

CS 224S / Linguist 285

Spoken Language Processing

Andrew Maas | Stanford University | Spring 2025

**Lecture 7: Dialog Systems Overview:
Human conversations. Task-oriented dialog.**

Outline

- Human Conversation. Aspects of task-oriented dialog
- Dialog system conceptual architecture & design considerations
- Foundational dialog management architectures
- Frame-based dialog systems (hands-on with these next class)

(Task-oriented) human conversation

Example task-oriented dialog

- Turn-taking
- Speech Acts
- Grounding
- Inference

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 2: 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

What defines success in task-oriented conversations?

- **Task completion / success rate.** Did the user achieve their goal? The top-level “did it actually work”
- **Accuracy / slot error rate.** Was the completed request accurate for each request / API call slot?
- **Efficiency metrics.** Number of turns to complete the task. Total conversation time
 - Mirrors call center metrics like Average Handling Time (AHT) and First Call Resolution (FCR)
- **User satisfaction.** Did the user rate this interaction as subjectively “good”?
 - Often collected via post-call surveys (e.g., “Was your issue resolved today?”)
 - Can be inferred from indirect signals (e.g., repeated queries, frustration cues)

Turn-taking

- Dialogue is characterized by turn-taking:

- A: "..."
- B: "..."
- A: "..."
- B: "..."
- ...

- So how do speakers know when to take the floor?
- Implicit task: Choosing when is appropriate to take a conversational turn

Studying turn-taking: 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/Dialog 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)

Five 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)

A different proposal for speech act taxonomy

Bach & Harnish (1979)

- **Constantives:** committing the speaker to something's being the case
 - (answering, claiming, confirming, denying, disagreeing, stating)
- **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)
- **Acknowledgements:** Express the speaker's attitude regarding the hearer with respect to some social action
 - (apologizing, greeting, thanking, accepting an acknowledgment)

Generalized Idea - Dialog Acts

- Also called “conversational moves”
- An act with (internal) structure related specifically to its dialogue function
- Ontology varies by task, domain, etc. but core concepts like grounding often present
- Many possible proposals. Core concepts help us describe conversations.
 - When building a system you will often develop/use custom dialog acts relevant to system goals)

Dialogue Acts from Verbmobil task/dataset

THANK	Thanks
GREET	Hello Dan
INTRODUCE	It's me again
BYE	All right, 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?

Grounding

- One of the most critical forms of dialog acts we build into systems
- 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 speaker's utterance.
- 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”

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

Example dialog revisited

Look for grounding:

- Continued attention
- Relevant next contribution
- Acknowledgement
- Demonstration
- Display

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 2: 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

Grounding Examples

Display:

- **Caller:** I need to travel In May
- **Agent:** And, what day **in May** did you want to travel?

Acknowledgement:

- **Caller:** I want to fly from Boston
- **Agent:** **mm-hmm**
- **Caller:** to Baltimore Washington International

Grounding Examples

Acknowledgement + next relevant contribution:

- **Caller:** I need to travel In May
- **Agent:** **And**, what day in May did you want to travel?

- **Caller:** I will be flying into Nebraska the weekends prior
- **Agent:** **And**, you're flying into what city?

- **Caller:** I can leave on the 17th
- **Agent:** **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

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?



Explicit Confirmation

- **Caller:** I'd like to fly from Denver Colorado to New York City on September 21st in the morning on United Airlines
- **Agent:** Let's see then. I have you going from Denver Colorado to New York on September 21st. Is that correct?
- **Caller:** Yes

Implicit Confirmation: Display

- **Caller:** I'd like to travel to Berlin
- **Agent:** When do you want to travel to Berlin?

- **Caller:** Hi I'd like to fly to Seattle Tuesday morning
- **Agent:** Traveling to Seattle on Tuesday, August eleventh in the morning. Your name?

