

Dialogue and Conversational Agents

Ling575
Spoken Dialog Systems
April 3, 2013

Roadmap

- Dialog and Dialog Systems
- Facets of Conversation:
 - Turn-taking
 - Speech Acts
 - Cooperativity
 - Grounding
- Spoken Dialogue Systems:
 - Pipeline Architecture
 - Finite-State, Frame-based, Information State Systems
 - Evaluation

Dialog Example

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.

Travel Planning

S1: Hello. You've reached the [Communicator]. Tell me your full name
U2: *hi i'd like to fly to seattle tuesday morning*
S3: Traveling to Seattle on Tuesday, August eleventh in the morning. Your full name?
U4: *[John Doe]*
S5: Hello, [John], I'm glad to hear from you again. You can now ask me for help.
S6: OK. Let's see I have a United flight departing Pittsburgh at ten forty five a.m. arriving into Seattle at one fifty seven p.m. Is that OK?
U7: *Yes*
S8: Will you return to Pittsburgh from Seattle?
U9: *no i will go to Los Angeles*
S10: Traveling to Los Angeles. on what date
U11: *Sorry can we go back to talk about the previous leg again*
S12: For the first leg, you are on a United flight 1446 departing Pittsburgh at ten forty five A.M. tomorrow, arriving in Seattle at one fifty seven P.M. Do you want to take that?
U13: *actually i want to leave on wednesday*
S14: a flight on wednesday.
S15: I have a flight departing Pittsburgh at ten forty five a.m., arrives Seattle at one fifty seven p.m. Is that OK?
U16: *Yes*

Figure 19.1 The travel domain: a fragment from a successful conversation between a user (U) and the Communicator system (S) of Xu and Rudnicky (2000).

AT&T's How May I Help You?

S: How may I help you?
J: *can you tell me how much it is to tokyo?*
S: You want to know the cost of a call?
J: *yes that's right*
S: Please hold on for rate information.

S: How may I help you?
J: *what's the area code for chicago?*
S: You want an area code.
J: *(silence)*
S: The area code for Chicago is 312.

Figure 19.2 The call routing/telephone information domain: two actual conversations between a user (U) and the AT&T How May I Help You (HMIHY) system (S) (Gorin et al., 1997).

ItSpoke Tutoring System

S: I didn't understand. What force acts on all objects within close proximity to earth?

U: *uh the force of gravity*

S: Fine. Besides the Earth's gravitational force, are there any other forces acting on the person, keys, and elevator after the release?

U: *no*

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

U: *it will be moving*

S: Fine. Not only are the person, keys, and elevator moving, they have only gravitational forces on them. When an object is falling and has only gravitational force on it, it is said to be in what?

U: *freefall*

Figure 19.3 The tutoring domain: part of an actual conversation between a student user (U) and the ITSPOKE system (S) of (Litman and Silliman, 2004), based on the Why2-Atlas text-based tutoring system (?).

Dialogue is Different

Dialogue is Different

- Two or more speakers
 - Primary focus on speech

Dialogue is Different

- Two or more speakers
 - Primary focus on speech
- Issues in multi-party spoken dialogue

Dialogue is Different

- Two or more speakers
 - Primary focus on speech
- Issues in multi-party spoken dialogue
 - Turn-taking – who speaks next, when?
 - Collaboration – clarification, feedback,...
 - Disfluencies
 - Adjacency pairs, dialogue acts

Conversations and Conversational Agents

- Conversation:
 - First and often most common form of language use
 - Context of language learning and use

Conversations and Conversational Agents

- Conversation:
 - First and often most common form of language use
 - Context of language learning and use
 - Goal:
 - Describe, characterize spoken interaction
 - Enable automatic recognition, understanding

Conversations and Conversational Agents

- Conversation:
 - First and often most common form of language use
 - Context of language learning and use
 - Goal:
 - Describe, characterize spoken interaction
 - Enable automatic recognition, understanding
- Conversational agents:
 - Spoken dialog systems, spoken language systems
 - Interact with users through speech
 - Tasks: travel arrangements, call routing, planning

Conversation

- Intricate, joint activity

Conversation

- Intricate, joint activity
 - Constructed from consecutive turns

Conversation

- Intricate, joint activity
 - Constructed from consecutive turns
 - Joint activity between speakers, hearer

Conversation

- Intricate, joint activity
 - Constructed from consecutive turns
 - Joint activity between speakers, hearer
 - Involves inferences about intended meaning

Conversation

- Intricate, joint activity
 - Constructed from consecutive turns
 - Joint activity between speakers, hearer
 - Involves inferences about intended meaning
- SDS: simpler, but hopefully consistent

Turn-Taking

- Multi-party discourse
 - Need to trade off speaker/hearer roles
 - Interpret reference from sequential utterances
- When?

Turn-Taking

- Multi-party discourse
 - Need to trade off speaker/hearer roles
 - Interpret reference from sequential utterances
- When?
 - End of sentence?

Turn-Taking

- Multi-party discourse
 - Need to trade off speaker/hearer roles
 - Interpret reference from sequential utterances
- When?
 - End of sentence?
 - No: multi-utterance turns
 - Silence?

Turn-Taking

- Multi-party discourse
 - Need to trade off speaker/hearer roles
 - Interpret reference from sequential utterances
- When?
 - End of sentence?
 - No: multi-utterance turns
 - Silence?
 - No: little silence in smooth dialogue: < 250ms
 - Gaps less than actual sentence planning time - anticipate
 - When other starts speaking?

Turn-Taking

- Multi-party discourse
 - Need to trade off speaker/hearer roles
 - Interpret reference from sequential utterances
- When?
 - End of sentence?
 - No: multi-utterance turns
 - Silence?
 - No: little silence in smooth dialogue: < 250ms
 - Gaps less than actual sentence planning time - anticipate
 - When other starts speaking?
 - No: relatively little overlap face-to-face: ~5%

Turn-taking: Who & How

- At each TRP in each turn (Sacks 1974)
 - If speaker has selected A to speak, A must take floor
 - If speaker has selected no one to speak, anyone can
 - If no one else takes the turn, the speaker can
- Selecting speaker A:

Turn-taking: Who & How

- At each TRP in each turn (Sacks 1974)
 - If speaker has selected A to speak, A must take floor
 - If speaker has selected no one to speak, anyone can
 - If no one else takes the turn, the speaker can
- Selecting speaker A:
 - By explicit/implicit mention: What about it, Bob?
 - By gaze, function
- Selecting others:

Turn-taking: Who & How

- At each TRP in each turn (Sacks 1974)
 - If speaker has selected A to speak, A must take floor
 - If speaker has selected no one to speak, anyone can
 - If no one else takes the turn, the speaker can
- Selecting speaker A:
 - By explicit/implicit mention: What about it, Bob?
 - By gaze, function
- Selecting others: questions, greetings, closing
 - (Traum et al., 2003)

Turns and Structure

- Some utterances select others:

Turns and Structure

- Some utterances select others:
 - Adjacency pairs:
 - Greeting – Greeting, Question – Answer,
 - Compliment – Downplayer

Turns and Structure

- Some utterances select others:
 - Adjacency pairs:
 - Greeting – Greeting, Question – Answer,
 - Compliment – Downplayer
 - Silence ‘dispreferred’ within adjacency pair
 - A: Is there something bothering you or not?
 - (1.0)
 - A: Yes or No?
 - (1.5)
 - A: Eh.
 - B: No.

