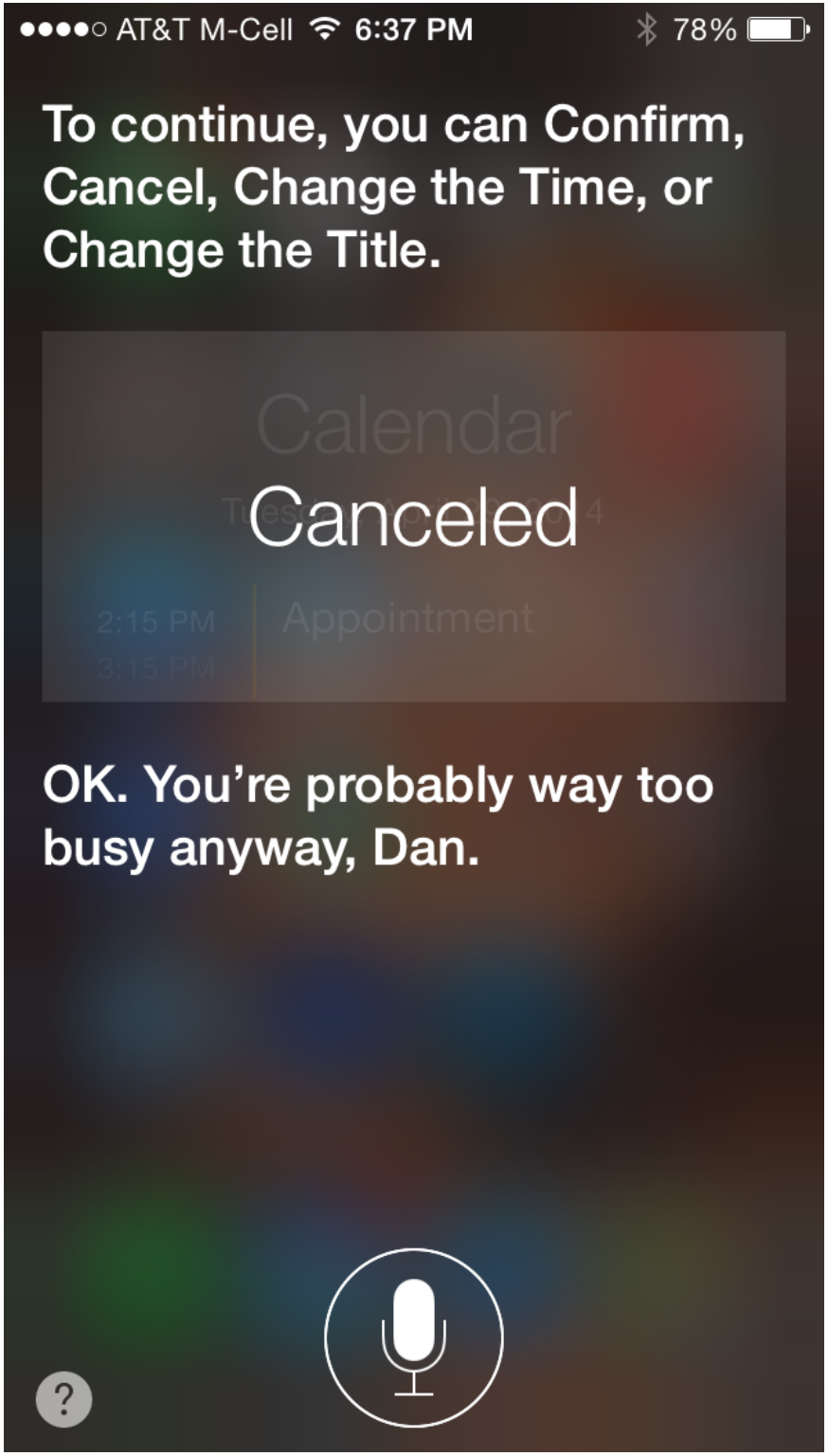
Conversational Agent Design Issues

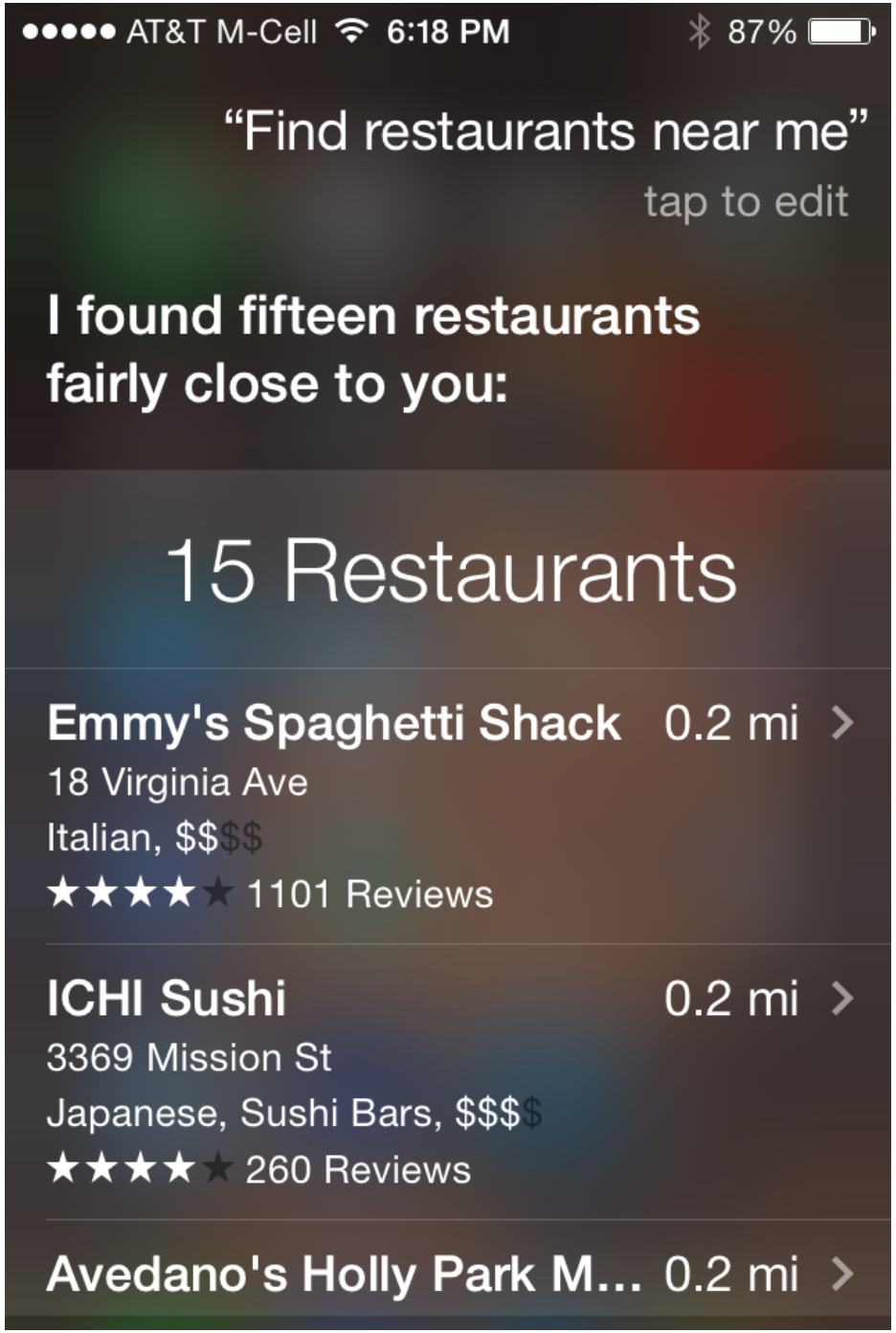
- Time to response (Synchronous?)
- Task complexity
  - What time is it?
  - Book me a flight and hotel for vacation in Greece
- Interaction complexity / number of turns
  - Single command/response
  - “I want new shoes” What kind? What color? What size?
- Initiative
  - User, System, Mixed
- Interaction modality
  - Purely spoken, Purely text, Mixing speech/text/media