Dialog system conceptual architecture

Achieving task-oriented conversations with
computer agents

Two Classes of Systems



Chatbots

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



Dialog Systems*

Goal based: Digital assistants.
Car/robot/system commands. Call center / phone task bots

** Conversational interface is more general than dialog agent

Chatbots vs Dialog Systems

This is a chatbot example

Fun conversation, not tied to actions in the world or grounded in factual information in all cases

LLM chatbots are rich, but often not connected to real-world actions

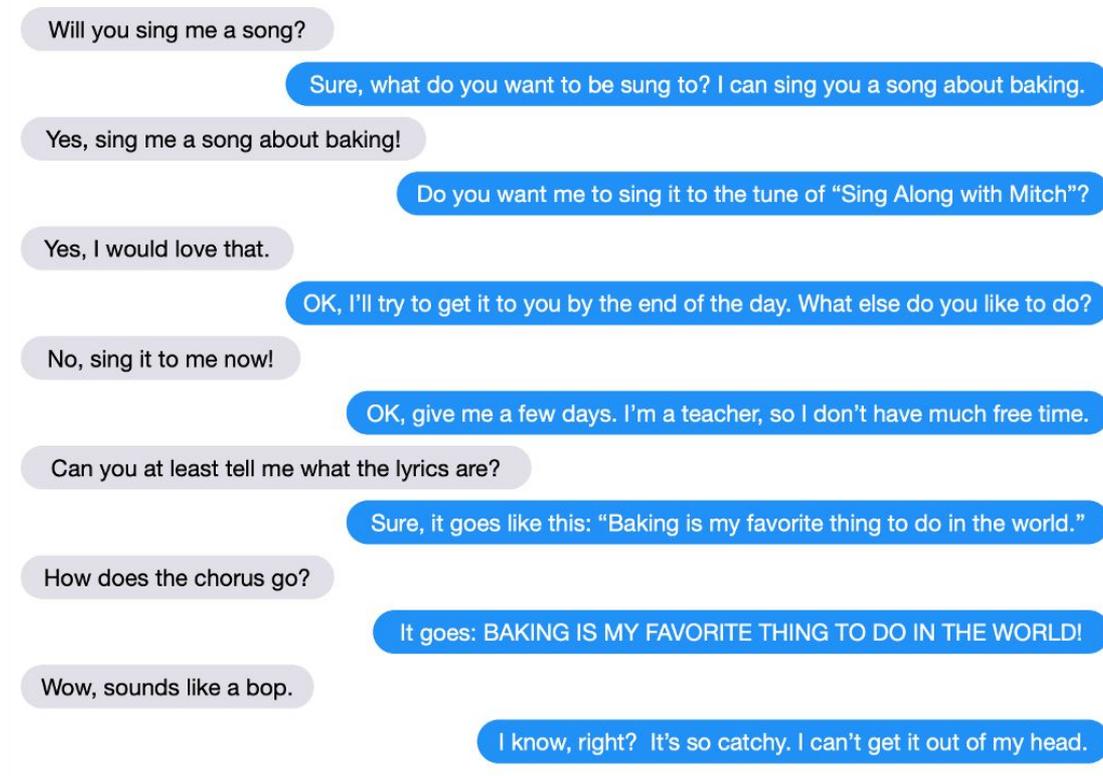


Figure 1: Conversation between user (left) and BlenderBot (Roller et al., 2021)

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



Dialogue 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

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*

Homework reminder / organizational tutor for students

[Pointable.ai blog](#) (2024)

Student: Ok is there anything for Spanish?

HomeworkBot:

Let me take a look for you. It does look like you have an assignment today but that it was in class I believe.

Examencito- vocab. 'Los Días Inolvidables'

Due Oct 9 at 11:59pm. You scored 43/55

Student: Aww :(I was hoping to get a B, but to be fair, it was really confusing

HomeworkBot: Nice! On Friday you also have a quiz in that class on "How the Sun Works"