Turns and Structure

- Some utterances select others:
 - Adjacency pairs:
 - Greeting – Greeting, Question – Answer,
 - Compliment – Downplayer
 - Silence ‘dispreferred’ within adjacency pair
 - A: Is there something bothering you or not?
 - (1.0)
 - A: Yes or No?
 - (1.5)
 - A: Eh.
 - B: No.

Turn-taking in HCI

- Human turn end:
 -

Turn-taking in HCI

- Human turn end:
 - Detected by 250ms (or longer) silence
- System turn end:

Turn-taking in HCI

- Human turn end:
 - Detected by 250ms (or longer) silence
- System turn end:
 - Signaled by end of speech
 - Indicated by any human sound
 - Barge-in
- Continued attention:

Turn-taking in HCI

- Human turn end:
 - Detected by 250ms (or longer) silence
- System turn end:
 - Signaled by end of speech
 - Indicated by any human sound
 - Barge-in
- Continued attention:
 - No signal
- Design problems create ambiguous silences
 - Problematic for SDS users
 - (Stifelman et al., 1993), (Yankelovich et al, 1995)

Speech Acts

- Utterance:
 - Action performed by the speaker (Austin, 1962)

Speech Acts

- Utterance:
 - Action performed by the speaker (Austin, 1962)
 - Performatives: *name, second*
 - *I name this ship the Titanic.*
 - *I second that motion.*
 - Extend to all utterances

Utterances as 3 Act Types

- Locutionary act:
 - utterance with some meaning
 - *“You can’t do that!”*

Utterances as 3 Act Types

- Locutionary act:
 - utterance with some meaning
 - *“You can’t do that!”*
- Illocutionary act:
 - Act of asking, promising, answering, in utterance

Utterances as 3 Act Types

- Locutionary act:
 - utterance with some meaning
 - *“You can’t do that!”*
- Illocutionary act:
 - Act of asking, promising, answering, in utterance
 - *Protesting*
- Perlocutionary act:
 - Production of effects on feeling, beliefs of addressee

Utterances as 3 Act Types

- Locutionary act:
 - utterance with some meaning
 - *“You can’t do that!”*
- Illocutionary act:
 - Act of asking, promising, answering, in utterance
 - *Protesting*
- Perlocutionary act:
 - Production of effects on feeling, beliefs of addressee
 - *Intend to prevent doing some action*
- Types: assertives, directives, commissives, expressives, declarations

The 3 levels of act revisited

	Locutionary Force	Illocutionary Force	Perlocutionary Force
Can I have the rest of your sandwich?			

The 3 levels of act revisited

	Locutionary Force	Illocutionary Force	Perlocutionary Force
Can I have the rest of your sandwich?	Question		

The 3 levels of act revisited

	Locutionary Force	Illocutionary Force	Perlocutionary Force
Can I have the rest of your sandwich?	Question	Request	

The 3 levels of act revisited

	Locutionary Force	Illocutionary Force	Perlocutionary Force
Can I have the rest of your sandwich?	Question	Request	Intent: You give me sandwich

The 3 levels of act revisited

	Locutionary Force	Illocutionary Force	Perlocutionary Force
Can I have the rest of your sandwich?	Question	Request	Intent: You give me sandwich
I want the rest of your sandwich			

The 3 levels of act revisited

	Locutionary Force	Illocutionary Force	Perlocutionary Force
Can I have the rest of your sandwich?	Question	Request	Intent: You give me sandwich
I want the rest of your sandwich	Declarative	Request	Intent: You give me sandwich
Give me your sandwich!			

The 3 levels of act revisited

	Locutionary Force	Illocutionary Force	Perlocutionary Force
Can I have the rest of your sandwich?	Question	Request	Intent: You give me sandwich
I want the rest of your sandwich	Declarative	Request	Intent: You give me sandwich
Give me your sandwich!	Imperative	Request	Intent: You give me sandwich

Collaborative Communication

- Speaker tries to establish and add to
 - “common ground” – “mutual belief”

Collaborative Communication

- Speaker tries to establish and add to
 - “common ground” – “mutual belief”
- Presumed a joint, collaborative activity
 - Make sure “mutually believe” the same thing

Collaborative Communication

- Speaker tries to establish and add to
 - “common ground” – “mutual belief”
- Presumed a joint, collaborative activity
 - Make sure “mutually believe” the same thing
- Hearer must ‘ground’ speaker’s utterances
 - Indicate heard and understood

Closure

- Principle of closure:
 - Agents performing an action require evidence of successful performance
 - Also important to indicate failure or understanding

Closure

- Principle of closure:
 - Agents performing an action require evidence of successful performance
 - Also important to indicate failure or understanding
- Non-speech closure:

Closure

- Principle of closure:
 - Agents performing an action require evidence of successful performance
 - Also important to indicate failure or understanding
- Non-speech closure:
 - Push elevator button -> Light turns on

Closure

- Principle of closure:
 - Agents performing an action require evidence of successful performance
 - Also important to indicate failure or understanding
- Non-speech closure:
 - Push elevator button -> Light turns on
- Two step process:
 - Presentation (speaker)
 - Acceptance (listener)

Degrees of Grounding

- Weakest to strongest

Degrees of Grounding

- Weakest to strongest
- Continued attention:
 - Silence implies consent

Degrees of Grounding

- Weakest to strongest
- Continued attention:
 - Silence implies consent
- Next relevant contribution

Degrees of Grounding

- Weakest to strongest
- Continued attention:
 - Silence implies consent
- Next relevant contribution
- Acknowledgment:
 - Minimal response, continuer: *yeah, uh-huh, okay; great*

Degrees of Grounding

- Weakest to strongest
- Continued attention:
 - Silence implies consent
- Next relevant contribution
- Acknowledgment:
 - Minimal response, continuer: *yeah, uh-huh, okay; great*
- Demonstrate:
 - Indicate understanding by reformulation, completion

Degrees of Grounding

- Weakest to strongest
- Continued attention:
 - Silence implies consent
- Next relevant contribution
- Acknowledgment:
 - Minimal response, continuer: *yeah, uh-huh, okay; great*
- Demonstrate:
 - Indicate understanding by reformulation, completion
- Display:
 - Repeat all or part

Dialog Example

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

- Display:
 - C: I need to travel in May.
 - A: And *what day* in May did you want to travel?

