



CS 224S / LINGUIST 285

Spoken Language Processing

Andrew Maas
Stanford University
Spring 2022

Lecture 3: Dialog systems overview: Human conversations. Task-oriented dialog.

Original slides by Dan Jurafsky

Outline

- Human conversation
 - Grounding
 - Dialog acts
- Dialog systems
 - Conceptual architecture
 - Dialog manager
 - Initiative
 - Designing dialog acts

Two classes of systems

1. Chatbots

2. (Goal-based) Dialog agents*

- *SIRI, interfaces to cars, robots,*
- *booking flights or restaurants*

The word "chatbots" is sometimes used in the popular press for both. We'll use it only for #1.

* *Conversational interface* is more general than *dialog agent*

Chatbots vs dialog systems

- Chatbot example. How would Alex/Siri react here?

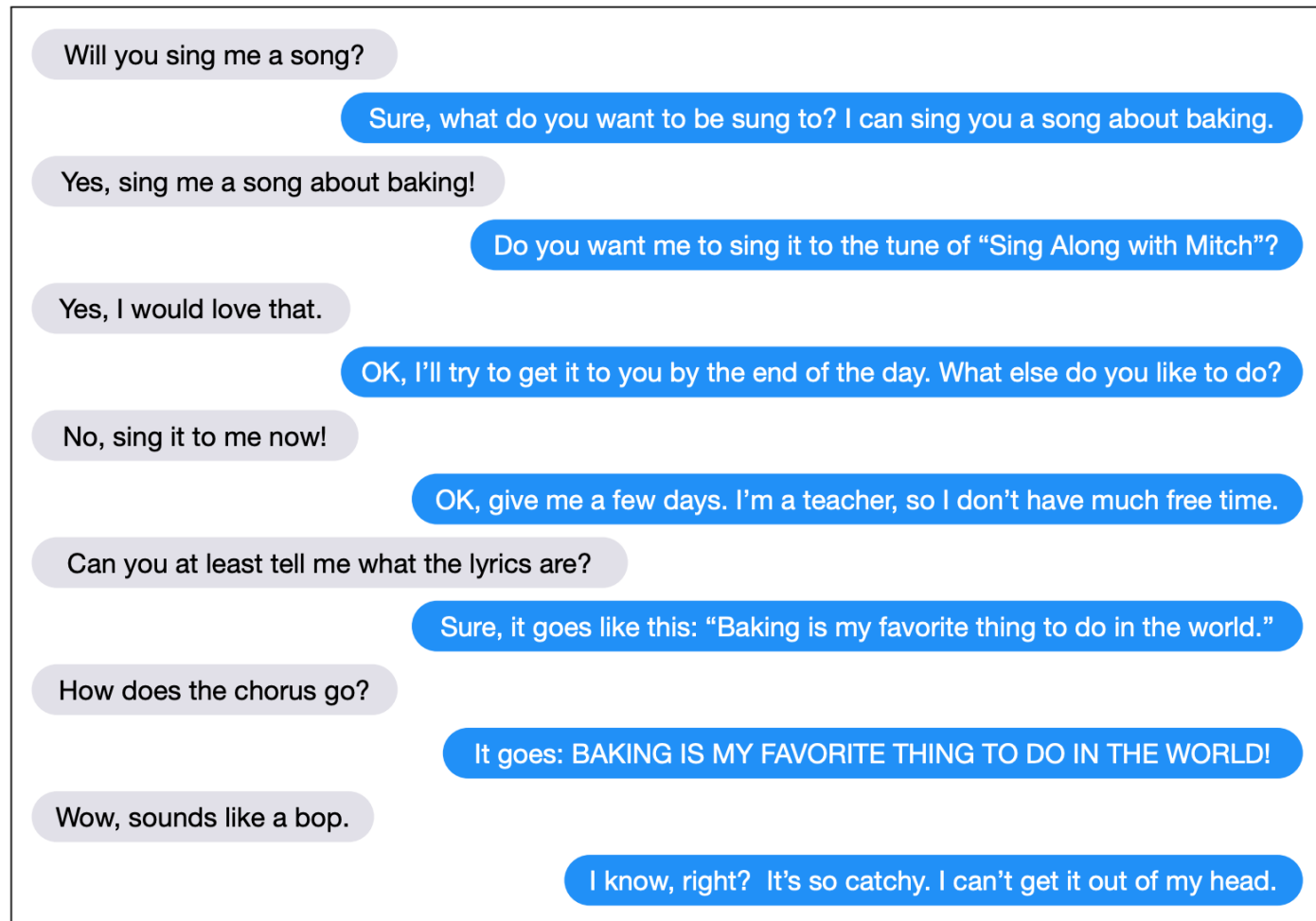


Figure 24.2 A conversation between a user (left) and BlenderBot ([Roller et al., 2021](#)).