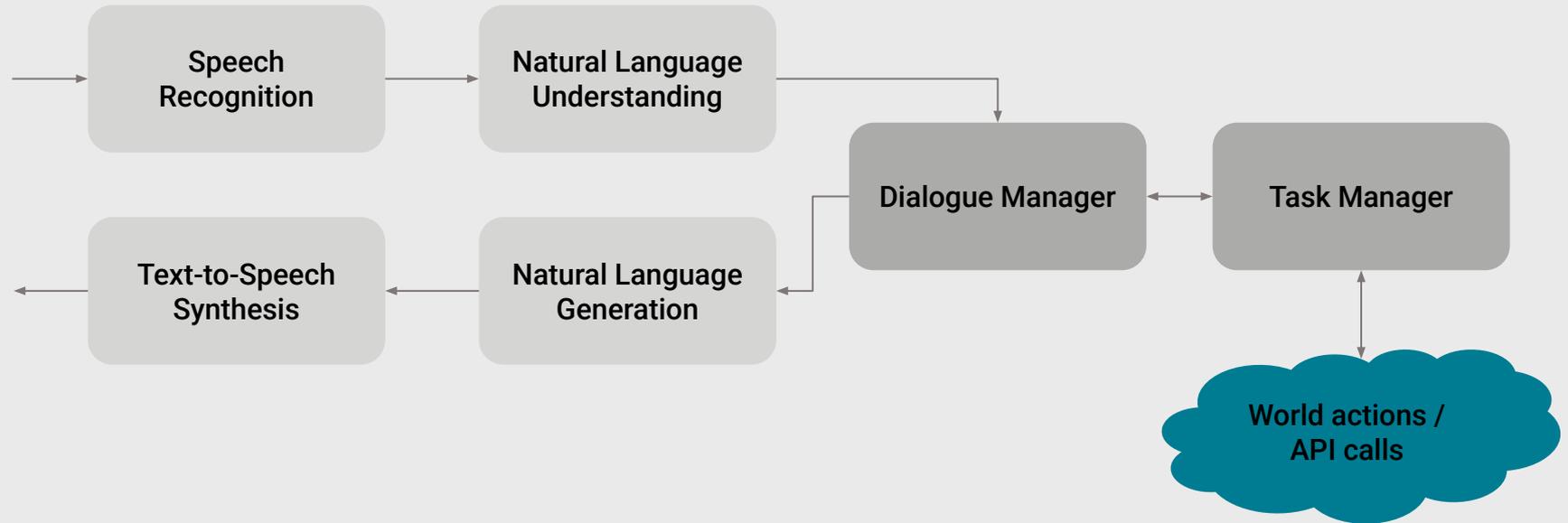
Student: I'm excited. Do I have any zeros outside of geometry which is to be ignored? In geometry, if it is classified under homework assignment it counts as 0% towards our grade I usually only do them if I feel I need extra help

HomeworkBot: I see a few zeros in Earth & Space Systems. You may want to take a look at those. Also English as well. Were there any unexpected issues in completing any of those assignments?

Student: What do I have for English?

HomeworkBot: There are some items marked "late" for The Most Dangerous Game Essay assignment. And a "missing" for Anticipation Guide for Wes Moore's The Other Wes Moore which was due today.

Dialog system information pipeline view



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)
- Something must play dialog manager role in order to:
 - Control the architecture and structure of dialogue. Decide what it knows, what to say next
 - Takes input from ASR/NLU component and connect it with structured task information/state
 - Maintain some sort of state of task and structured data (not just conversation history)
 - Choose actions and send information to NLG to produce a response. Take actions via task interfaces

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*

Dialog system conceptual architecture:

Design choices of conversational interactions. Initiative & dialog act sets

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
- Design choice: When building a dialog system you decide what interaction style it should have, and balance desired user experience with implementation challenges

User Initiative

- **User directs the system**
 - Asks a single question, system answers
- **Examples: Voice web search, voice commands (“Turn on the lights”, “Call David”)**
- **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
 - o Known words: Better performance from ASR
 - o 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:

- **Caller:** Hi, I'd like to fly from Seattle Tuesday morning

- **Caller:** I want a flight from Milwaukee to Orlando one way leaving after 5 p.m. on Wednesday.

Single Initiative (voice commands) + Universals

- User issues single request/command. System processes and responds as needed
- We can give users a little more flexibility by adding universals:
Single-initiative-type commands you can say anywhere during dialog
- As if we augmented every state of a state machine with these:
 - Help
 - Start over
 - Correct
- This describes many implemented systems, 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

- **Agent:** Which city do you want to leave from?
- **Caller:** Baltimore
- **Agent:** Do you want to leave from Baltimore?
- **Caller:** Yes

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)
- Early systems: all-implicit or all-explicit. Modern systems: adaptive
- How to decide when to be explicit?
 - ASR system can give confidence metric on its transcription of the speech
 - If high confidence, use implicit confirmation. If low confidence, use explicit confirmation
 - Dialog manager system sets a *policy* for when to use implicit vs explicit dialog acts