Grounding

- Display:
 - C: I need to travel in May.
 - A: *And what day* in May did you want to travel?
- Acknowledgment + Next relevant contribution:
 - *And what day* in May did you want to travel?
 - *And you are flying into what city?*
 - *And what time* would you like to leave Pittsburgh?

Travel Planning

S1: Hello. You've reached the [Communicator]. Tell me your full name
U2: *hi i'd like to fly to seattle tuesday morning*
S3: Traveling to Seattle on Tuesday, August eleventh in the morning. Your full name?
U4: *[John Doe]*
S5: Hello, [John], I'm glad to hear from you again. You can now ask me for help.
S6: OK. Let's see I have a United flight departing Pittsburgh at ten forty five a.m. arriving into Seattle at one fifty seven p.m. Is that OK?
U7: *Yes*
S8: Will you return to Pittsburgh from Seattle?
U9: *no i will go to Los Angeles*
S10: Traveling to Los Angeles. on what date
U11: *Sorry can we go back to talk about the previous leg again*
S12: For the first leg, you are on a United flight 1446 departing Pittsburgh at ten forty five A.M. tomorrow, arriving in Seattle at one fifty seven P.M. Do you want to take that?
U13: *actually i want to leave on wednesday*
S14: a flight on wednesday.
S15: I have a flight departing Pittsburgh at ten forty five a.m., arrives Seattle at one fifty seven p.m. Is that OK?
U16: *Yes*

Figure 19.1 The travel domain: a fragment from a successful conversation between a user (U) and the Communicator system (S) of Xu and Rudnicky (2000).

Grounding in HCI

- Key factor in HCI:
 - Users confused if system fails to ground, confirm
 - (Stifelman et al., 1993), (Yankelovich et al, 1995)
 - S: Did you want to review some more of your profile?
 - U: No.
 - S: What's next?

Grounding in HCI

- Key factor in HCI:
 - Users confused if system fails to ground, confirm
 - (Stifelman et al., 1993), (Yankelovich et al, 1995)
 - S: Did you want to review some more of your profile?
 - U: No.
 - S: What's next?
- S: Did you want to review some more of your profile?
- U: No.
- S: Okay, what's next?

Conversational Implicature

- Meaning more than just literal contribution
 - *A: And, what day in May did you want to travel?*
 - *C: OK uh I need to be there for a meeting the 12-15th*
 - Appropriate?

Conversational Implicature

- Meaning more than just literal contribution
 - *A: And, what day in May did you want to travel?*
 - *C: OK uh I need to be there for a meeting the 12-15th*
 - Appropriate? Yes
 - Why?

Conversational Implicature

- Meaning more than just literal contribution
 - *A: And, what day in May did you want to travel?*
 - *C: OK uh I need to be there for a meeting the 12-15th*
 - Appropriate? Yes
 - Why?
- Inference guides

Grice's Maxims

- Cooperative principle:
 - Tacit agreement b/t conversants to cooperate

Grice's Maxims

- Cooperative principle:
 - Tacit agreement b/t conversants to cooperate
- Grice's Maxims
 - Quantity: Be as informative as required

Grice's Maxims

- Cooperative principle:
 - Tacit agreement b/t conversants to cooperate
- Grice's Maxims
 - Quantity: Be as informative as required
 - Quality: Be truthful
 - Don't lie, or say things without evidence

Grice's Maxims

- Cooperative principle:
 - Tacit agreement b/t conversants to cooperate
- Grice's Maxims
 - Quantity: Be as informative as required
 - Quality: Be truthful
 - Don't lie, or say things without evidence
 - Relevance: Be relevant
 - Manner: "Be perspicuous"
 - Don't be obscure, ambiguous, prolix, or disorderly

Relevance

- Client: **I need to be there for a meeting that's from the 12th to the 15th**
 - Hearer thinks:

Relevance

- Client: **I need to be there for a meeting that's from the 12th to the 15th**
 - Hearer thinks: **Speaker is following maxims, would only have mentioned meeting if it was relevant. How could meeting be relevant? If client meant me to understand that he had to depart in time for the mtg.**

Quantity

- A: How much money do you have on you?
- B: I have 5 dollars
 - Implication

Quantity

- A: How much money do you have on you?
- B: I have 5 dollars
 - Implication: not 6 dollars
- A: Did you do the reading for today's class?
- B: I intended to
 - Implication:

Quantity

- A: How much money do you have on you?
- B: I have 5 dollars
 - Implication: not 6 dollars
- A: Did you do the reading for today's class?
- B: I intended to
 - Implication: No
 - B's answer would be true if B intended to do the reading AND did the reading, but would then violate maxim

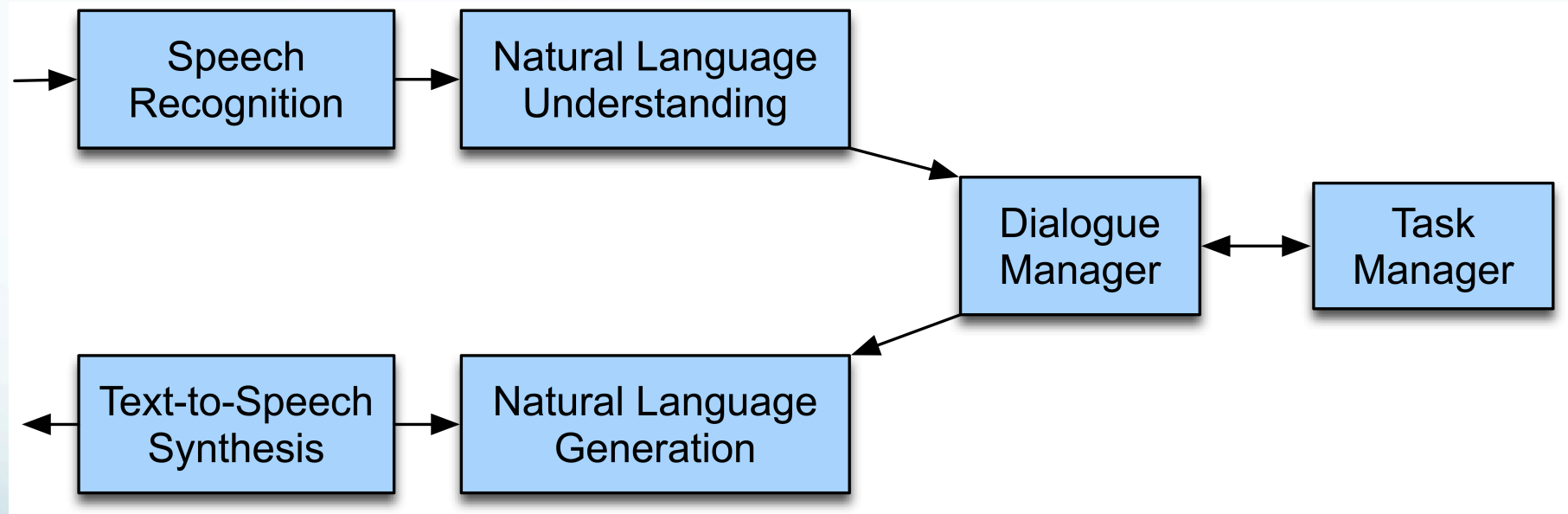
From Human to Computer

- Conversational agents
 - Systems that (try to) participate in dialogues
 - Examples: Directory assistance, travel info, weather, restaurant and navigation info
- Issues:

