Challenges in Dialog & Dialog Systems

Ling575
Discourse and Dialog
April 27, 2011

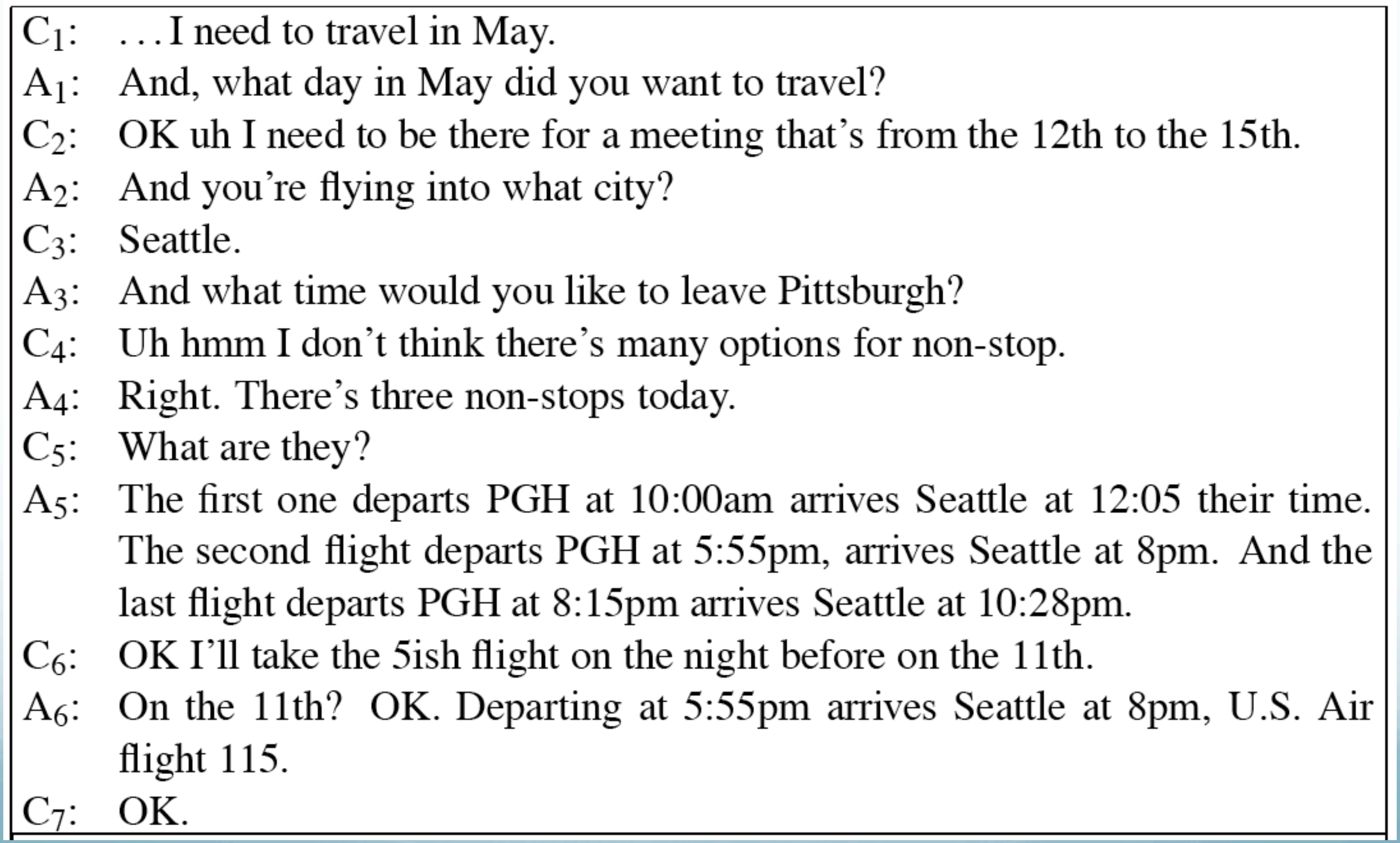
Roadmap

- Issues in Dialog & Dialog Systems
  - Dialog vs General Discourse
  - Linguistics of Conversation
    - Turn-taking
    - Grounding
    - Implicature
  - Dialog Systems
    - Architecture
    - Components
    - Evaluation

Dialogue vs General Discourse

- Key contrast: Two or more speakers
  - Primary focus on speech

Dialog Example



C_1: …I need to travel in May.

