

CS 6120/CS 4120: Natural Language Processing

Instructor: Prof. Lu Wang

College of Computer and Information Science

Northeastern University

Webpage: www.ccs.neu.edu/home/luwang

What is Natural Language Dialogue?

- Communication involving
 - Multiple contributions
 - Coherent interaction
 - More than one participant
- Interaction modalities
 - Input: Speech, typing, writing, gesture
 - Output: Speech, text, graphical display, animated face/body (embodied virtual agent)

What is involved in NL dialogue

- Understanding
 - What does a person say?
 - Identify words from speech signal
 - “Please close the window”
 - What does the speech mean?
 - Identify semantic content
 - Request (subject: close (object: window))
 - What were the speaker’s intentions?
 - Speaker requests an action in a physical world

What is involved in NL dialogue

- Managing interaction
 - Internal representation of the domain
 - Identify new information
 - Identifying which action to perform given new information
 - “close the window”, “set a thermostat” -> physical action
 - “what is the weather like outside?” -> call the weather API
 - Determining a response
 - “OK”, “I can’t do it”
 - Provide an answer
 - Ask a clarification question

What is involved in NL dialogue

- Access to information
- To process a request “Please close the window” you (or the system) needs to know:
 - There is a window
 - Window is currently opened
 - Window can/can not be closed

What is involved in NL dialogue

- Producing language
 - Deciding when to speak
 - Deciding what to say
 - Choosing the appropriate meaning
 - Deciding how to present information
 - So partner understands it
 - So expression seems natural

When is automatic dialogue system useful?

- When hands-free interaction is needed
 - In-car interface
 - In-field assistant system
 - Command-and-control interface
 - Language tutoring
 - Immersive training
- When speaking is easier than typing
 - Voice search interface
 - Virtual assistant (Siri, Google Now)
- Replacing human agents (cutting cost for companies)
 - Call routing
 - Menu-based customer help
 - Voice interface for customer assistance

Visions of dialogue from science fiction

- Hal “2001: A Space Odyssey” (1968)
 - Naturally conversing computer
- Star Trek (original 1966)
 - Natural language command and control
- Her (2013)
 - A virtual partner with natural dialogue capabilities