From Human to Computer

- Conversational agents
 - Systems that (try to) participate in dialogues
 - Examples: Directory assistance, travel info, weather, restaurant and navigation info
- Issues:
 - Limited understanding: ASR errors, interpretation
 - Computational costs

Dialogue System Architecture



Speech Recognition

- (aka ASR)
- Input: acoustic waveform
 - Telephone, microphone, and smartphone

Speech Recognition

- (aka ASR)
- Input: acoustic waveform
 - Telephone, microphone, and smartphone
- Output: recognized word string

Speech Recognition

- (aka ASR)
- Input: acoustic waveform
 - Telephone, microphone, and smartphone
- Output: recognized word string
- Requirements:

Speech Recognition

- (aka ASR)
- Input: acoustic waveform
 - Telephone, microphone, and smartphone
- Output: recognized word string
- Requirements:
 - Acoustic models: map acoustics to phone [ae] [k]
 - Pronunciation dictionary: words to phones: cat: [k][ae][t]
 - Grammar: legal word sequences
 - Search procedure: best word sequence given audio

Recognition in SDS

Recognition in SDS

- Create domain specific vocabulary, grammar
 - Typically hand-crafted in most commercial systems
 - Based on human-human interactions
 - Grammars: finite-state, context-free, language model

Recognition in SDS

- Create domain specific vocabulary, grammar
 - Typically hand-crafted in most commercial systems
 - Based on human-human interactions
 - Grammars: finite-state, context-free, language model
- Activate only portion of grammar based on dialog state
 - E.g. Where are you leaving from?

Recognition in SDS

- Create domain specific vocabulary, grammar
 - Typically hand-crafted in most commercial systems
 - Based on human-human interactions
 - Grammars: finite-state, context-free, language model
- Activate only portion of grammar based on dialog state
 - E.g. Where are you leaving from?
 - {I want to (leave|depart) from} CITYNAME {STATENAME}
 - ‘Yes/No’ grammar for confirmations

Natural Language Understanding

- Most systems use frame-slot semantics
Show me morning flights from Boston to SFO on Tuesday
Alternatives:
 - Full parser with semantic attachments
 - Domain-specific analyzers
- SHOW:
- FLIGHTS:
 - ORIGIN:
 - CITY: Boston
 - DATE:
 - DAY-OF-WEEK: Tuesday
 - TIME:
 - PART-OF-DAY: Morning
 - DEST:
 - CITY: San Francisco

Generation and TTS

- Generation:
 - Identify concepts to express
 - Convert to words
 - Assign appropriate prosody, intonation

Generation and TTS

- Generation:
 - Identify concepts to express
 - Convert to words
 - Assign appropriate prosody, intonation
- TTS:
 - Input words, prosodic markup
 - Synthesize acoustic waveform

Generation

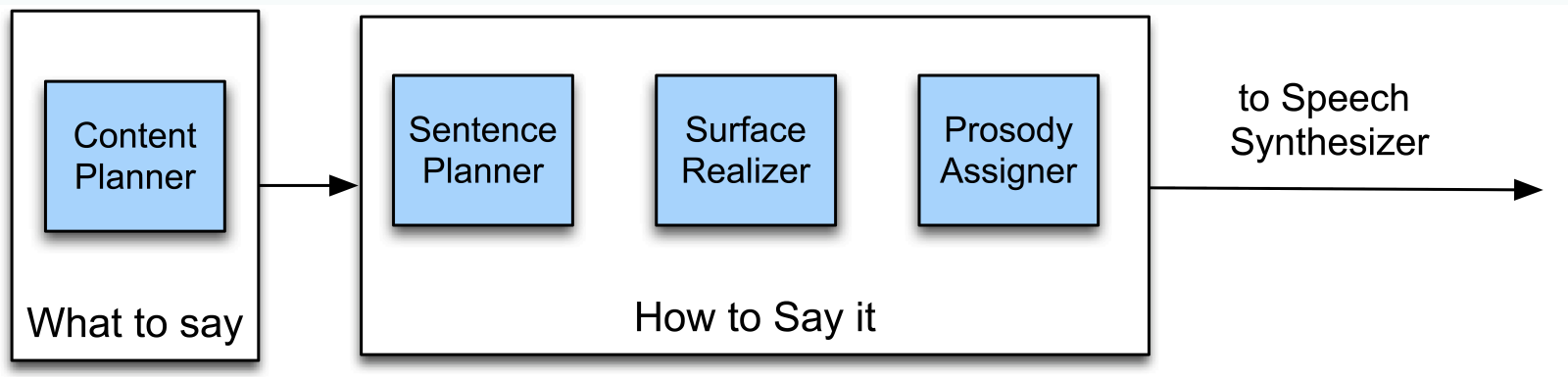
- Content planning:
 - What to say:
 - Question, answer, etc?
 - Often merged with dialog manager

Generation

- Content planning:
 - What to say:
 - Question, answer, etc?
 - Often merged with dialog manager
- Language generation:
 - How to say it
 - Select syntactic structure and words
 - Most common: Template-based generation (prompts)
 - Templates with variable: When do you want to leave CITY?

Full NLG

- Converts representation from dialog manager



Dialogue Manager

- Holds system together: Governs interaction style

Dialogue Manager

- Holds system together: Governs interaction style
 - Takes input from ASR/NLU

Dialogue Manager

- Holds system together: Governs interaction style
 - Takes input from ASR/NLU
- Maintains dialog state, history
 - Incremental frame construction
 - Reference, ellipsis resolution
 - Determines what system does next

Dialogue Manager

- Holds system together: Governs interaction style
 - Takes input from ASR/NLU
 - Maintains dialog state, history
 - Incremental frame construction
 - Reference, ellipsis resolution
 - Determines what system does next
 - Interfaces with task manager/backend app