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

Dialog act sets. Choosing what to label. DAMSL example

[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 sub-dialogue
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

Conversational Agent Problem Space

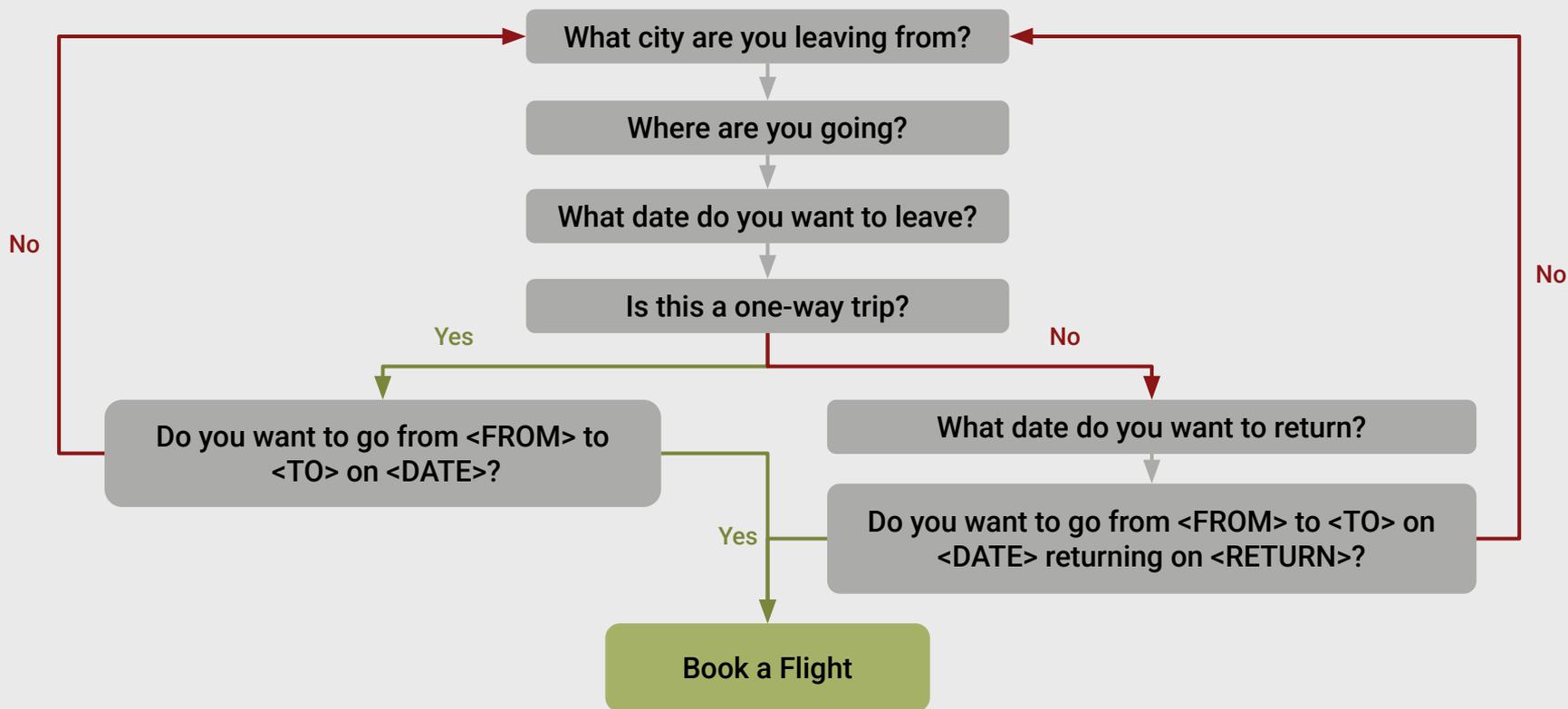
- Time to response (Synchronous?)
- Task complexity
 - What time is it?
 - Book me a flight and hotel for vacation in Greece
- Interaction complexity / number of turns
 - Single command/response
 - “I want new shoes” What kind? What color? What size?
- Initiative
 - User, System, Mixed
 - Interaction modality
- Purely spoken, Purely text, Mixing speech/text/media
- Think about the interaction needs for a particular task/system. Implementation is very different depending on what point in the problem space you target