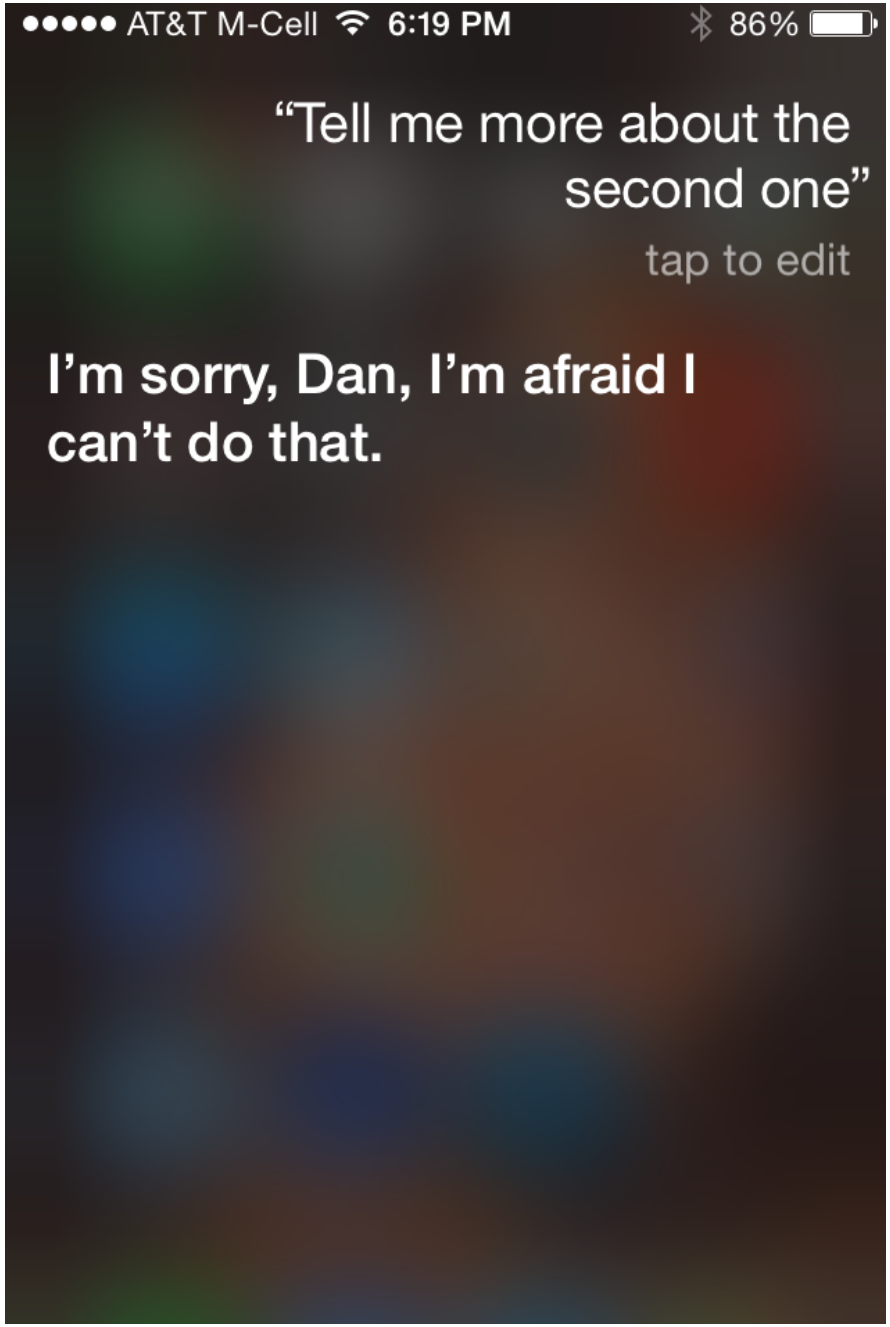
Spoken Synchronous Personal Assistants

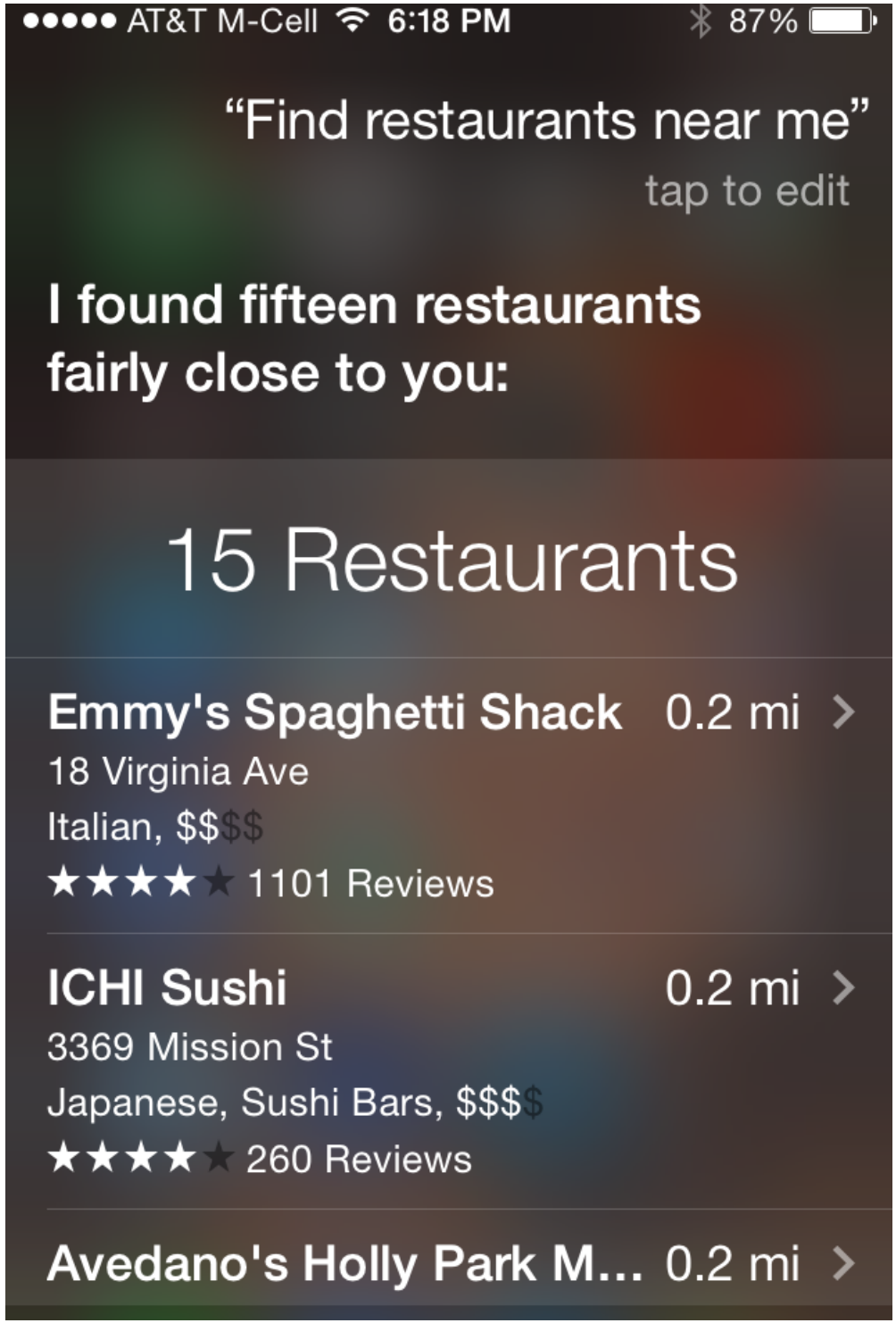
- Siri
- Google Now
- Microsoft Cortana
- Amazon Alexa