Conversational Agents aka Dialogue Agents

Phone-based Personal Assistants

SIRI, Cortana, Google Now

Talking to your car

Communicating with robots

Clinical uses for mental health

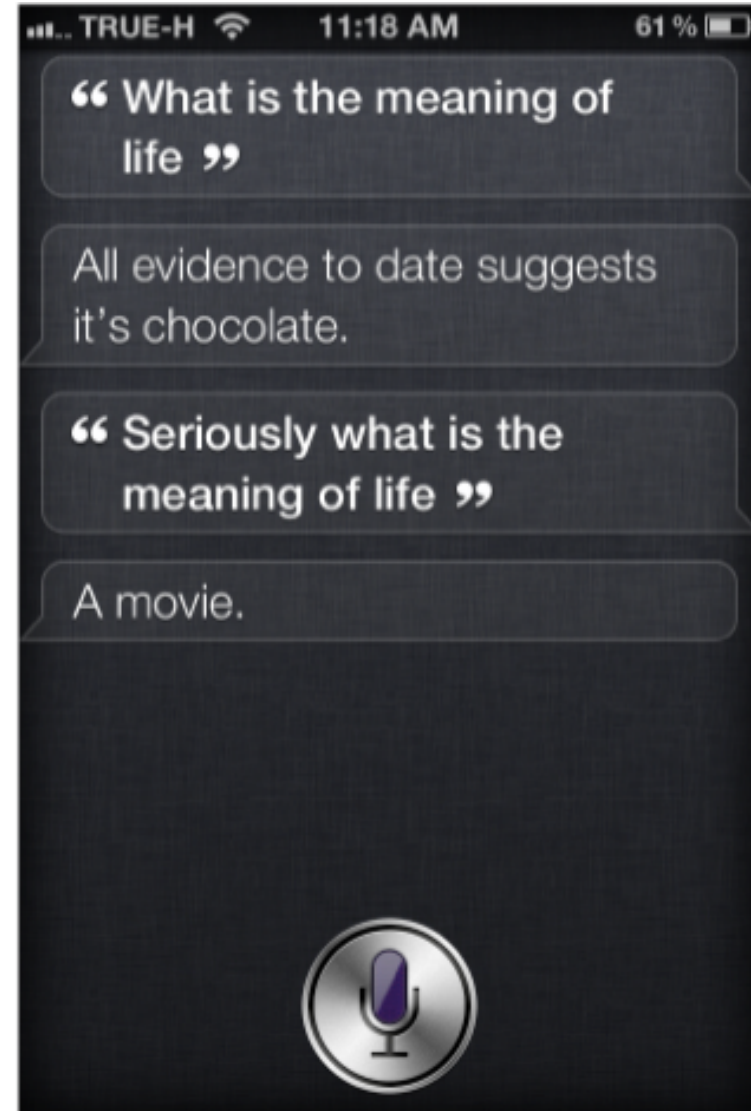
Chatting for fun

Two classes of systems

1. (Goal-based) Dialogue agents
 - *SIRI, interfaces to cars, robots*
 - *booking flights or restaurants*
2. Chatbots

Examples of modern Virtual Assistant dialogue systems

- Apple Siri (2010)
 - Supports questions in a set of domains
 - Answers open-end questions
 - Cute “Easter egg” responses



“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”

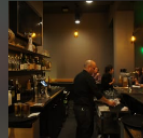


"Find restaurants near me"
tap to edit


Here's what I found:

15 Results
Nearby


Ichi Sushi & Ni Bar
 Japanese · 800 feet · Closed now
 ★★★★★ (311) on Yelp · \$\$\$




Coco's Ramen
 Vietnamese · 800 feet · Closed now
 ★★★★★ (280) on Yelp · \$\$



Blue Plate
 Restaurant · 0.2 miles · Closed now
 ★★★★★ (1251) on OpenTable · \$\$\$



Jasmine Tea House
 Chinese · 900 feet · Closed now



"Tell me more about the first one"
tap to edit

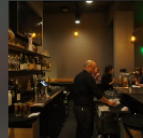
Ichi Sushi & Ni Bar is a Japanese restaurant on Mission St 800 ft to your west. It averages 3½ stars and is somewhat pricey.

"Find restaurants near me"
tap to edit


Here's what I found:

15 Results
Nearby


Ichi Sushi & Ni Bar
 Japanese · 800 feet · Closed now
 ★★★★★ (311) on Yelp · \$\$\$



Coco's Ramen
 Vietnamese · 800 feet · Closed now
 ★★★★★ (280) on Yelp · \$\$



Blue Plate
 Restaurant · 0.2 miles · Closed now
 ★★★★★ (1251) on OpenTable · \$\$\$



Jasmine Tea House
 Chinese · 900 feet · Closed now

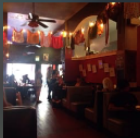


"Are any of them Italian"
tap to edit

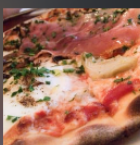
OK, here's what I found:

15 Results
Nearby

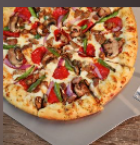
Emmy's Spaghetti Shack
 Italian · 0.2 miles · Closed now
 ★★★★★ (233) on Yelp · \$\$




Vega
 Pizza · 0.2 miles · Closed now
 ★★★★★ (423) on Yelp · \$\$



Pizza Hut
 Pizza · 800 feet
 ★★★★★ (69) on Yelp · \$



La Ciccia
 Sardinian · 0.4 miles



Examples of modern Virtual Assistant dialogue systems

- Android Google Now (2013)
 - Predictive search assistant
- Windows Cortana (2014)
 - Works across different Windows devices
 - Aims to be able to “talk about anything”