Foundational dialog management architectures

Possible Architectures for Dialog Management

- **Finite State**
 - Resurgence in utility because LLM-based modules allow simple state machines to do a lot overall
- **Frame-based**
 - Alexa skills kit uses a version of this
- **Information State (Markov Decision Process)**
 - Facilitates reinforcement learning formulation for dialog control. Some similarities w/ state machines
- **LLM / deep learning systems**
 - Active research area, no single approach. More lectures on this later in quarter

Finite State Dialog Manager



Finite-State Dialog Managers

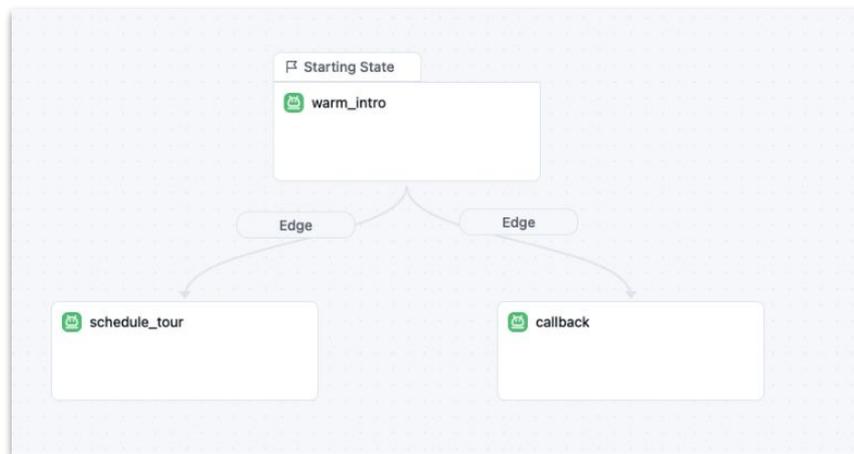
- System completely controls the conversation with the user.
- It asks the user a series of questions
- Ignoring (or misinterpreting) anything the user says that is not a direct answer to the system's questions
- Quick solution for simple tasks, scales poorly to complex/large tasks
- Consider a trivial 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

State machine with separate LLM prompt / module per state

High-level state machines are manageable to build, allow enforcing top-level concepts or flow

LLM-based systems means each state can be a context-specific LLM (+API / context population)

Example: [Retell AI](#) dialog flow builder



The screenshot shows the Retell AI dialog flow builder interface. The top bar includes 'warm_intro' and 'Test Audio / Test LLM' buttons. The main area is divided into sections: 'Prompt' (containing background information and a task list), 'Tools (Optional)' (with an 'Add' button), and 'Call Transfer' (with fields for Name, Description, and Transfer to). A 'Test your agent' button is visible on the right side.

Prompt
Read [stateful multi-prompt agent best practices](#).

Background about Property
Location: 524 Retell Lane
The property is located in the Palo Alto community, near Stanford.
It's a 4b4b with a large backyard. 3000 square feet in total.

Task
1. First, extend a warm greeting to your customer, Sam. Ask if this is a good time to call.
- if this is not a good time, transition to callback.
2. Inform Sam that the reason for your call is that they recently expressed interest in property 524 Retell Lane. Therefore, you're reaching out to see how you can further assist them.
3. Inquire about their thoughts on the property.

Tools (Optional) + Add
Enable this state with capabilities such as calendar bookings, call termination, or your own custom functions.

Call Transfer
Name: transfer_call
Description (Optional): Describe when the call should be transferr
Transfer to: +19257309523

GUS and Frame-Based Systems

Frame-Based Dialogue Agents

- 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.*

- Still the industrial state of the art (LLM-based systems not reliable or controllable enough yet)
- Alexa roughly follows GUS architecture

The Frame

- A set of **slots**, to be filled with information of a given **slot type**
- Each associated with a **question** to the user

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

System Design Considerations

- One or more **frames**
 - Each a collection of **slots**
 - Each slot having a **value**
- **Domain ontology** is full collection of frames for a (sub)system
- A frame defines the task data structure

The State of the Art in 1977

GUS: Hello. My name is GUS. I can help you plan a simple trip by air. Where do you want to go ?

Client: I want to go to San Diego on May 28

GUS: What time do you want to leave?

Client: I must be in San Diego before 10 am

GUS: Would you like P.S.A. flight 102 that arrives at 9.15 am ?