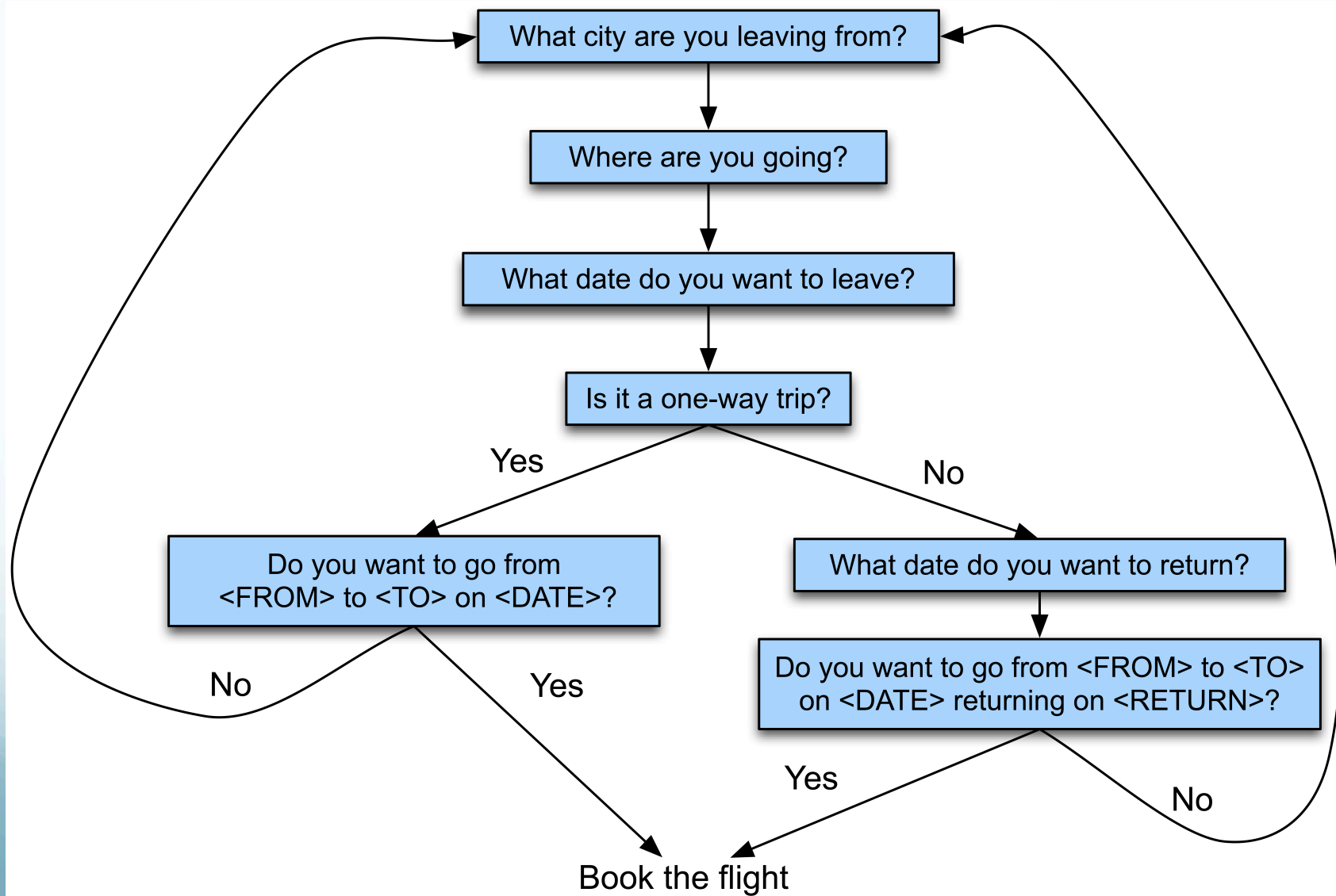
Dialogue Manager

- Holds system together: Governs interaction style
 - Takes input from ASR/NLU
 - Maintains dialog state, history
 - Incremental frame construction
 - Reference, ellipsis resolution
 - Determines what system does next
 - Interfaces with task manager/backend app
 - Formulates basic response, passes to NLG,TTS

Dialog Management Types

- Finite-State Dialog Management
- Frame-based Dialog Management
- Information State Manager
- Statistical Dialog Management

Finite-State Management



Finite-State Dialogue Management

- Simplest type of dialogue management
 - States:
 - Questions system asks user
 - Arcs:
 - User responses

Finite-State Dialogue Management

- Simplest type of dialogue management
 - States:
 - Questions system asks user
 - Arcs:
 - User responses
- System controls interactions:
 - Interprets all input based on current state
 - Assumes any user input is response to last question

Finite-State Dialogue Management

- Initiative:
 - Control of the interaction
- Who's in control here?

Finite-State Dialogue Management

- Initiative:
 - Control of the interaction
- Who's in control here?
 - System!
 - “system initiative”/”single initiative”
 - Natural?

Finite-State Dialogue Management

- Initiative:
 - Control of the interaction
- Who's in control here?
 - System!
 - “system initiative” / “single initiative”
 - Natural? No!
 - Human conversation goes back and forth
- Deploy targeted vocabulary / grammar for state
 - Add ‘universals’ – accessible anywhere in dialog
 - ‘Help’, ‘Start over’

Pros and Cons

- Advantages

Pros and Cons

- Advantages
 - Straightforward to encode
 - Clear mapping of interaction to model
 - Well-suited to simple information access
 - System initiative
- Disadvantages

Pros and Cons

- Advantages
 - Straightforward to encode
 - Clear mapping of interaction to model
 - Well-suited to simple information access
 - System initiative
- Disadvantages
 - Limited flexibility of interaction
 - Constrained input – single item
 - Fully system controlled
 - Restrictive dialogue structure, order
 - Ill-suited to complex problem-solving

Frame-based Dialogue Management

- Essentially form-filling
 - User can include any/all of the pieces of form
 - System must determine which entered, remain
 - Rules determine next action, question, information presentation

Frame-based Dialogue Management

- Essentially form-filling
 - User can include any/all of the pieces of form
 - System must determine which entered, remain
 - Rules determine next action, question, information presentation

Slot	Question
ORIGIN CITY	“From what city are you leaving?”
DESTINATION CITY	“Where are you going?”
DEPARTURE TIME	“When would you like to leave?”
ARRIVAL TIME	“When do you want to arrive?”

Frames and Initiative

- Mixed initiative systems:
 - A) User/System can shift control arbitrarily, any time
 - Difficult to achieve
 - B) Mix of control based on prompt type

Frames and Initiative

- Mixed initiative systems:
 - A) User/System can shift control arbitrarily, any time
 - Difficult to achieve
 - B) Mix of control based on prompt type
- Prompts:
 - Open prompt: 'How may I help you?'

Frames and Initiative

- Mixed initiative systems:
 - A) User/System can shift control arbitrarily, any time
 - Difficult to achieve
 - B) Mix of control based on prompt type
- Prompts:
 - Open prompt: ‘How may I help you?’
 - Open-ended, user can respond in any way
 - Directive prompt: ‘Say yes to accept call, or no o.w.’