Embedded devices with dialogue capabilities

- Amazon Echo (2014) – home assistant device
 - Plays music
 - With voice commands
 - Question answering
 - Get weather, news
 - More complex questions, like
 - “how many spoons are in a cup?”
 - Setting timer
 - Manages TODO lists



Architectures for Practical Dialogue Systems

- ➔ • Finite-State
 - Simple information: e.g., passwords or credit cards*
- Frame-Based
 - All commercial and academic system (SIRI etc.)*

Finite-State Dialog Management

Consider a trivial airline travel system:

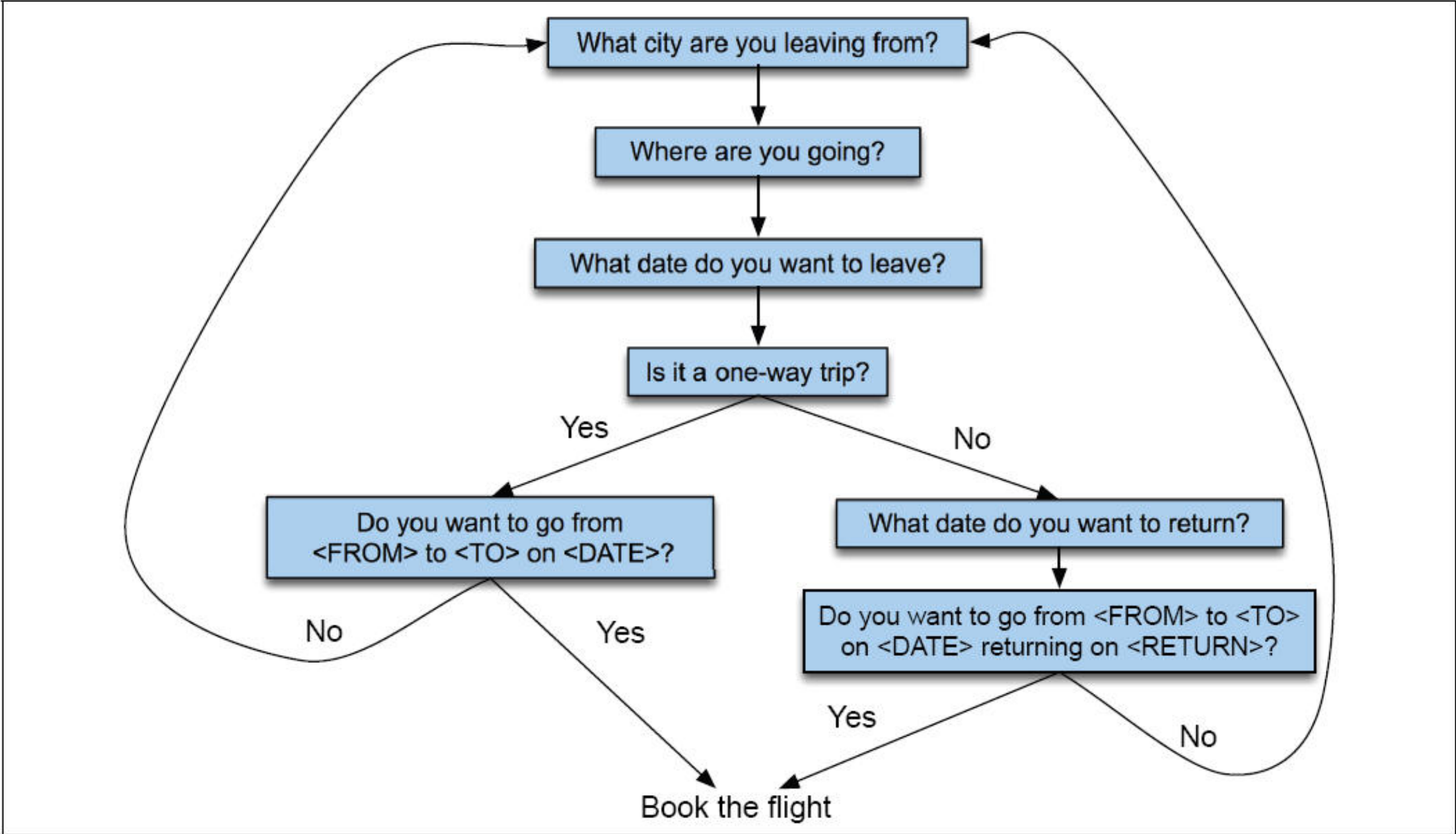
Ask the user for a departure city

Ask for a destination city

Ask for a time

Ask whether the trip is round-trip or not

Finite State Dialog Manager



Finite-state dialogue 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 called **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 (NL understanding)
- 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