A_1: And, what day in May did you want to travel?

C_2: OK uh I need to be there for a meeting that's from the 12th to the 15th.

A_2: And you're flying into what city?

C_3: Seattle.

A_3: And what time would you like to leave Pittsburgh?

C_4: Uh hmm I don't think there's many options for non-stop.

A_4: Right. There's three non-stops today.

C_5: What are they?

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.

C_6: OK I'll take the 5ish flight on the night before on the 11th.

A_6: On the 11th? OK. Departing at 5:55pm arrives Seattle at 8pm, U.S. Air flight 115.

C_7: OK.

Dialogue vs General Discourse

- Key contrast: 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

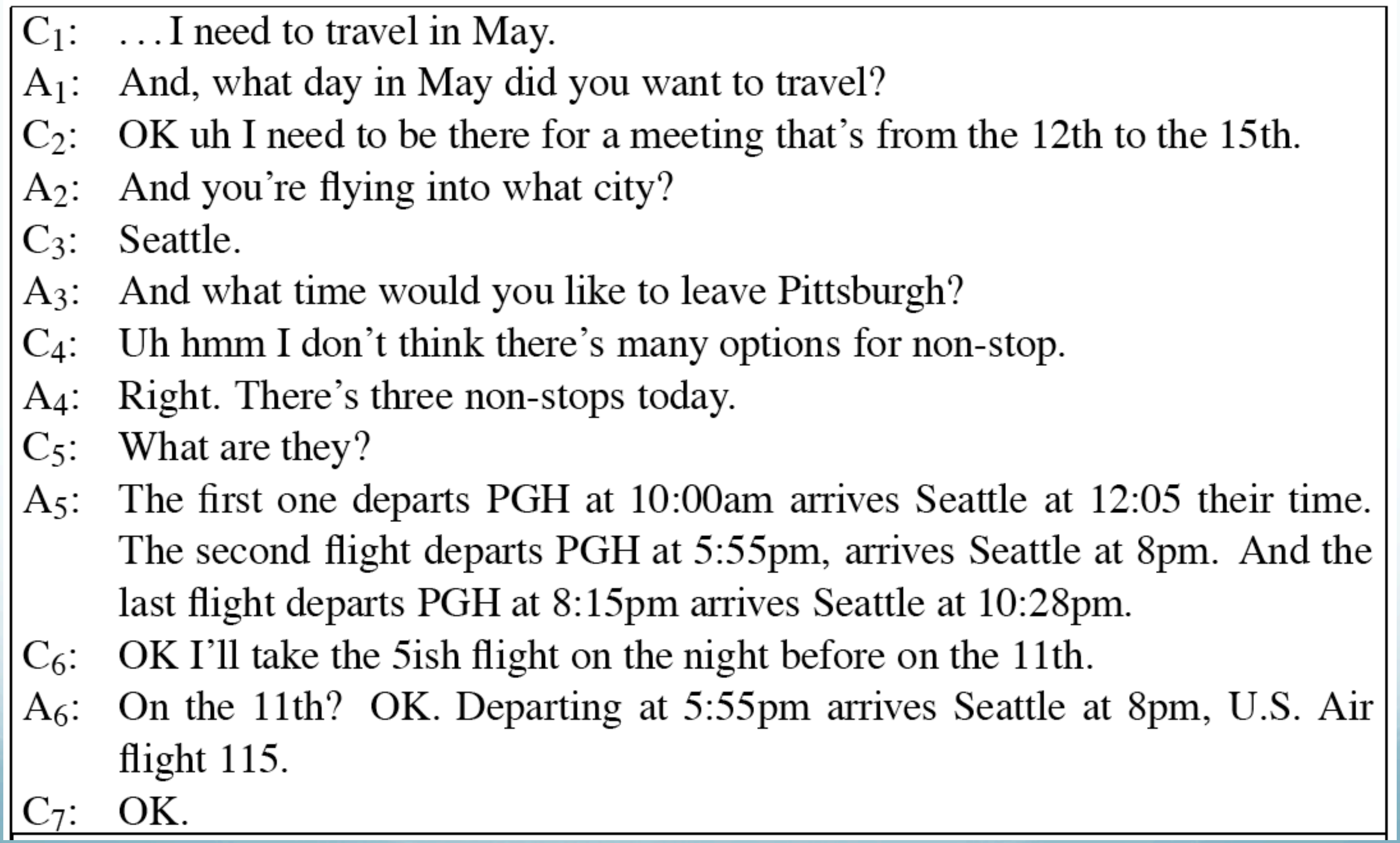
- Conversational agents:
  - Spoken dialog systems, spoken language systems