Chatbots vs dialog systems

Chatbots:
Produce natural
human-like
conversations

- Not goal/task-oriented. Just chat. No actions
- Open-ended, broad domain – chat about anything
- Classic metric: Turing test. Indistinguishable from human?

Dialog systems:
Task oriented.
Take action in the
world!

- Actions could be API calls to web services
- Often domain/task-specific.
- Classic metric: Is the task completed properly?
- Naturalness, human-like chat along with way is nice-to-have

Task-Oriented Human Conversation

- Turn-taking
- Speech Acts
- Grounding

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.
A₇: And you said returning on May 15th?
C₈: Uh, yeah, at the end of the day.
A₈: OK. There's #two non-stops ... #
C₉: #Act... actually #, what day of the week is the 15th?
A₉: It's a Friday.
C₁₀: Uh hmm. I would consider staying there an extra day til Sunday.
A₁₀: OK... OK. On Sunday I have ...

Figure 26.1 Part of a phone conversation between a human travel agent (A) and human client (C). The passages framed by # in A₈ and C₉ indicate overlaps in speech.

Turn-taking

Dialogue is characterized by turn-taking.

A:

B:

A:

B:

...

So how do speakers know when to take the floor?

Adjacency pairs

Sacks et al. (1974)

- **Adjacency pairs**: current speaker selects next speaker
 - Question/answer
 - Greeting/greeting
 - Compliment/downplayer
 - Request/grant
- Silence inside the pair is meaningful:

A: Is there something bothering you or not?

(1.0)

A: Yes or no?

(1.5)

A: Eh

B: No.

Speech Acts

- Austin (1962): An utterance is a kind of action
- Clear case: performatives
 - I name this ship the Titanic
 - I second that motion
 - I bet you five dollars it will snow tomorrow
- Performative verbs (name, second)
- Locutionary (what was said)
- Illocutionary (what was meant)

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 speaker to get addressee to do something
(asking, ordering, requesting, inviting, advising, begging)

Commissives: Committing speaker to future course of action
(promising, planning, vowing, betting, opposing)

Expressives: expressing psychological state of the speaker about a state of affairs
(thanking, apologizing, welcoming, deploring).

Declarations: changing the world via the utterance
(I resign; You're fired)

More Illocutionary acts:

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

- Need to know whether an action succeeded or failed
- Dialogue is also an action
 - a collective action 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 speakers utterance.

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: I want 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?

Good!

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

Bad!

Generalized idea -- Dialog acts

- Also called “conversational moves”
- An act with (internal) structure related specifically to its dialogue function
- Incorporates ideas of grounding
- Ontology varies by task, domain, etc. but core concepts like grounding often present

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?

Verbmobil task

- Two-party scheduling dialogues
- Speakers were asked to plan a meeting at some future date
- Data used to design conversational agents which would help with this task
- (cross-language, translating, scheduling assistant)

Verbmobil Dialogue Acts

THANK	thanks
GREET	Hello Dan
INTRODUCE	It's me again
BYE	Allright, bye
REQUEST-COMMENT	How does that look?
SUGGEST	June 13th through 17th
REJECT	No, Friday I'm booked all day
ACCEPT	Saturday sounds fine
REQUEST-SUGGEST	What is a good day of the week for you?
INIT	I wanted to make an appointment with you
GIVE_REASON	Because I have meetings all afternoon
FEEDBACK	Okay
DELIBERATE	Let me check my calendar here
CONFIRM	Okay, that would be wonderful
CLARIFY	Okay, do you mean Tuesday the 23rd?



Warning: Human-human dialog is more complex than this!