Frames and Initiative

- Mixed initiative systems:
 - A) User/System can shift control arbitrarily, any time
 - Difficult to achieve
 - B) Mix of control based on prompt type
- Prompts:
 - Open prompt: ‘How may I help you?’
 - Open-ended, user can respond in any way
 - Directive prompt: ‘Say yes to accept call, or no o.w.’
 - Stipulates user response type, form

Dialogue Management: Confirmation

- Miscommunication common in SDS
 - “Error spirals” of sequential errors
 - Highly problematic
 - Recognition, recovery crucial
- Confirmation strategies can detect, mitigate
 - Explicit confirmation:

Dialog Example

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.

Travel Planning

S1: Hello. You've reached the [Communicator]. Tell me your full name
U2: *hi i'd like to fly to seattle tuesday morning*
S3: Traveling to Seattle on Tuesday, August eleventh in the morning. Your full name?
U4: *[John Doe]*
S5: Hello, [John], I'm glad to hear from you again. You can now ask me for help.
S6: OK. Let's see I have a United flight departing Pittsburgh at ten forty five a.m. arriving into Seattle at one fifty seven p.m. Is that OK?
U7: *Yes*
S8: Will you return to Pittsburgh from Seattle?
U9: *no i will go to Los Angeles*
S10: Traveling to Los Angeles. on what date
U11: *Sorry can we go back to talk about the previous leg again*
S12: For the first leg, you are on a United flight 1446 departing Pittsburgh at ten forty five A.M. tomorrow, arriving in Seattle at one fifty seven P.M. Do you want to take that?
U13: *actually i want to leave on wednesday*
S14: a flight on wednesday.
S15: I have a flight departing Pittsburgh at ten forty five a.m., arrives Seattle at one fifty seven p.m. Is that OK?
U16: *Yes*

Figure 19.1 The travel domain: a fragment from a successful conversation between a user (U) and the Communicator system (S) of Xu and Rudnicky (2000).

Dialogue Management: Confirmation

- Miscommunication common in SDS
 - “Error spirals” of sequential errors
 - Highly problematic
 - Recognition, recovery crucial
- Confirmation strategies can detect, mitigate
 - Explicit confirmation:
 - Ask for verification of each input
 - Implicit confirmation:

Dialogue Management: Confirmation

- Miscommunication common in SDS
 - “Error spirals” of sequential errors
 - Highly problematic
 - Recognition, recovery crucial
- Confirmation strategies can detect, mitigate
 - Explicit confirmation:
 - Ask for verification of each input
 - Implicit confirmation:
 - Include input information in subsequent prompt

Confirmation Strategies

- Explicit:

S: Which city do you want to leave from?

U: Baltimore.

S: Do you want to leave from Baltimore?

U: Yes.

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

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

U: Yes

Confirmation Strategy

- Implicit:

U: I want to travel to Berlin

S: When do you want to travel to Berlin?

U2: Hi I'd like to fly to Seattle Tuesday Morning

A3: Traveling to Seattle on Tuesday, August eleventh in the morning.
Your full name?

Pros and Cons

- Grounding of user input
 - Weakest grounding
 - I.e. continued att'n, next relevant contribution

Pros and Cons

- Grounding of user input
 - Weakest grounding insufficient
 - I.e. continued att'n, next relevant contribution
 - Explicit:

Pros and Cons

- Grounding of user input
 - Weakest grounding insufficient
 - I.e. continued att'n, next relevant contribution
 - Explicit: highest: repetition
 - Implicit:

Pros and Cons

- Grounding of user input
 - Weakest grounding insufficient
 - I.e. continued att'n, next relevant contribution
 - Explicit: highest: repetition
 - Implicit: demonstration, display
- Explicit;

Pros and Cons

- Grounding of user input
 - Weakest grounding insufficient
 - I.e. continued att'n, next relevant contribution
 - Explicit: highest: repetition
 - Implicit: demonstration, display
- Explicit;
 - Pro: easier to correct; Con: verbose, awkward, non-human
- Implicit:

Pros and Cons

- Grounding of user input
 - Weakest grounding insufficient
 - I.e. continued att'n, next relevant contribution
 - Explicit: highest: repetition
 - Implicit: demonstration, display
- Explicit;
 - Pro: easier to correct; Con: verbose, awkward, non-human
- Implicit:
 - Pro: more natural, efficient; Con: less easy to correct

VoiceXML

- W3C standard for simple frame-based dialogues
 - Fairly common in commercial settings
- Construct forms, menus
 - Forms get field data
 - Using attached prompts
 - With specified grammar (CFG)
 - With simple semantic attachments

Simple VoiceXML Example

```
<form>
  <field name="transporttype">
    <prompt>
      Please choose airline, hotel, or rental car.
    </prompt>
    <grammar type="application/x-nuance-gsl">
      [airline hotel "rental car"]
    </grammar>
  </field>
  <block>
    <prompt>
      You have chosen <value expr="transporttype">.
    </prompt>
  </block>
</form>
```