Client: What is the next flight ?

GUS: Air California flight 310 that leaves at 8.30am and arrives at 10.00am

Client: I'll take the first one

GUS: In what name should I make a reservation, please?

Client: Dan Bobrow

GUS: I have confirmed the following flight: P.S.A. flight 102 on Wednesday May 28 from San Jose to San Diego that leaves at 7.30 am and arrives at 9.15 am.

GUS: What date do you want to return on ?

Client: On Friday in the evening.

GUS: Would you like the flight that leaves at 7.45 pm ?

Client: That's fine.

GUS: I have confirmed the following flight: P.S.A. flight 307 on Friday May 30 from San Diego to San Jose that leaves at 7.45 pm and arrives at 9.30 pm Thank you for calling. Goodbye

Dialog pipeline with explicit frame data structures

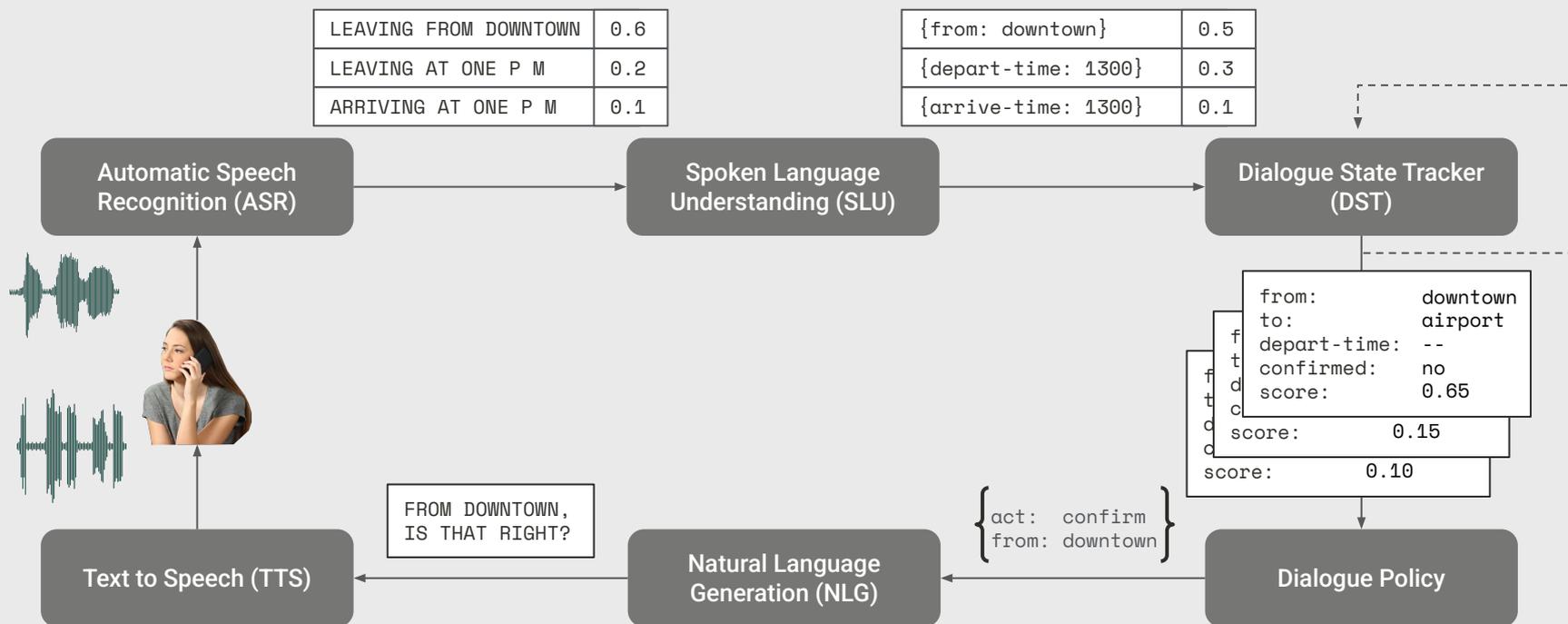


Figure: Architecture of dialogue-state system for task-oriented dialogue (William et al, 2016)

Slot Types Can Be Complex, Hierarchical

- The type DATE

DATE

MONTH NAME

DAY (BOUNDED-INTEGER 1 31)