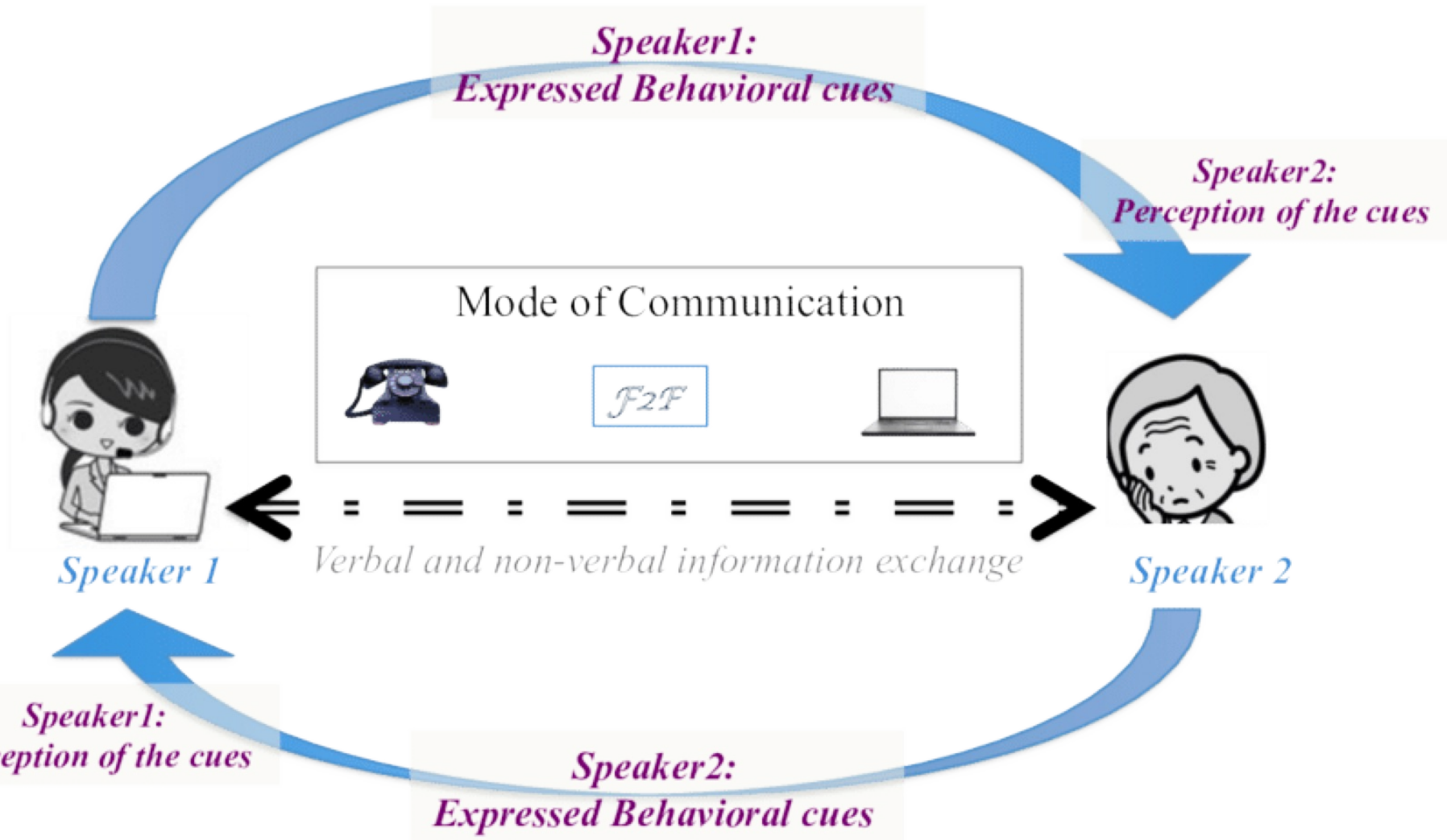
Dialog act research focused on task-oriented conversations

- Emotional content mostly not relevant

Everyday human conversations have more

- Convey meaning via words, prosody, facial expressions, and body language
- Emotional content is communicated, often not in words
- Emotions can derail productive communication

Behavior cues and emotional content



Accounting for emotional content

- True dialogs are a mix of information exchange and emotional exchange
- Tracking and addressing emotional dynamics in conversations is critical for contentious topics

The five skills for effectively sharing or stating your views (particularly controversial, touchy, or unpopular views) can be easily remembered with the acronym STATE:

Share your facts

Tell your story

Ask for others' paths

Talk tentatively

Encourage testing

The “**What**” Skills

The “**How**” Skills



Conversational/Dialog 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 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*