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

Human-Human Example



C_1: . . . I need to travel in May.

A_1: And, what day in May did you want to travel?

C_2: OK uh I need to be there for a meeting that's from the 12th to the 15th.

A_2: And you're flying into what city?

C_3: Seattle.

A_3: And what time would you like to leave Pittsburgh?

C_4: Uh hmm I don't think there's many options for non-stop.

A_4: Right. There's three non-stops today.

C_5: What are they?

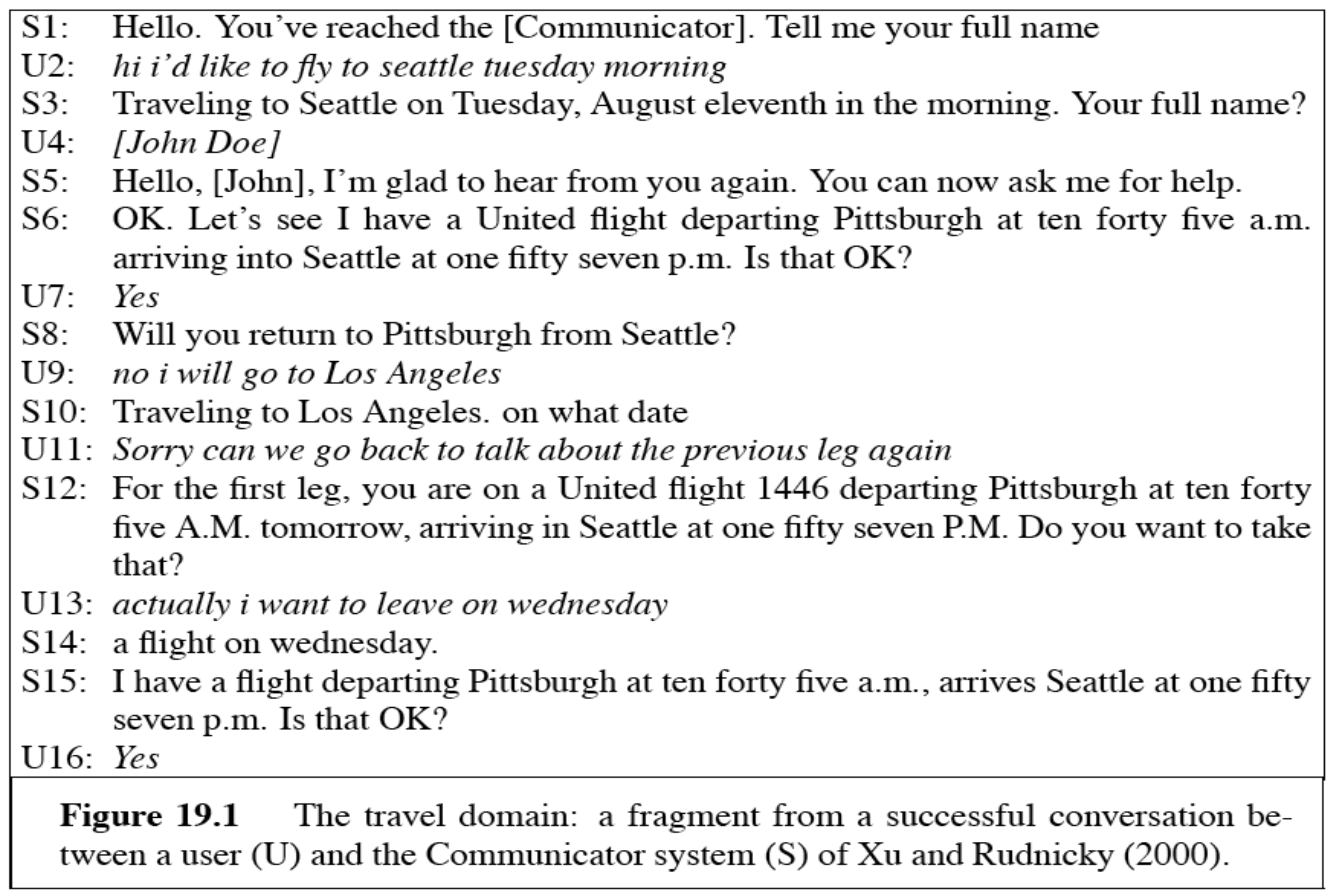
A_5: The first one departs PGH at 10:00am arrives Seattle at 12:05 their time. The second flight departs PGH at 5:55pm, arrives Seattle at 8pm. And the last flight departs PGH at 8:15pm arrives Seattle at 10:28pm.