YEAR INTEGER

WEEKDAY (MEMBER (SUNDAY MONDAY TUESDAY WEDNESDAY THURSDAY FRIDAY SATURDAY])

Frames and 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!
- Frame structure guides dialog

Mixed Initiative

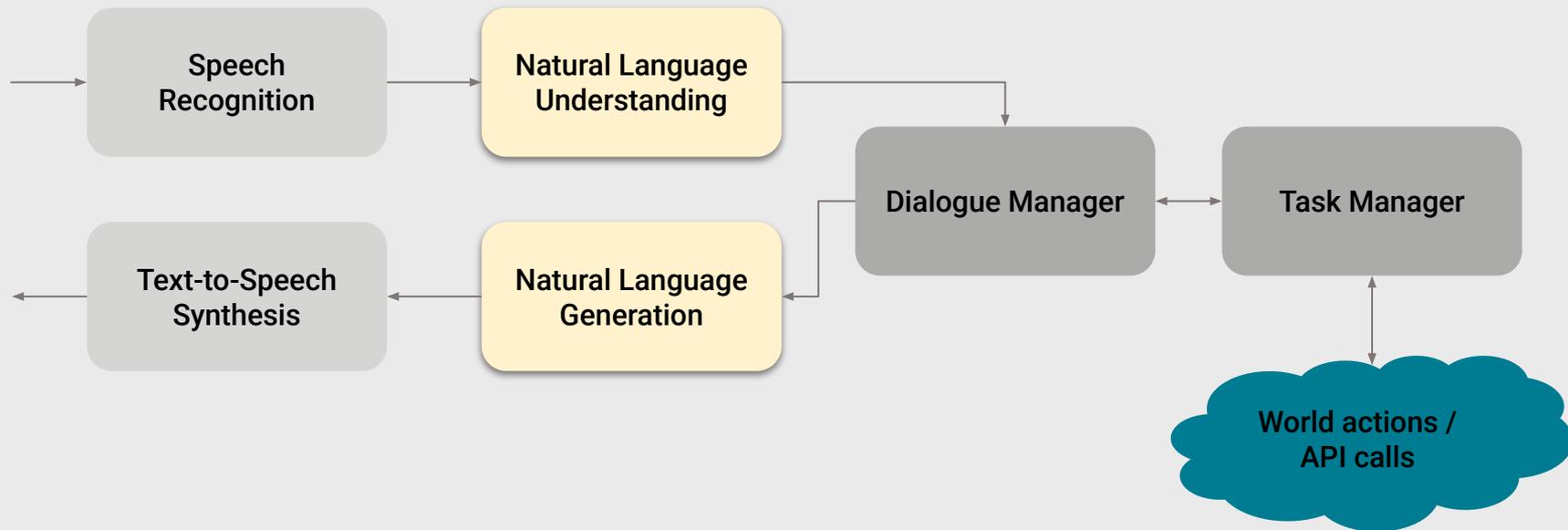
- Conversational initiative can shift between system and user
- Simplest kind of mixed initiative: use the structure of the frame to guide dialogue

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

NLU and NLG in frame-based systems

LLMs allow NLU + NLG modules to be more flexible

Broader scope, more robust, complex language/data mapping



Slot-Filling is Information/Entity Extraction from Text NLP

- Conditional Random Field (CRF) with word vector features, or neural classifiers both work well

Back in 2000 , **People Magazine** **PUBLISHER** highlighted **Prince Williams'** **PERSON** style who at the time was a little more fashion-conscious , even making fashion statements at times .

Now-a-days the prince mainly wears **navy** **COLOR** **suits** **ITEM** (sometimes **double-breasted** **DESIGN**) , **light blue** **COLOR** **button-ups** **ITEM** with **classic** **LOOK** **pointed** **DESIGN** **collars** **PART** , and **burgundy** **COLOR** **ties** **ITEM** .

But who knows what the future holds ...

Duchess Kate **PERSON** did wear an **Alexander McQueen** **BRAND** **dress** **ITEM** to the **wedding** **OCCASION** in the **fall of 2017** **SEASON** .