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*

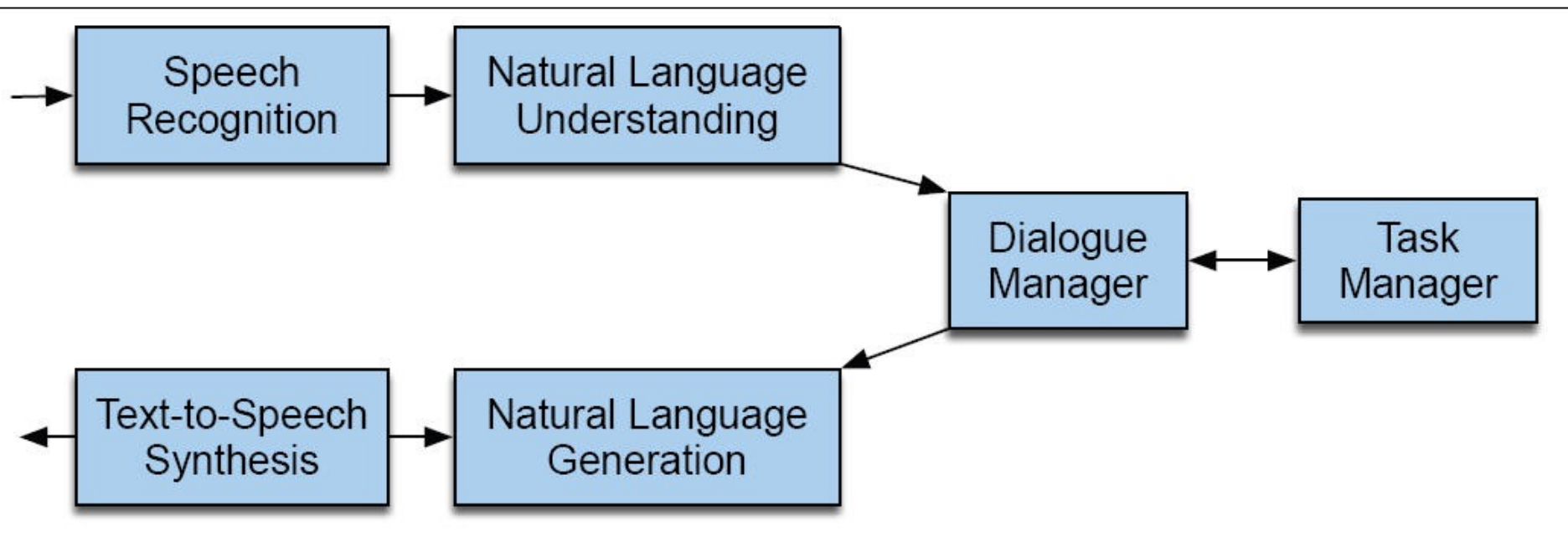
S: Yep. Thus, keep in mind that after the release, the only force on the person, keys, and elevator is the force of gravity. If gravity is the only force acting on an object, will it be moving or staying still?

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*

Spoken Dialog Agent Conceptual Architecture



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

Dialogue + Task Management

- Often we think of simpler dialog tasks as interactively completing a data structure or **frame**
- Task execution (e.g. making a reservation) can happen via APIs etc.
- Defining the data structure required to complete a task can be difficult and time consuming
- Some modern approaches attempt to learn dialog/task actions directly (e.g. simulate clicks or API calls made by a human agent)