●●●●● AT&T M-Cell 6:37 PM 79%
“Make an appointment for Tuesday”
What time is your appointment?
“215”
OK, I can create your meeting. Note that you already have an appointment at 2:15 pm. Shall I schedule it anyway?
“No”


●●●●○ AT&T M-Cell 6:37 PM 78%
To continue, you can Confirm, Cancel, Change the Time, or Change the Title.
Calendar
Canceled
2:15 PM
3:15 PM
Appointment
OK. You’re probably way too busy anyway, Dan.

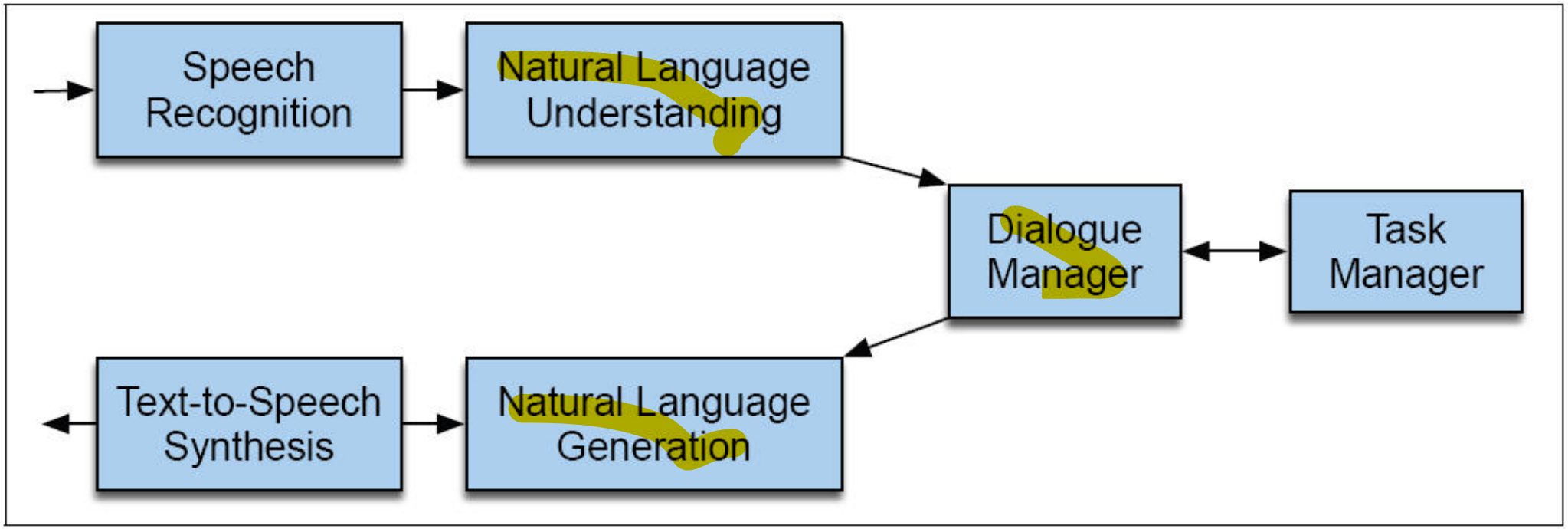

●●●●● AT&T M-Cell 6:18 PM 87%
“Find restaurants near me”
tap to edit
I found fifteen restaurants fairly close to you:
15 Restaurants
Emmy's Spaghetti Shack 0.2 mi
18 Virginia Ave
Italian, $$$$
★★★★★ 1101 Reviews
ICHI Sushi 0.2 mi
3369 Mission St
Japanese, Sushi Bars, $$$$
★★★★★ 260 Reviews
Avedano's Holly Park M… 0.2 mi


●●●●● AT&T M-Cell 6:19 PM 86%
“Tell me more about the second one”
tap to edit
I’m sorry, Dan, I’m afraid I can’t do that.


AT&T M-Cell 6:18 PM 87%
"Find restaurants near me"
tap to edit
I found fifteen restaurants fairly close to you:
15 Restaurants
Emmy's Spaghetti Shack 0.2 mi
18 Virginia Ave
Italian, $$$$
★★★★★ 1101 Reviews
ICHI Sushi 0.2 mi
3369 Mission St
Japanese, Sushi Bars, $$$$
★★★★★ 260 Reviews
Avedano's Holly Park M... 0.2 mi