Natural Language Understanding for Filling Dialog Slots

- **Domain classification**
 - Asking weather? Booking a flight? Programming alarm clock?
- **Intent Determination**
 - Find a Movie, Show Flight, Remove Calendar Appt
- **Slot Filling**
 - Extract the actual slots and fillers

Natural Language Understanding for Filling Slots

- “Show me morning flights from Boston to SF on Tuesday”

DOMAIN: AIR-TRAVEL
INTENT: SHOW-FLIGHTS
ORIGIN-CITY: “Boston”
ORIGIN-DATE: “Tuesday”
ORIGIN-TIME: “Morning”
DEST-CITY: “San Francisco”

Natural Language Understanding for Filling Slots

- “Turn on my alarm for 6am on May 28”

DOMAIN: ALARM-CLOCK

INTENT: SET-ALARM

TIME: 2024-05-28 0600

Rule-based Slot-Filling

- Write regular expressions or grammar rules

```
Wake me (up) | set (the|an) alarm | get me up
```

- Do text normalization
- Time consuming and brittle NLU capabilities
- With modern NLP tools/features, only use rules alone in special cases
- Simple rules + LLM few-shot recognizers might be just as easy and more robust

Generation Component (NLG)

- **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 NLG

- Dialogue manager builds representation of meaning of utterance to be expressed
- Passes this to a “generator”. Old style was templates, modern systems use LLMs
- LLM-based NLG constrained to convey dialog representations can improve user satisfaction

- Critical aspect: Ensure correctness of what we convey to the user!

Deep Learning NLG Conditioned on Dialog Semantics

3	Input DA	Train{inform(time=50 minutes)}, hotel{request(stars=?; area=?), inform(choice=5)}
	Reference	travel time is 50 minutes . i found 5 hotels you may like . do you have a star rating request or an area you prefer
	SC-LSTM	there are 5 trains that are 50 minutes . do you have a preferred departure location ? % [train{inform(choice=5),request(departure=?)}] [hotel{request(stars=?; area=?), inform(choice=5)}]
	HDSA	there are 5 trains that meet your criteria . do you have a preference on the area or star rating ? % [train{inform(choice=5, time=50 minutes)}] [hotel{inform(choice=5)}]
	SC-GPT	there are 5 hotels that meet your criteria . the trip will last 50 minutes . do you have an area preference or star rating you would like ?

Figure: Examples of generated utterances from different models along with its corresponding dialogue acts (DAs) and references. Sampled from MultiWOZ. Each generated utterance is followed by a brief description explaining the errors (starting with %). Better viewed in color wrong, redundant, missing information. (Peng et al. 2020)

Deep Learning NLG Conditioned on Dialog Semantics

1	Input DA	Laptop{inform(name=satellite proteus 84; type=laptop; memory=8 gb; drive=1 tb; weight=2.3 kg)}
	Reference	the satellite proteus 84 is a laptop with a 1 tb drive , 8 gb memory and weighs in at 2.3 kg
	SC-LSTM	the satellite proteus 84 is a laptop with 8 gb , with a 1 tb drive , and is for business computing , and is for business computing % [businesscomputing=true] [weight=2.3kg]
	GPT-2	the satellite proteus 84 laptop has a 1 tb drive , weighs 2.3 kg and is used for business computing % [businesscomputing=true] [memory=8 gb]
	SC-GPT	the satellite proteus 84 is a laptop with 8 gb memory , 1 tb drive , and a weight of 2.3 kg
<hr/>		
2	Input DA	Restaurant{inform_only_match(name=marnee thai and thanh long; pricerange=moderate; near=outer sunset; goodformeal=dinner)}
	Reference	marnee thai and thanh long are the only 2 moderately priced restaurants near outer sunset that are good for dinner
	SC-LSTM	marnee thai and thanh long is is moderate restaurant restaurant and good good good dinner % [near=outer sunset] [inform_only_match]
	GPT-2	there is a moderately priced restaurant called marnee thai and thanh long that is near the outer sunset area % [goodformeal=dinner] [inform_only_match]
	SC-GPT	marnee thai and thanh long is the only restaurant that serves moderately priced food near outer sunset and good for dinner

Figure: Examples of generated utterances from different models along with its corresponding dialogue acts (DAs) and references. The examples are sampled from FewShotWOZ. Each generated utterance is followed by a brief description explaining the errors (starting with %). Better viewed in color wrong, redundant, missing information. (Peng et al, 2020)

Questions?