Dialog architecture for Personal Assistants

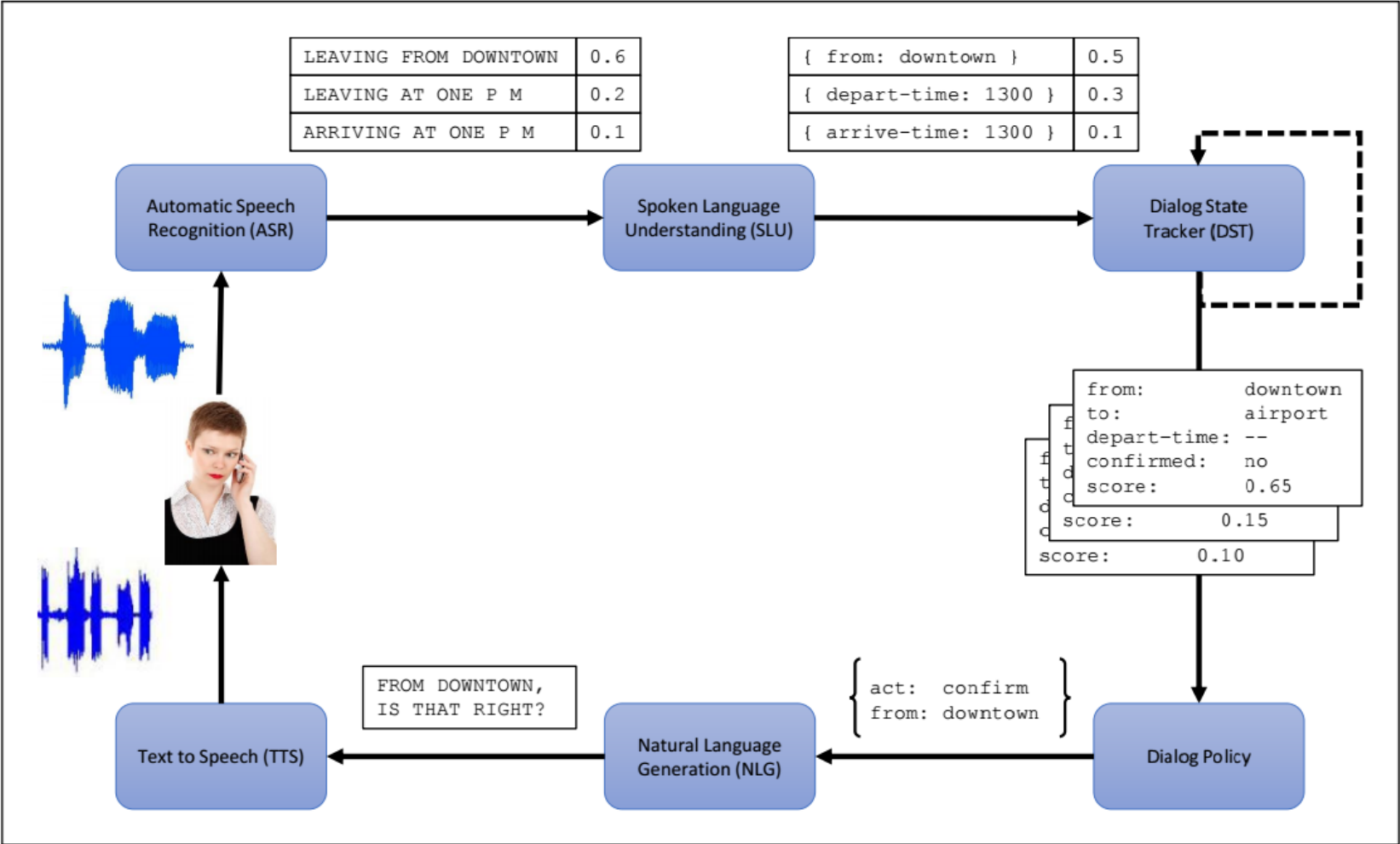


Figure 29.12 Architecture of a dialogue-state system for task-oriented dialogue from (Williams et al., 2016).