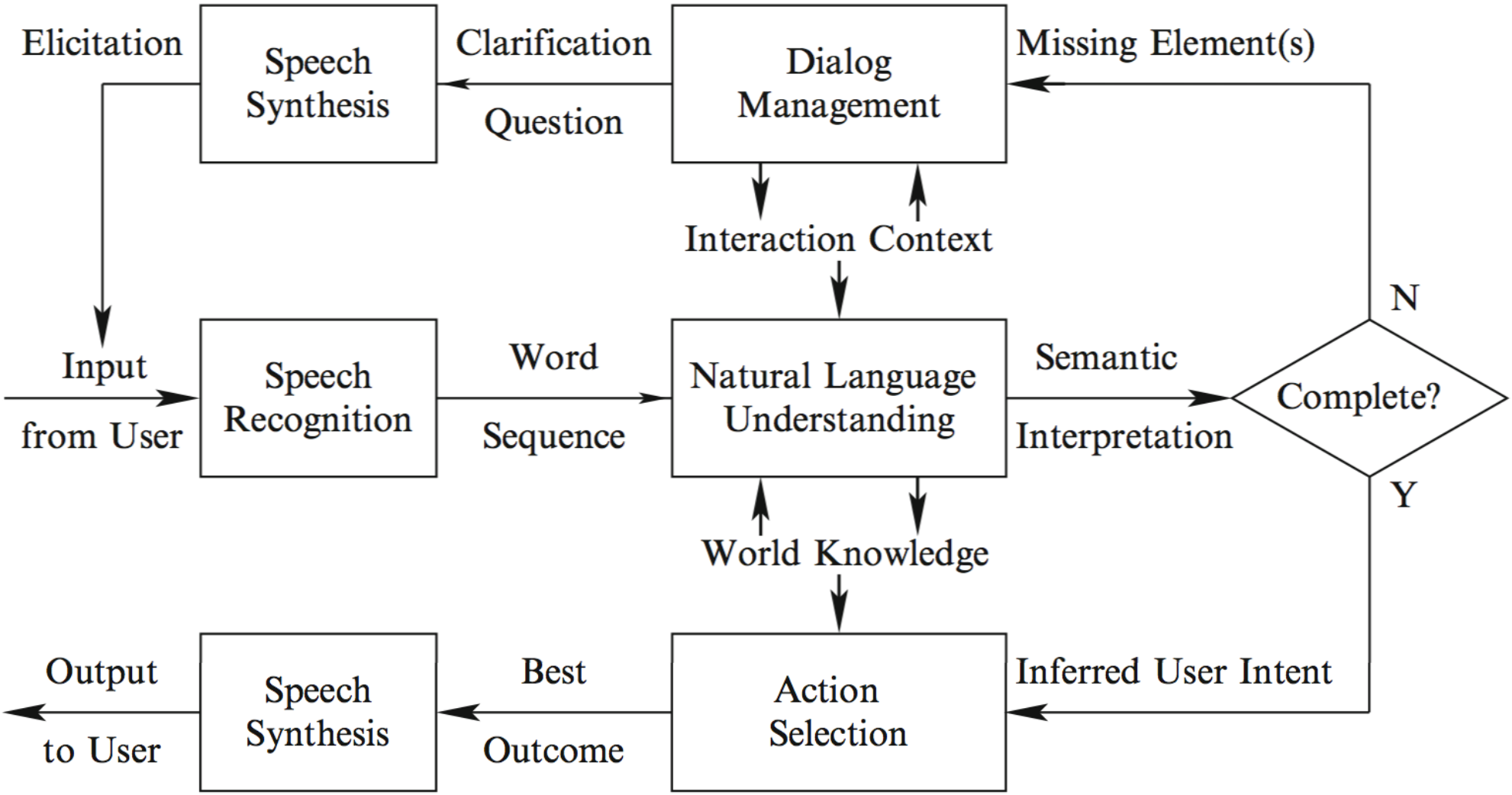

AT&T M-Cell 6:19 PM 86%
"Are any of them Italian"
tap to edit
My web search turned this up:
Web Search
Are any of them Italian
any - Dizionario inglese-italiano WordReference
www.wordreference.com
English-Italian Dictionary | any ... of any sort adj (of an unspecified variety) di qualsiasi
Italian language - Wikipedia, the free encyclopedia
en.wikipedia.org

Dialogue System Architecture



Dialog architecture for Personal Assistants

Bellegarda



Dialog architecture for Personal Assistants

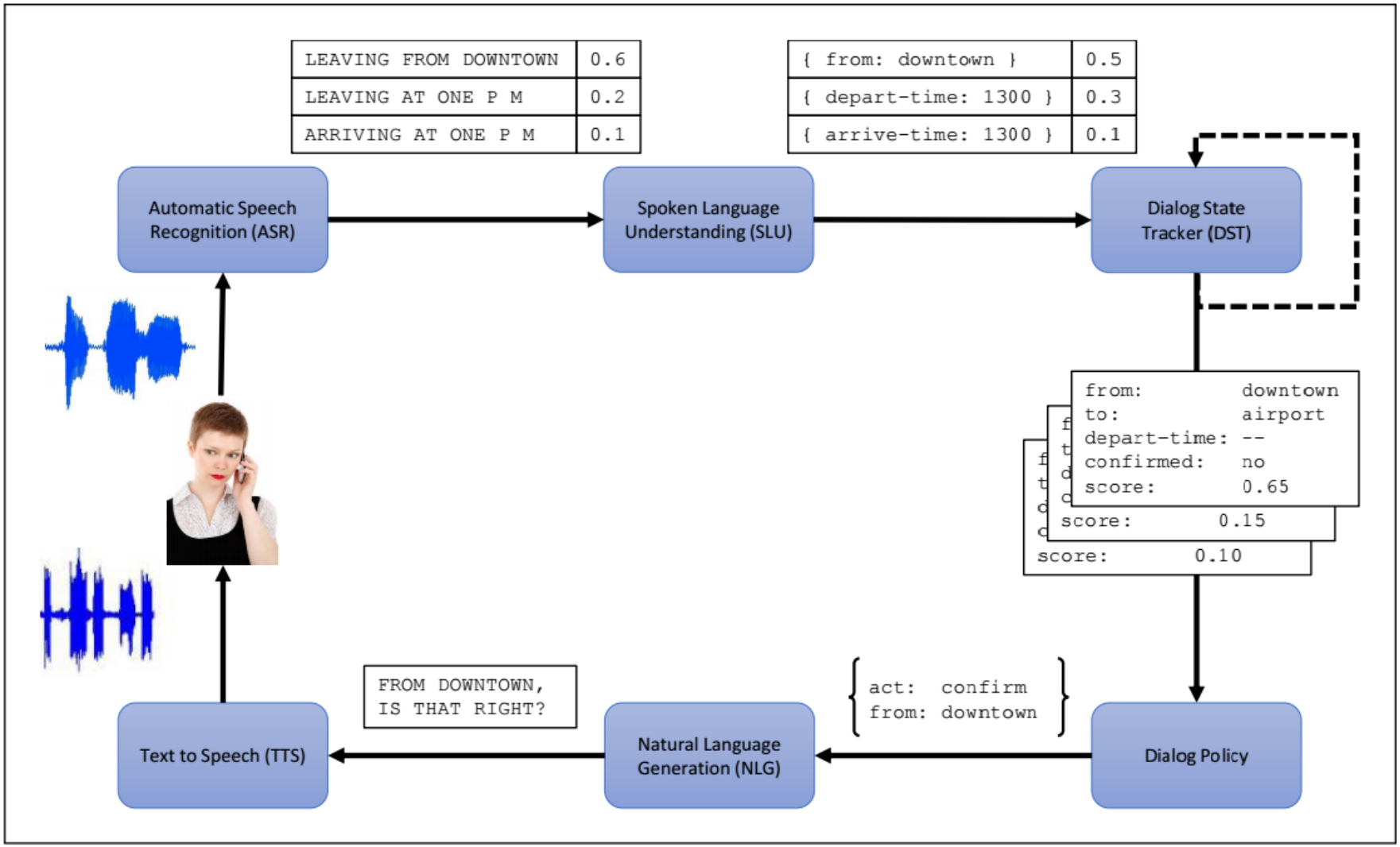


Figure 29.12 Architecture of a dialogue-state system for task-oriented dialogue from (Williams et al., 2016).