Architectures for Practical Dialogue Systems

- Finite-State

Simple information: e.g., passwords or credit cards

- • Frame-Based

All commercial and academic system (SIRI etc.)

Instead, the state of the art:
Frame-based dialogue

- A kind of *mixed initiative*
 - The conversational initiative shifts between system and user
- The structure of the **frame** guides dialogue

Frame-based dialogue

- Invented up the hill in 1977:

GUS, A Frame-Driven Dialog System¹

**Daniel G. Bobrow, Ronald M. Kaplan, Martin Kay,
Donald A. Norman, Henry Thompson and
Terry Winograd**

*Xerox Palo Alto Research Center, 3333 Coyote Hill Road,
Palo Alto, CA 94304, U.S.A.*

Artificial Intelligence Journal, 1977

- Still the state of the art (in real world systems)
 - SIRI based on GUS architecture

The Frame

- A set of slots, to be filled with specific information
- Each associated with a question to the user

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

- 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 can fill 3 slots and not ask these questions again!

The Natural Language Understanding Component

Show me morning flights from Boston to SF on Tuesday.

SHOW:

FLIGHTS:

ORIGIN:

CITY: Boston

DATE: Tuesday

TIME: morning

DEST:

CITY: San Francisco

Often called "dialogue state" detection

Dialogue state: representation of what the user wants at any point in a dialogue

- Which slots got filled in the last sentence?
- What is the current state of the frame?
 - All the values of the filled slots
- What is the user's last "dialogue act":
 - Did they ask me a **question**?
 - **Inform** me of something?

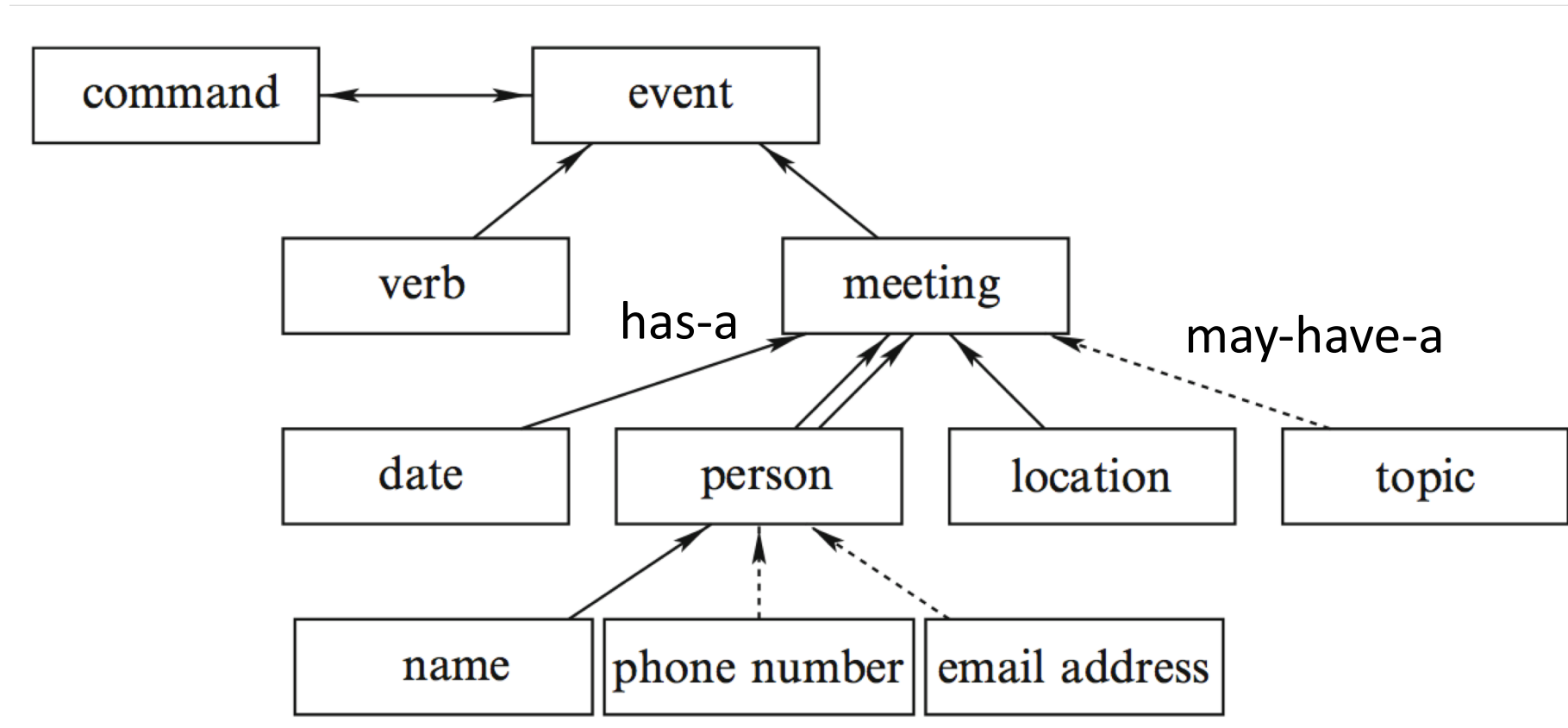
How to do Frame-based Natural Language Understanding?