C_6: OK I'll take the 5ish flight on the night before on the 11th.

A_6: On the 11th? OK. Departing at 5:55pm arrives Seattle at 8pm, U.S. Air flight 115.

C_7: OK.

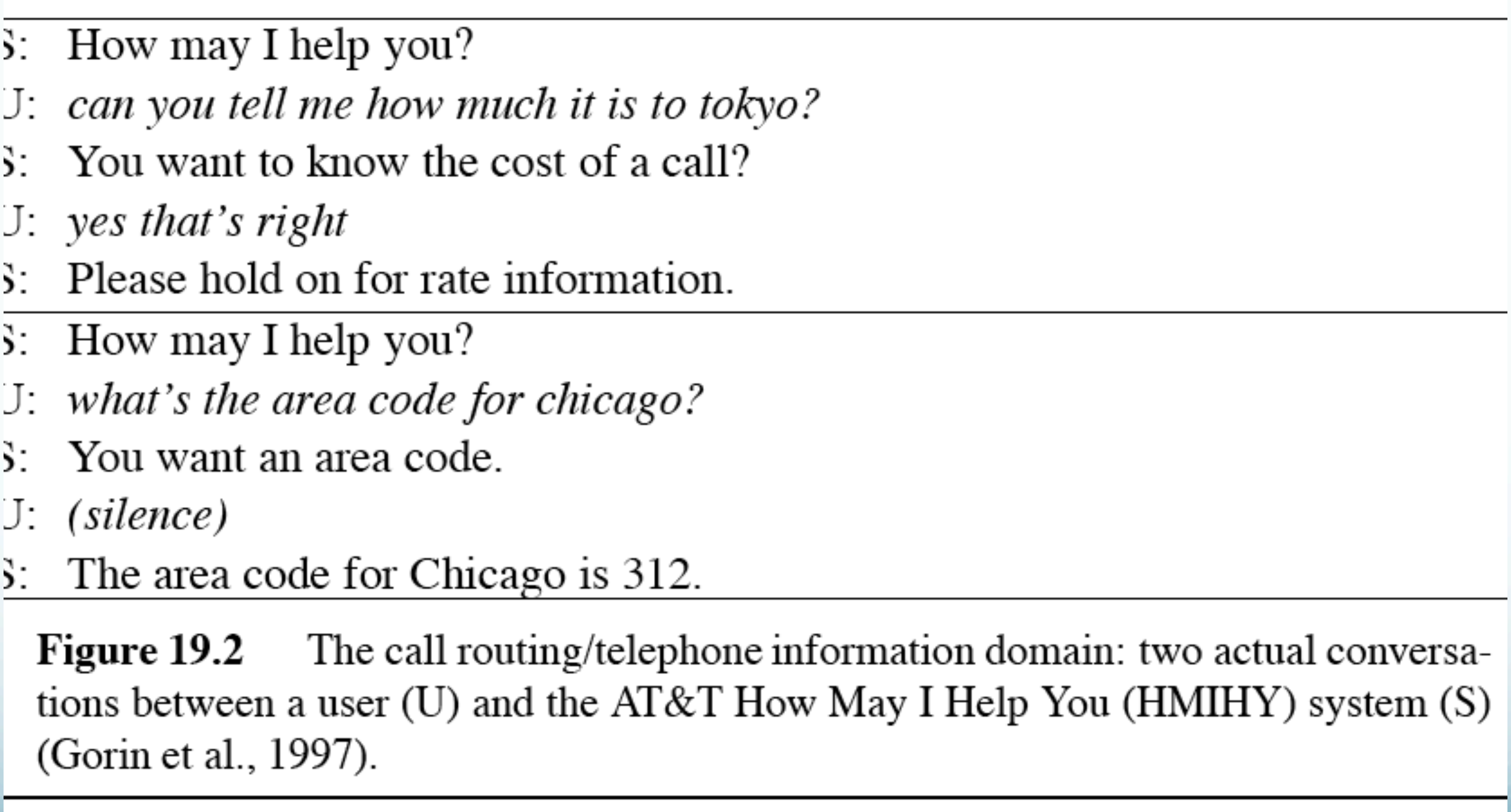
Travel Planning



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?