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

Possible architectures for dialog management

Finite State

Frame-based

Information State (Markov Decision Process)

Classic AI Planning

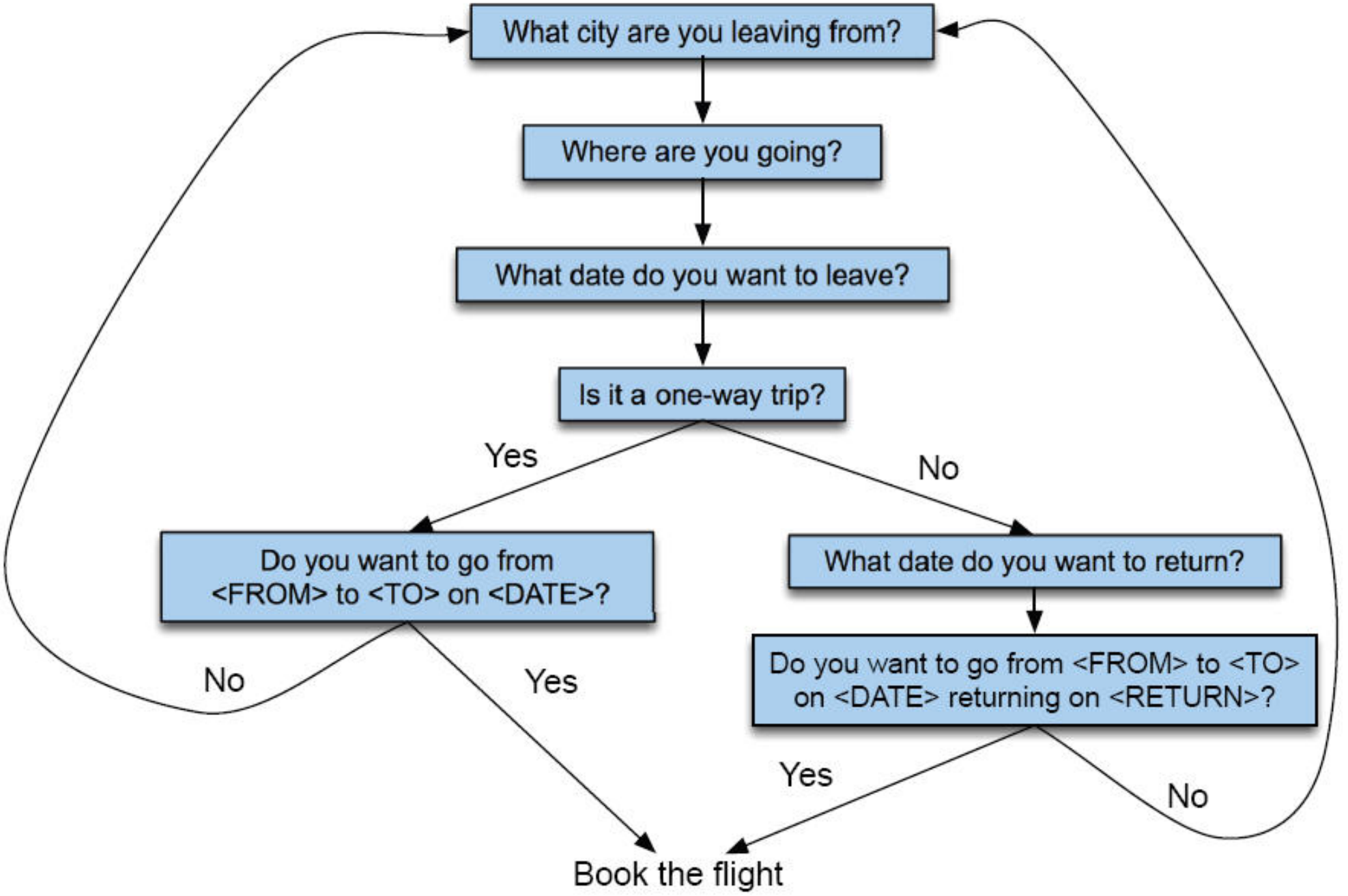
Distributional / neural network

Finite-State Dialog Management

Consider a trivial airline travel system:

- Ask the user for a departure city
- Ask for a destination city
- Ask for a time
- Ask whether the trip is round-trip or not

Finite State Dialog Manager



Finite-state dialog managers

- 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

System completely controls the conversation

- 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)

+

−

- Too limited

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 **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 much flexibility

User Initiative

- User directs the system
  - Asks a single question, system answers
- Examples: **Voice web search**
- But system can't:
  - ask questions back,
  - engage in clarification dialogue,
  - engage in confirmation dialogue

Mixed Initiative

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

An example of a frame

FLIGHT FRAME:

ORIGIN:

CITY: Boston

DATE: Tuesday

TIME: morning

DEST:

CITY: San Francisco

AIRLINE:

…

Mixed Initiative

- Conversational initiative can shift between system and user
- Simplest kind of mixed initiative: use the structure of the **frame** 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?

Frames are 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!
  - Avoids strict constraints on order of the finite-state architecture.

Multiple frames

- flights, hotels, rental cars
- Flight legs: Each flight can have multiple legs, which might need to be discussed separately
- Presenting the flights (If there are multiple flights meeting users constraints)
  - It has slots like 1ST_FLIGHT or 2ND_FLIGHT so user can ask “how much is the second one”
- General route information:
  - Which airlines fly from Boston to San Francisco
- Airfare practices:
  - Do I have to stay over Saturday to get a decent airfare?

Natural Language Understanding

- 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

Semantics for a sentence

LIST FLIGHTS ORIGIN

Show me flights from Boston

DESTINATION DEPARTDATE

to San Francisco on Tuesday

DEPARTTIME

morning

Idea: HMMs for semantics

- Hidden units are slot names

  ORIGIN

  DESTCITY

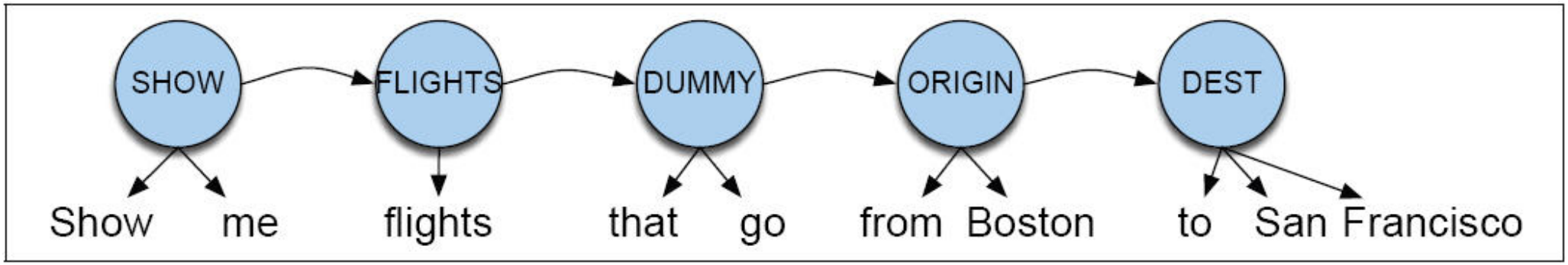
  DEPARTTIME

- Observations are word sequences

  on Tuesday

HMM model of semantics

Pieraccini et al (1991)



Semantic HMM

- Goal of HMM model:

  To compute labeling of semantic roles C = c1,c2,…,cn

  (C for ‘cases’ or ‘concepts’)

  that is most probable given words W

$$\operatorname*{argmax}_C P(C \mid W) = \operatorname*{argmax}_C \frac{P(W \mid C)P(C)}{P(W)}$$

$$= \operatorname*{argmax}_C P(W \mid C)P(C)$$

$$= \operatorname*{argmax}_C \prod_{i=2}^{N} P(w_i \mid w_{i-1} \ldots w_1, C) P(w_1 \mid C) \prod_{i=2}^{M} P(c_i \mid c_{i-1} \ldots c_1)$$

Semantic HMM

- From previous slide:

$$= \underset{C}{\operatorname{argmax}} \prod_{i=2}^{N} P(w_i \mid w_{i-1} \ldots w_1, C) P(w_1 \mid C) \prod_{i=2}^{M} P(c_i \mid c_{i-1} \ldots c_1)$$