Frame-based Systems: Pros and Cons

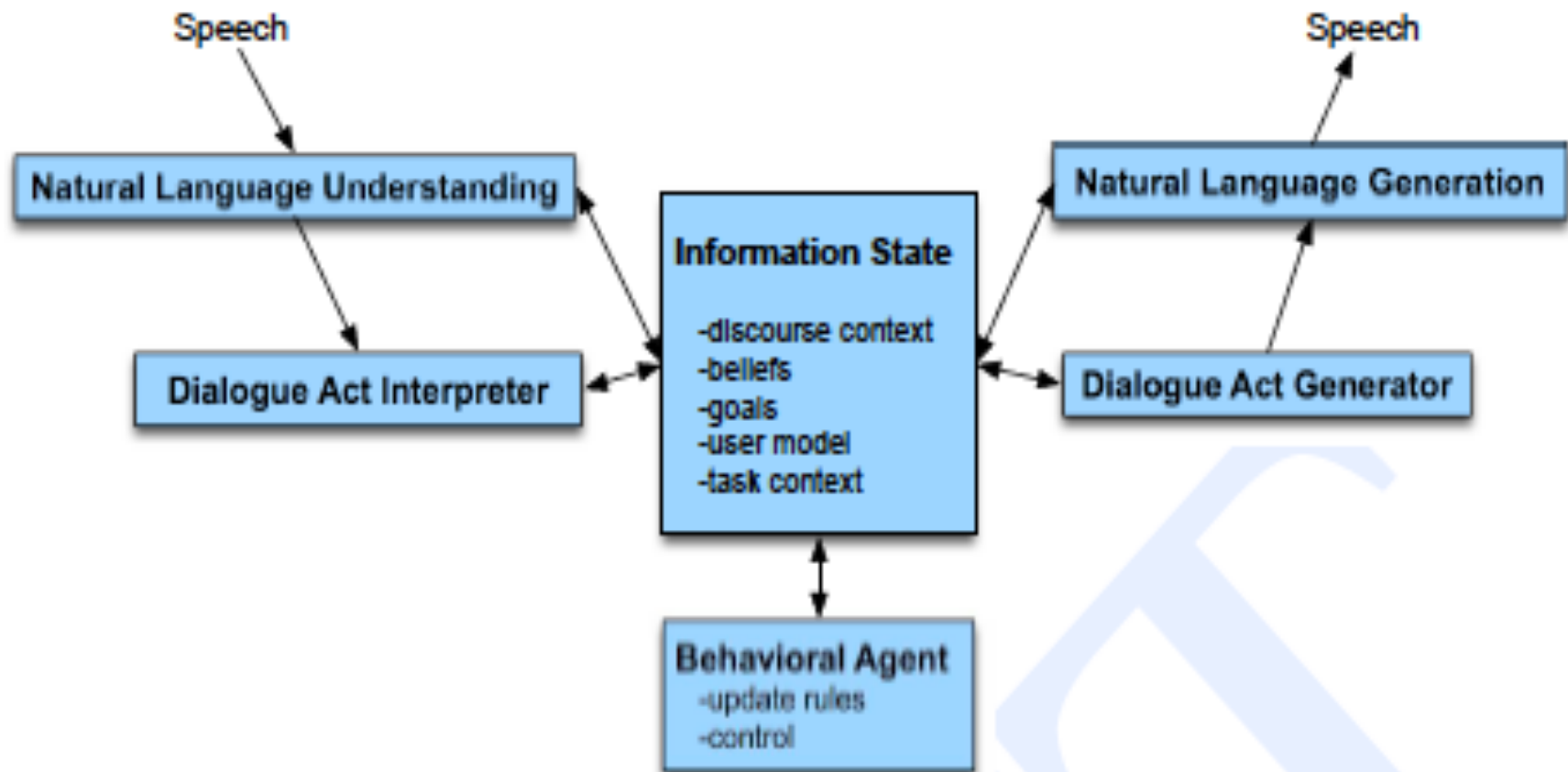
- Advantages
 - Relatively flexible input – multiple inputs, orders
 - Well-suited to complex information access (air)
 - Supports different types of initiative
- Disadvantages
 - Ill-suited to more complex problem-solving
 - Form-filling applications

Information State Dialogue Management

- Problem: Not every task is equivalent to form-filling
- Real tasks require:
 - Proposing ideas, refinement, rejection, grounding, clarification, elaboration, etc
- Information state models include:
 - Information state
 - Dialogue act interpreter
 - Dialogue act generator
 - Update rules
 - Control structure

Information State Architecture

- Simple ideas, complex execution



Dialogue Acts

- Extension of speech acts
 - Adds structure related to conversational phenomena
 - Grounding, adjacency pairs, etc
- Many proposed tagsets

Dialogue Act Interpretation

- Automatically tag utterances in dialogue
- Some simple cases:
 - **YES-NO-Q:** Will breakfast be served on USAir 1557?
 - I don't care about lunch.
 - Show be flights from L.A. to Orlando

Dialogue Act Interpretation

- Automatically tag utterances in dialogue
- Some simple cases:
 - **YES-NO-Q:** Will breakfast be served on USAir 1557?
 - **Statement:** I don't care about lunch.
 - Show be flights from L.A. to Orlando

Dialogue Act Interpretation

- Automatically tag utterances in dialogue
- Some simple cases:
 - **YES-NO-Q:** Will breakfast be served on USAir 1557?
 - **Statement:** I don't care about lunch.
 - **Command:** Show be flights from L.A. to Orlando
- Is it always that easy?
 - Can you give me the flights from Atlanta to Boston?
 - Yeah.

Dialogue Act Interpretation

- Automatically tag utterances in dialogue
- Some simple cases:
 - **YES-NO-Q:** Will breakfast be served on USAir 1557?
 - **Statement:** I don't care about lunch.
 - **Command:** Show be flights from L.A. to Orlando
- Is it always that easy?
 - Can you give me the flights from Atlanta to Boston?
 - Yeah.
 - Depends on context: Y/N answer; agreement; back-channel

Detecting Correction Acts

- Miscommunication is common in SDS
 - Utterances after errors misrecognized >2x as often
 - Frequently repetition or paraphrase of original input

Detecting Correction Acts

- Miscommunication is common in SDS
 - Utterances after errors misrecognized >2x as often
 - Frequently repetition or paraphrase of original input
- Systems need to detect, correct

Detecting Correction Acts

- Miscommunication is common in SDS
 - Utterances after errors misrecognized >2x as often
 - Frequently repetition or paraphrase of original input
- Systems need to detect, correct
- Corrections are spoken differently:
 - Hyperarticulated (slower, clearer) -> lower ASR conf.
 - Some word cues: 'No', 'I meant', swearing..

Detecting Correction Acts

- Miscommunication is common in SDS
 - Utterances after errors misrecognized >2x as often
 - Frequently repetition or paraphrase of original input
- Systems need to detect, correct
- Corrections are spoken differently:
 - Hyperarticulated (slower, clearer) -> lower ASR conf.
 - Some word cues: 'No', 'I meant', swearing..
- Can train classifiers to recognize with good acc.

Designing Dialog

- Apply user-centered design

Designing Dialog

- Apply user-centered design
 - Study user and task: How?

Designing Dialog

- Apply user-centered design
 - Study user and task: How?
 - Interview potential users, record human-human tasks
 - Study how the user interacts with the system

Designing Dialog

- Apply user-centered design
 - Study user and task: How?
 - Interview potential users, recorded human-human tasks
 - Study how the user interacts with the system
 - But it's not built yet....

Designing Dialog

- Apply user-centered design
 - Study user and task: How?
 - Interview potential users, recorded human-human tasks
 - Study how the user interacts with the system
 - But it's not built yet....
 - Wizard-of-Oz systems: Simulations
 - User thinks they're interacting with a system, but it's driven by a human
 - Prototypes

Designing Dialog

- Apply user-centered design
 - Study user and task: How?
 - Interview potential users, recorded human-human tasks
 - Study how the user interacts with the system
 - But it's not built yet....
 - Wizard-of-Oz systems: Simulations
 - User thinks they're interacting with a system, but it's driven by a human
 - Prototypes
 - Iterative redesign:
 - Test system: see how users really react, what problems occur, correct, repeat

SDS Evaluation

- Goal: Determine overall user satisfaction
 - Highlight systems problems; help tune

SDS Evaluation

- Goal: Determine overall user satisfaction
 - Highlight systems problems; help tune
- Classically: Conduct user surveys

SDS Evaluation

- Goal: Determine overall user satisfaction
 - Highlight systems problems; help tune
- Classically: Conduct user surveys

TTS Performance	Was the system easy to understand ?
ASR Performance	Did the system understand what you said?
Task Ease	Was it easy to find the message/flight/train you wanted?
Interaction Pace	Was the pace of interaction with the system appropriate?
User Expertise	Did you know what you could say at each point?
System Response	How often was the system sluggish and slow to reply to you?
Expected Behavior	Did the system work the way you expected it to?
Future Use	Do you think you'd use the system in the future?