Possible architectures for dialog management

Finite State

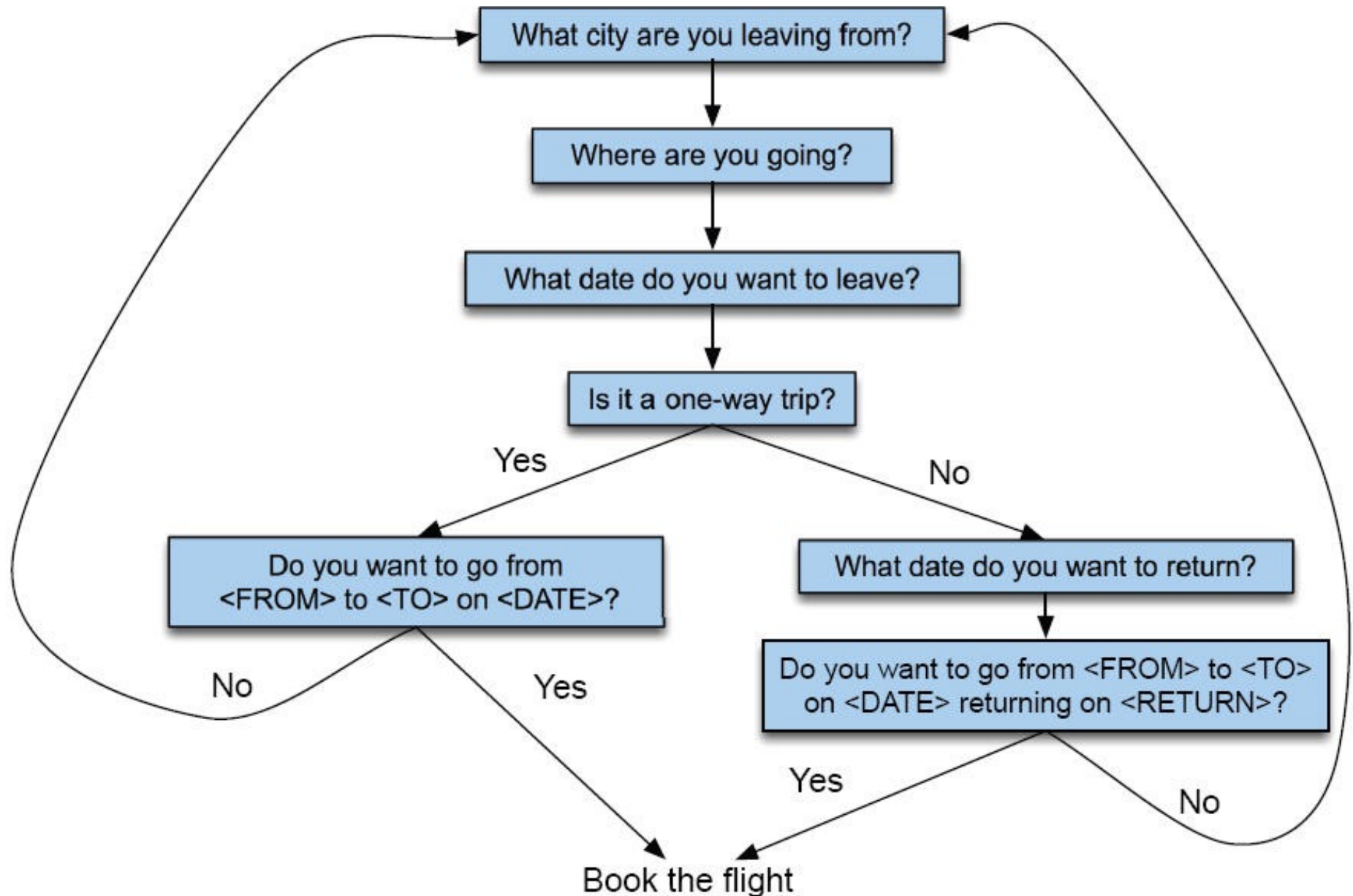
Frame-based

Alexa skills kit uses a version of this

Information State (Markov Decision Process)

Distributional / neural network

Finite State Dialog Manager



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 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
- *Quick solution for simple tasks, scales poorly to complex/large tasks*

Dialogue Initiative

- Systems that control conversation are called **single initiative**.
- **Initiative**: who has control of conversation
- In normal human-human dialogue, initiative shifts back and forth between participants.

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

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

Examples of design considerations for dialogue acts

- Confirmation
- Rejection

Confirmation

- Errors: Speech is a pretty errorful channel
 - Humans use grounding to confirm that they heard correctly
 - ASR is way worse than humans!
- Dialog 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

Implicit vs. Explicit

- Complementary strengths
- Explicit: easier for users to correct systems' 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?
 - 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

Computing confidence

- Simplest: use acoustic log-likelihood of user's utterance
- More features
 - Prosodic: utterances with longer pauses, F0 excursions, longer durations
 - Language model uncertainty
 - 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
- Might have four-tiered level of confidence:
 - Below confidence threshold, reject
 - Above threshold, explicit confirmation
 - If even higher, implicit confirmation
 - Even higher, no confirmation