- Assume simplification:

$$P(w_i \mid w_{i-1} \ldots w_1, C) = P(w_i \mid w_{i-1}, \ldots, w_{i-N+1}, c_i)$$

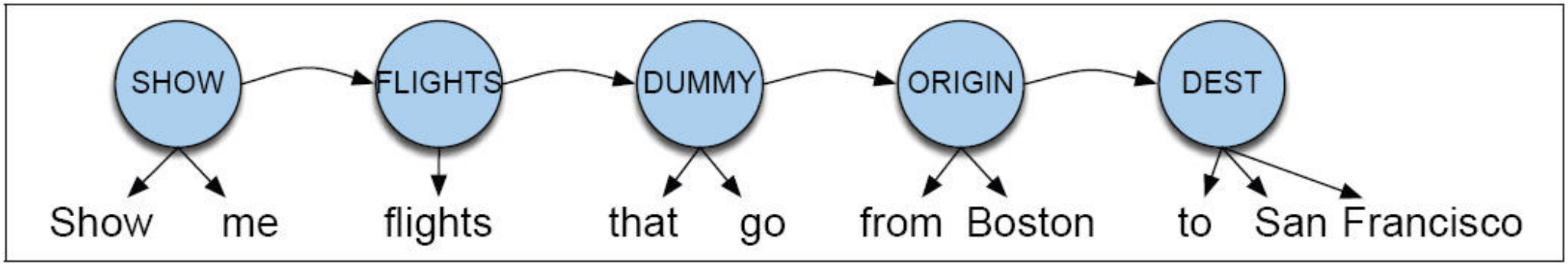
$$P(c_i \mid c_{i-1} \ldots c_1, C) = P(c_i \mid c_{i-1}, \ldots, c_{i-M+1})$$

- Final form:

$$= \underset{C}{\operatorname{argmax}} \prod_{i=2}^{N} P(w_i \mid w_{i-1} \ldots w_{i-N+1}, c_i) \prod_{i=2}^{M} P(c_i \mid c_{i-1} \ldots c_{i-M+1})$$

semi-HMM model of semantics

Pieraccini et al (1991)



P(W|C) =

P(me|show,SHOW) P(show|SHOW) P(flights|FLIGHTS)…

P(FLIGHTS|SHOW) P(DUMMY|FLIGHTS)…

Semi-HMMs

- Each hidden state
  - Can generate multiple observations
- By contrast, a traditional HMM
  - One observation per hidden state
  - Need to loop to have multiple observations with the same state label

How to train

- Supervised training
- Label and segment each sentence with frame fillers
- Essentially learning an N-gram grammar for each slot