- Rule-based models
- Statistical models

Siri uses GUS architecture: 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)`

Part of ontology for meeting task



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

Statistical Natural Language Understanding

- Statistical classifiers to map words to semantic frame-fillers
- Given a set of labeled sentences
 - “I want to fly to San Francisco on Tuesday”
Destination: San Francisco
Depart-date: Tuesday
- Requirements: Lots of labeled data

Statistical Slot filling

- Given a sentence:

I want to go from Boston to SF

- Classifier predicts which slot the user wants to fill

Output: (ORIGIN, DESTINATION, DEPARTURE-DATE, AIRLINE)

Statistical Slot filling

- Given a sentence:

I want to go from Boston to SF

- Classifier predicts which slot the user wants to fill

Output: (ORIGIN, DESTINATION, DEPARTURE-DATE, AIRLINE)

Features: Words, Named Entities

- Classifier or sequence model predicts the filler:

ORIGIN DEST

I want to go from Boston to SF

Evaluation

1. Slot Error Rate for a Sentence

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

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

Evaluation

“Make an appointment with David at 11:30 in Rm 905”

Slot	Filler
PERSON	David
TIME	11:30 p.m.
ROOM	Rm 905

Slot error rate: 1/3

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

Dialogue Act Detection

- Dialogue Act: The dialogue function of the sentence
 - Question
 - Command
 - Suggestion
- Given a user's sentence:
 - Context: David wants to eat Italian food.
 - He asked: *How many Italian restaurants are in walking distance?*
- Was that a question?

Dialogue Act detection is hard

Can you give me a list of the flights from Atlanta to Boston?

- This looks like an QUESTION.
 - It has a question-mark, starts with "can you"
- If so, the answer is:
 - YES.
- But really it's a COMMAND, a polite form of:

Please give me a list of the flights...
- What looks like a QUESTION can be a COMMAND

Dialogue Act detection

- Rules-based models
- Statistical models/Machine learning models

Another example of dialogue act detection: Correction Detection

- If system misrecognizes an utterance, and either
 - Rejects
 - Via confirmation, displays its misunderstanding
- Then user has a chance to make a **correction**
 - Repeat themselves
 - Rephrasing
 - Saying “no” to the confirmation question.