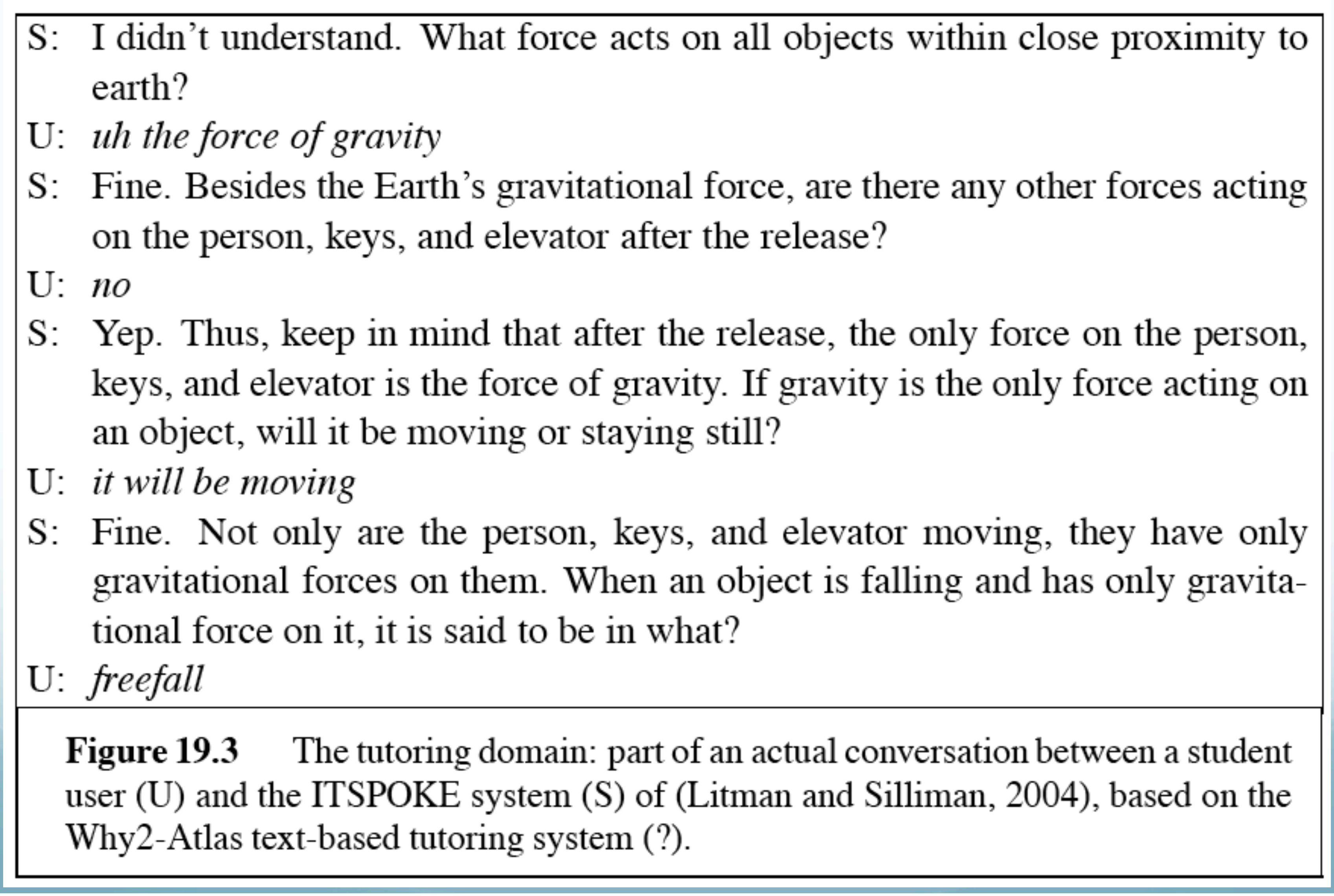


S: How may I help you?
U: *can you tell me how much it is to tokyo?*
S: You want to know the cost of a call?
U: *yes that's right*
S: Please hold on for rate information.

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

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 (?).

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

- Rule-governed behavior
  - Possibly multiple legal turn change times
    - Aka transition-relevance places (TRP)

Turn-taking

- Rule-governed behavior
  - Possibly multiple legal turn change times
    - Aka transition-relevance places (TRP)
    - Generally at utterance boundaries
      - Utterance not necessarily sentence
      - In fact, utterance/sentence boundaries not obvious in speech
        - Don't necessarily pause between sentences

Turn-taking

- Rule-governed behavior
  - Possibly multiple legal turn change times
    - Aka transition-relevance places (TRP)
    - Generally at utterance boundaries
      - Utterance not necessarily sentence
      - In fact, utterance/sentence boundaries not obvious in speech
        - Don't necessarily pause between sentences
    - Automatic utterance boundary detection
      - Cue words (okay, so,..); POS sequences; prosody

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

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

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
  - Linkage 'disprefers' silences within adjacency pair
    - More acceptable between

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

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)
    - 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
  - *Protesting*

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*

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?	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

Dialogue Acts

- (aka Conversational moves)
  - Enriched set of speech acts
    - Capture full range of conversational functions