LIST	FLIGHTS	DUMMY	ORIGIN	DEST
Show me	flights	that go	from Boston	to SF

Another way to do NLU: 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

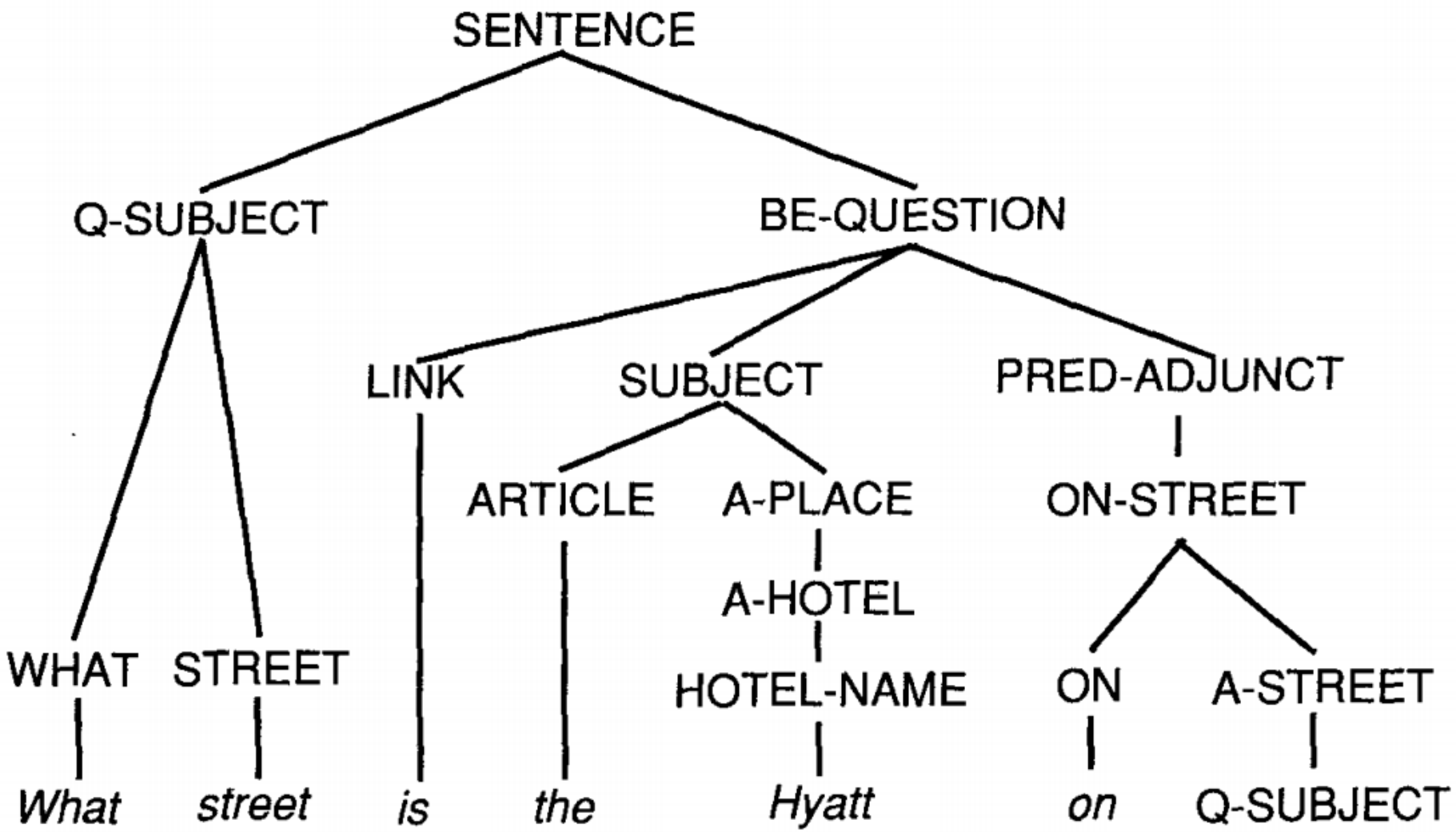
ORIGIN -> from CITY

DESTINATION -> to CITY

CITY -> Boston | San Francisco | Denver | Washington

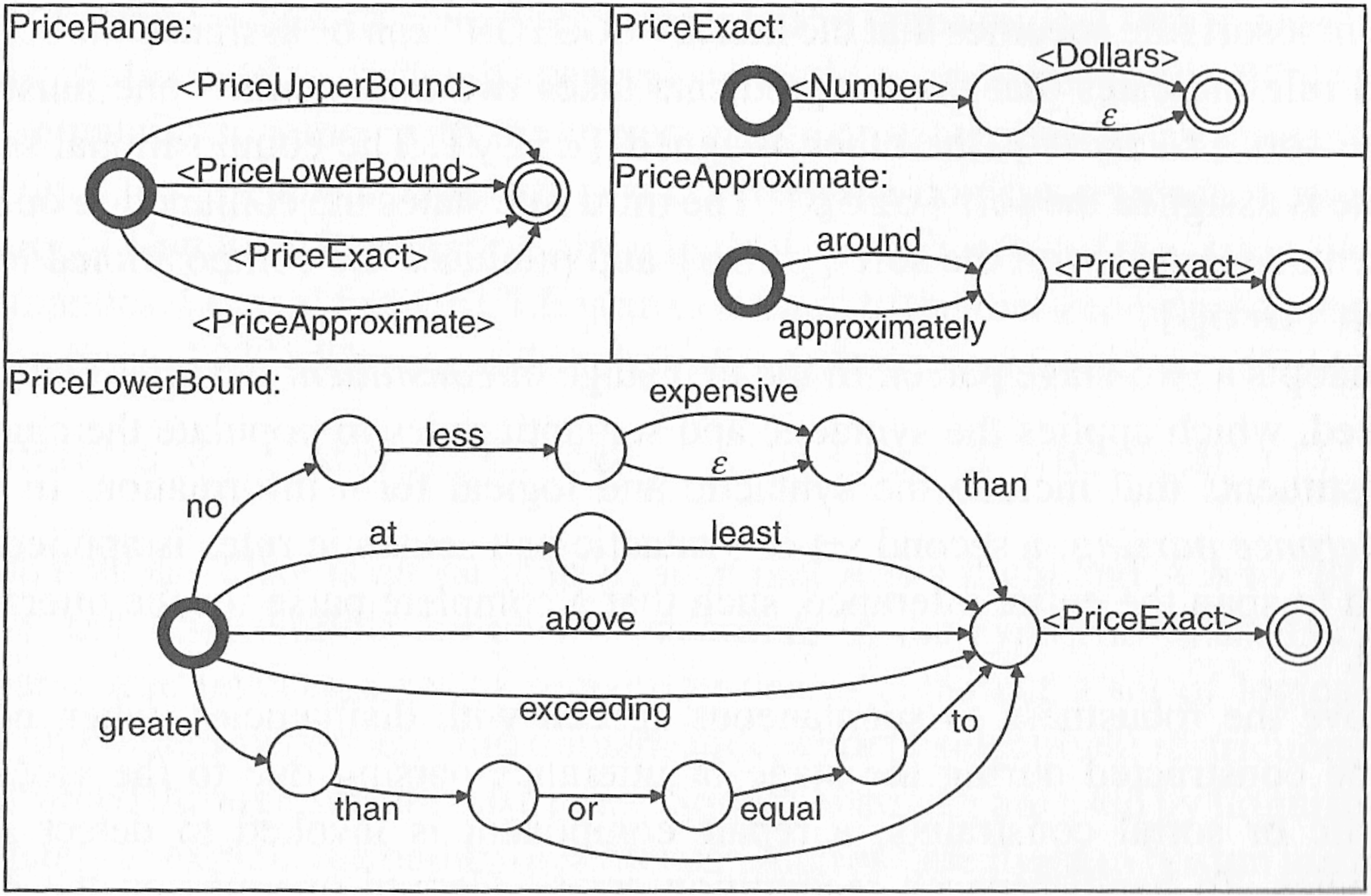
Tina parse tree with semantic rules

Seneff 1992



Phoenix SLU system: Recursive Transition Network

Ward 1991, figure from Wang, Deng, Acero



Modern Approach: Semantic Parsing

- System translates natural language into logical forms
- System can act on structured logical forms
- Modern approaches mix hand engineered grammar generation with machine learning to map input text to output structured form

Semantic Parsing Output: Database Query

- Directly map natural language to database queries
- Potentially time consuming to build/train for a new schema, but a clean, clear formalism

which country had the highest carbon emissions last year

```
SELECT    country.name
FROM      country, co2_emissions
WHERE     country.id = co2_emissions.country_id
AND       co2_emissions.year = 2014
ORDER BY  co2_emissions.volume DESC
LIMIT     1;
```

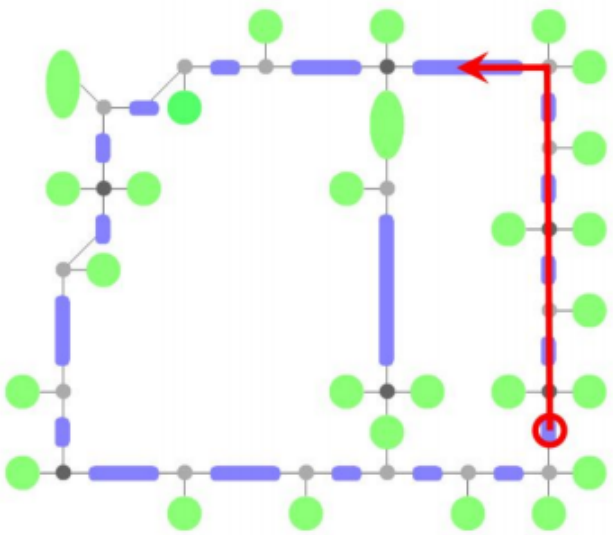
Slide from Bill McCartney CS224U

Semantic Parsing Output: Procedural Languages

- Express concept, nested states or action sequences
- Designing set of possible actions and composition rules can get very complex
- How much can a user reasonably specify in one utterance?

Go to the third junction and take a left.

```
(do-sequentially
  (do-n-times 3
    (do-sequentially
      (move-to forward-loc)
      (do-until
        (junction current-loc)
        (move-to forward-loc))))
  (turn-left))
```



Slide from Bill McCartney CS224U

Semantic Parsing Output: Intents and Arguments

- Personal assistant voice commands are simple and need to scale to many domains
- Simplicity helps with robustness and scale, just recognize what *action* and what required *arguments* for that action

directions to SF by train

```
(TravelQuery
  (Destination /m/0d6lp)
  (Mode TRANSIT))
```

angelina jolie net worth

```
(FactoidQuery
  (Entity /m/0f4vbz)
  (Attribute /person/net_worth))
```

weather friday austin tx

```
(WeatherQuery
  (Location /m/0vzm)
  (Date 2013-12-13))
```

text my wife on my way

```
(SendMessage
  (Recipient 0x31cbf492)
  (MessageType SMS)
  (Subject "on my way"))
```

play sunny by boney m

```
(PlayMedia
  (MediaType MUSIC)
  (SongTitle "sunny")
  (MusicArtist /m/017mh))
```

is REI open on sunday

```
(LocalQuery
  (QueryType OPENING_HOURS)
  (Location /m/02nx4d)
  (Date 2013-12-15))
```

Slide from Bill McCartney CS224U

Semantic Parsing Approach Outline

- Very active area of research
- Define possible syntactic structures using a context-free grammar
- Construct semantics bottom-up, following syntactic structure
- Score parses with a (log-linear) model that was fit on training input, action/output pairs
- Use external annotators to recognize names, dates, places, etc.
- Grammar induction if possible, or lots of grammar engineering