Figure 24.14 User satisfaction survey, adapted from Walker et al. (2001).

SDS Evaluation

- User evaluation issues:

SDS Evaluation

- User evaluation issues:
 - Expensive; often unrealistic; hard to get real user to do
- Create model correlated with human satisfaction
- Criteria:

SDS Evaluation

- User evaluation issues:
 - Expensive; often unrealistic; hard to get real user to do
- Create model correlated with human satisfaction
- Criteria:
 - Maximize task success
 - Measure task completion: % subgoals; Kappa of frame values
 - Minimize task costs
 - Efficiency costs: time elapsed; # turns; # error correction turns
 - Quality costs: # rejections; # barge-in; concept error rate

PARADISE Model

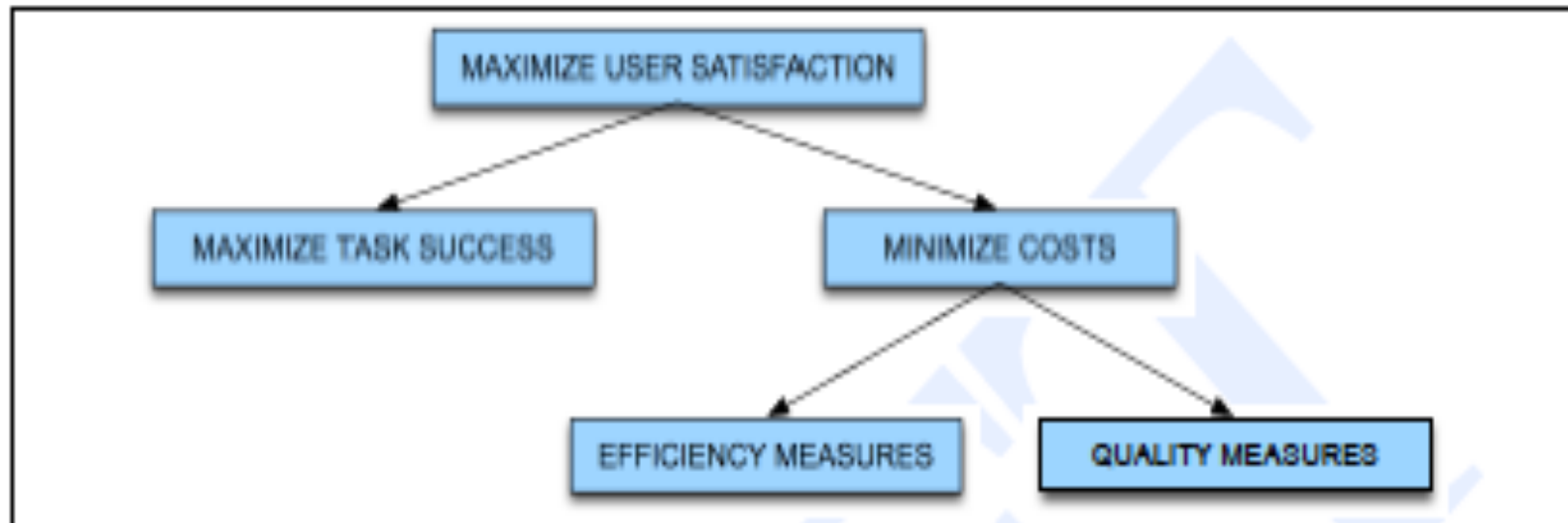


Figure 24.15 PARADISE's structure of objectives for spoken dialogue performance. After Walker et al. (1997).

PARADISE Model

- Compute user satisfaction with questionnaires
- Extract task success and costs measures from corresponding dialogs
 - Automatically or manually
- Perform multiple regression:
 - Assign weights to all factors of contribution to Usat
 - Task success, Concept accuracy key
- Allows prediction of accuracy on new dialog

Summary

- Spoken Dialogue Systems:
 - Build on existing text-based NLP techniques, but
 - Incorporate dialogue specific factors:
 - Turn-taking, grounding, dialogue acts
 - Affected by computational and modal constraints
 - Recognition errors, processing speed, etc.
 - Speech transience, slowness
 - Becoming more widespread and more flexible

Semantic Grammars

- Alternatives:
 - Full parser with semantic attachments
 - Domain-specific analyzers
- 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

Result

- **SHOW FLIGHT ORIGIN DEST DEP_DATE DEP_TIME**
- *Show me flights from Boston to SFO on Tuesday morning*

Verbmobil DA

- 18 high level tags

Tag	Example
THANK	<i>Thanks</i>
GREET	<i>Hello Dan</i>
INTRODUCE	<i>It's me again</i>
BYE	<i>Allright bye</i>
REQUEST-COMMENT	<i>How does that look?</i>
SUGGEST	<i>from thirteenth through seventeenth June</i>
REJECT	<i>No Friday I'm booked all day</i>
ACCEPT	<i>Saturday sounds fine,</i>
REQUEST-SUGGEST	<i>What is a good day of the week for you?</i>
INIT	<i>I wanted to make an appointment with you</i>
GIVE_REASON	<i>Because I have meetings all afternoon</i>
FEEDBACK	<i>Okay</i>
DELIBERATE	<i>Let me check my calendar here</i>
CONFIRM	<i>Okay, that would be wonderful</i>
CLARIFY	<i>Okay, do you mean Tuesday the 23rd?</i>
DIGRESS	<i>[we could meet for lunch] and eat lots of ice cream</i>
MOTIVATE	<i>We should go to visit our subsidiary in Munich</i>
GARBAGE	<i>Oops, I-</i>

Figure 24.17 The 18 high-level dialogue acts used in Verbmobil-1, abstracted over a total of 43 more specific dialogue acts. Examples are from Jekat et al. (1995).

Dialogue Act Ambiguity

- Indirect speech acts

A	OPEN-OPTION	I was wanting to make some arrangements for a trip that I'm going to be taking uh to LA uh beginning of the week after next.
B	HOLD	OK uh let me pull up your profile and I'll be right with you here. [pause]
B	CHECK	And you said you wanted to travel next week?
A	ACCEPT	Uh yes.

Performance Functions for 3 Systems

- ELVIS User Sat.= $.21 * \text{COMP} + .47 * \text{MRS} - .15 * \text{ET}$
 - TOOT User Sat.= $.35 * \text{COMP} + .45 * \text{MRS} - .14 * \text{ET}$
 - ANNIE User Sat.= $.33 * \text{COMP} + .25 * \text{MRS} + .33 * \text{Help}$
-
- COMP: User perception of task completion (task success)
 - MRS: Mean (concept) recognition accuracy (cost)
 - ET: Elapsed time (cost)
 - Help: Help requests (cost)