Corrections

- Unfortunately, corrections are harder to recognize than normal sentences!
 - Swerts et al (2000): corrections misrecognized twice as often as non-corrections!!!
 - Why?
 - Prosody seems to be largest factor: **hyperarticulation**
 - Liz Shriberg example:
 - “NO, I am DE-PAR-TING from Jacksonville”



Machine learning to detect user corrections: features

- Lexical information (no, correction, I don't, swear words)
- Prosodic indicators of hyperarticulation
 - pause duration, word duration
- Length
- LM probability
- Various dialogue features (repetition)

Deciding what to say:

Dialogue act generation/prediction

Should I do a "*Confirmation*" dialog act?

Natural Language Generation

Given that I'm confirming, what exactly should I say?

Grounding

- Why do elevator buttons light up?
- Clark (1996) (after Norman 1988)
Principle of closure. Agents performing an action require evidence, sufficient for current purposes, that they have succeeded in performing it
- What is the linguistic correlate of this?

Grounding and Confirmation

- We need to know whether an action succeeded or failed
- Talking is an action!
- I need to know if my action succeeded
 - i.e. the hearer understood my turn!

How do speakers ground?

Clark and Schaefer

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

A human-human conversation

C₁: ...I need to travel in May.
A₁: And, what day in May did you want to travel?
C₂: OK uh I need to be there for a meeting that's from the 12th to the 15th.
A₂: And you're flying into what city?
C₃: Seattle.
A₃: And what time would you like to leave Pittsburgh?
C₄: Uh hmm I don't think there's many options for non-stop.
A₄: Right. There's three non-stops today.
C₅: What are they?
A₅: 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.
C₆: OK I'll take the 5ish flight on the night before on the 11th.
A₆: On the 11th? OK. Departing at 5:55pm arrives Seattle at 8pm, U.S. Air flight 115.
C₇: OK.

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

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.

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

Confirmation

- Errors: Speech is a pretty errorful channel
 - Humans use grounding to confirm that they heard correctly
 - Automatic speech recognition (ASR) is way worse than humans!
- Dialogue systems need to do even more grounding and confirmation than humans
 - Users are confused when system doesn't give explicit acknowledgement signal.

Stifelman et al. (1993), Yankelovich et al. (1995)

Explicit confirmation

S: Which city do you want to leave from?

U: Baltimore

S: Do you want to leave from Baltimore?

U: Yes

Explicit confirmation

U: I'd like to fly from Denver Colorado to New York City on September 21st in the morning on United Airlines

S: Let's see then. I have you going from Denver Colorado to New York on September 21st. Is that correct?

U: Yes

Implicit confirmation: display

U: I'd like to travel to Berlin

S: When do you want to travel to Berlin?

U: Hi I'd like to fly to Seattle Tuesday morning

S: Traveling to Seattle on Tuesday, August eleventh in the morning. Your name?

Implicit vs. Explicit

- Complementary strengths
- Explicit: easier for users to correct systems's mistakes (can just say "no")
- But explicit is cumbersome and long
- Implicit: much more natural, quicker, simpler (if system guesses right).

Implicit and Explicit

- Early systems: all-implicit or all-explicit
- Modern systems: adaptive
- How to decide?

Implicit and Explicit

- Early systems: all-implicit or all-explicit
- Modern systems: adaptive
- How to decide?
 - ASR system can give confidence metric.
 - This expresses how convinced system is of its transcription of the speech
 - If high confidence, use implicit confirmation
 - If low confidence, use explicit confirmation

Should also consider cost of an error: Explicit confirmation before moving money or booking flights

Rejection

- “I’m sorry, I didn’t understand that.”
- Reject when:
 - ASR confidence is low
 - Best interpretation is semantically ill-formed (e.g., where is sushi (Sochi))
- Might have four-tiered level of confidence:
 - Below confidence threshold, reject
 - Above threshold, explicit confirmation
 - If even higher, implicit confirmation
 - Even higher, no confirmation