Slide from Bill McCartney CS224U

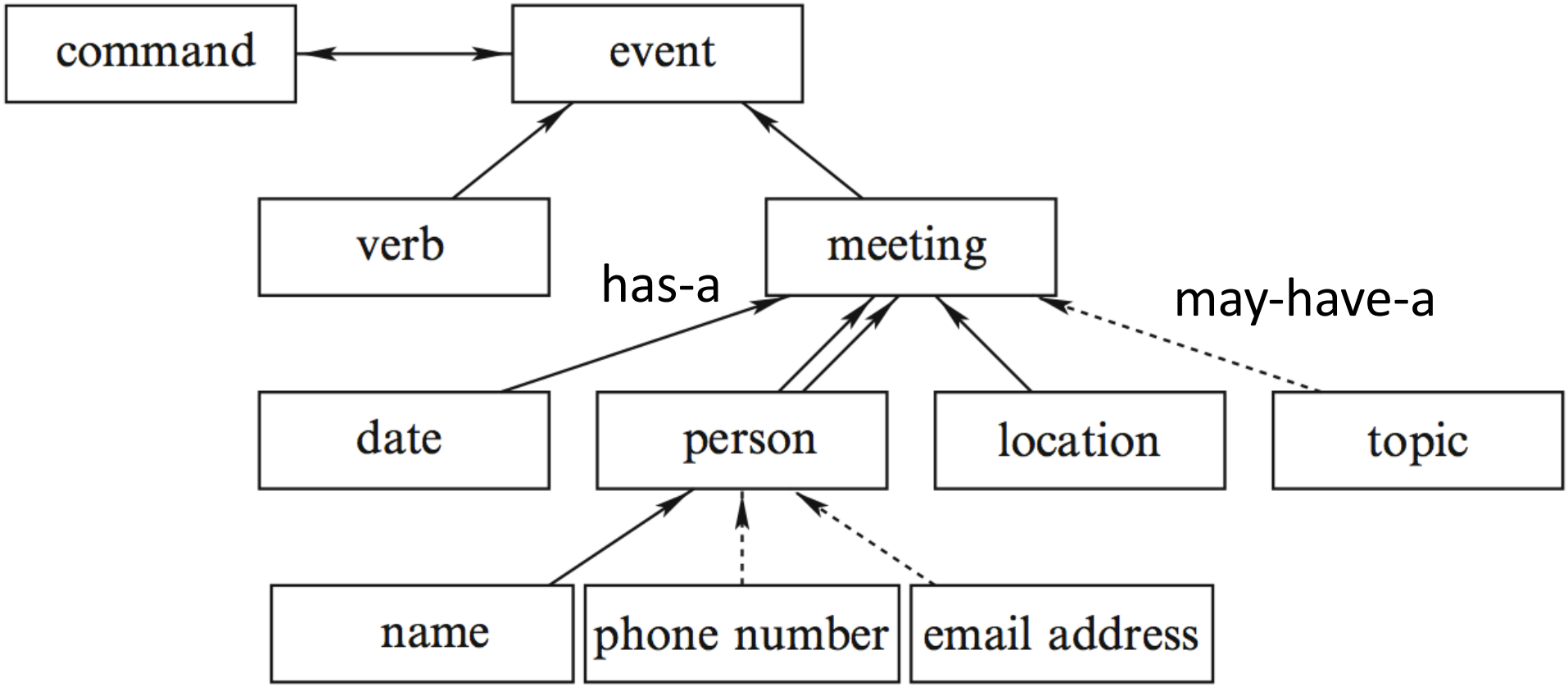
A final way to do NLU: Condition-Action Rules

- Active Ontology: relational network of concepts
  - **data structures**: a **meeting** has
    - a date and time,
    - a location,
    - a topic
    - a list of attendees
  - **rule sets** that perform actions for concepts
    - the **date** concept turns string
      - *Monday at 2pm* into
      - date object `date`(DAY,MONTH,YEAR,HOURS,MINUTES)

Rule sets

- Collections of **rules** consisting of:
  - condition
  - action
- When user input is processed, facts added to store and
  - rule conditions are evaluated
  - relevant actions executed

Part of ontology for meeting task



meeting concept: if you don’t yet have a location, ask for a location

Other components

ASR: Language Models for dialogue

- Often based on hand-written Context-Free or finite-state grammars rather than N-grams
- Why?
  - Need for understanding; we need to constrain user to say things that we know what to do with.

ASR: Language Models for Dialogue

- We can have LM specific to a dialogue state
- If system just asked "What city are you departing from?"
- LM can be
  - City names only
  - FSA: (I want to (leave|depart)) (from) [CITYNAME]
  - N-grams trained on answers to "Cityname" questions from labeled data
- A LM that is constrained in this way is technically called a "restricted grammar" or "restricted LM"

Generation Component

- Content Planner
  - Decides what content to express to user

    (ask a question, present an answer, etc)
  - Often merged with dialogue manager
- Language Generation
  - Chooses syntax and words
  - TTS
- **In practice**: Template-based w/most words prespecified

  What time do you want to leave CITY-ORIG?

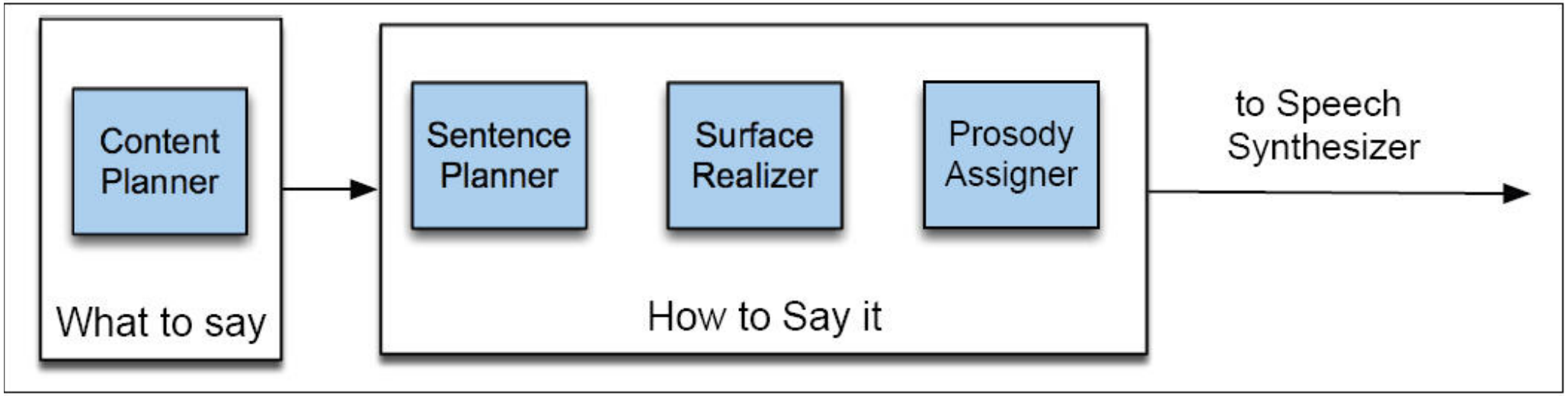
  Will you return to CITY-ORIG from CITY-DEST?

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

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 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: …

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
- 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

Evaluation

1. Slot Error Rate for a Sentence

$$\frac{\text{\# of inserted/deleted/subsituted slots}}{\text{\# of total reference slots for sentence}}$$

2. End-to-end evaluation (Task Success)

Evaluation Metrics

"Make an appointment with Chris at 10:30 in Gates 104"

Slot	Filler
PERSON	Chris
TIME	11:30 a.m.
ROOM	Gates 104

Slot error rate: 1/3

Task success: At end, was the correct meeting added to the calendar?