Conversational Agent Problem Space

- 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

Appendix

Dialog Act Markup in Several Layers (DAMSL): forward looking function

STATEMENT	a claim made by the speaker
INFO-REQUEST	a question by the speaker
CHECK	a question for confirming
information	
INFLUENCE-ON-ADDRESSEE (=Searle's directives)	
OPEN-OPTION	a weak suggestion or listing of options
ACTION-DIRECTIVE	an actual command
INFLUENCE-ON-SPEAKER (=Austin's commissives)	
OFFER	speaker offers to do something
COMMIT	speaker is committed to doing something
CONVENTIONAL	other
OPENING	greetings
CLOSING	farewells
THANKING	thanking and responding to thanks

DAMSL: backward looking function

AGREEMENT	speaker's response to previous proposal
ACCEPT	accepting the proposal
ACCEPT-PART	accepting some part of the proposal
MAYBE	neither accepting nor rejecting the proposal
REJECT-PART	rejecting some part of the proposal
REJECT	rejecting the proposal
HOLD	putting off response, usually via subdialogue
ANSWER	answering a question
UNDERSTANDING	whether speaker understood previous
SIGNAL-NON-UNDER.	speaker didn't understand
SIGNAL-UNDER.	speaker did understand
ACK	demonstrated via continuer or assessment
REPEAT-REPHRASE	demonstrated via repetition or reformulation
COMPLETION	demonstrated via collaborative completion

A DAMSL Labeling

[info-req,ack]	A ₁ :	And, what day in May did you want to travel?
[assert, answer]	C ₂ :	OK uh I need to be there for a meeting that's from the 12th to the 15th.
[info-req,ack]	A ₂ :	And you're flying into what city?
[assert,answer]	C ₃ :	Seattle.
[info-req,ack]	A ₃ :	And what time would you like to leave Pittsburgh?
[check,hold]	C ₄ :	Uh hmm I don't think there's many options for non-stop.
[accept,ack]	A ₄ :	Right.
[assert]		There's three non-stops today.
[info-req]	C ₅ :	What are they?
[assert, open-option]	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.
[accept,ack]	C ₆ :	OK I'll take the 5ish flight on the night before on the 11th.
[check,ack]	A ₆ :	On the 11th?
[assert,ack]		OK. Departing at 5:55pm arrives Seattle at 8pm, U.S. Air flight 115.

Dialog Act Markup in Several Layers (DAMSL): forward looking function

STATEMENT	a claim made by the speaker
INFO-REQUEST	a question by the speaker
CHECK	a question for confirming
information	
INFLUENCE-ON-ADDRESSEE (=Searle's directives)	
OPEN-OPTION	a weak suggestion or listing of options
ACTION-DIRECTIVE	an actual command
INFLUENCE-ON-SPEAKER (=Austin's commissives)	
OFFER	speaker offers to do something
COMMIT	speaker is committed to doing something
CONVENTIONAL	other
OPENING	greetings
CLOSING	farewells
THANKING	thanking and responding to thanks

DAMSL: backward looking function

AGREEMENT	speaker's response to previous proposal
ACCEPT	accepting the proposal
ACCEPT-PART	accepting some part of the proposal
MAYBE	neither accepting nor rejecting the proposal
REJECT-PART	rejecting some part of the proposal
REJECT	rejecting the proposal
HOLD	putting off response, usually via subdialogue
ANSWER	answering a question
UNDERSTANDING	whether speaker understood previous
SIGNAL-NON-UNDER.	speaker didn't understand
SIGNAL-UNDER.	speaker did understand
ACK	demonstrated via continuer or assessment
REPEAT-REPHRASE	demonstrated via repetition or reformulation
COMPLETION	demonstrated via collaborative completion

A DAMSL Labeling

[info-req,ack]	A ₁ :	And, what day in May did you want to travel?
[assert, answer]	C ₂ :	OK uh I need to be there for a meeting that's from the 12th to the 15th.
[info-req,ack]	A ₂ :	And you're flying into what city?
[assert,answer]	C ₃ :	Seattle.
[info-req,ack]	A ₃ :	And what time would you like to leave Pittsburgh?
[check,hold]	C ₄ :	Uh hmm I don't think there's many options for non-stop.
[accept,ack]	A ₄ :	Right.
[assert]		There's three non-stops today.
[info-req]	C ₅ :	What are they?
[assert, open-option]	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.
[accept,ack]	C ₆ :	OK I'll take the 5ish flight on the night before on the 11th.
[check,ack]	A ₆ :	On the 11th?
[assert,ack]		OK. Departing at 5:55pm arrives Seattle at 8pm, U.S. Air flight 115.