Dialogue Acts

- (aka Conversational moves)
  - Enriched set of speech acts
    - Capture full range of conversational functions
  - Adjacency pairs: Many two-part structures
    - E.g. Question-Answer, Greeting-Greeting, Request-Grant, etc…

Dialogue Acts

- (aka Conversational moves)
  - Enriched set of speech acts
    - Capture full range of conversational functions
  - Adjacency pairs: Many two-part structures
    - E.g. Question-Answer, Greeting-Greeting, Request-Grant, etc…
    - Paired for speaker-hearer dyads
      - Contrast with rhetorical relations in monologue

DAMSL

- Dialogue Act Tagging framework
  - Builds on Adjacency pairs

DAMSL

- Dialogue Act Tagging framework
  - Builds on Adjacency pairs
- Forward looking functions
  - Statement, info-request, commit, closing, etc

DAMSL

- Dialogue Act Tagging framework
  - Builds on Adjacency pairs
- Forward looking functions
  - Statement, info-request, commit, closing, etc
- Backward looking functions
  - Focus on link to prior speaker utterance
    - Agreement, answer, accept, etc..

[assert] C1: . . . I need to travel in May.

[inforeq,ack] A1: And, what day in May did you want to travel?

[assert,answer] C2: OK uh I need to be there for a meeting that's from the 12th to the 15th.

[inforeq,ack] A2: And you're flying into what city?

[assert,answer]C3: Seattle.

[inforeq,ack] A3: And what time would you like to leave Pittsburgh?

[check,hold] C4: Uh hmm I dont think theres many options for nonstop.

[accept,ack] A4: Right.

[assert] There's three non-stops today.

[info-req] C5: What are they?

[assert,open-option] A5: The first one departs PGH at...

Dialogue Act Recognition

- Goal: Identify dialogue act tag(s) from surface form

Dialogue Act Recognition

- Goal: Identify dialogue act tag(s) from surface form
- Challenge: Surface form can be ambiguous
  - “Can you X?” – yes/no question, or info-request
    - “Flying on the 11th, at what time?” – check, statement

Dialogue Act Recognition

- Goal: Identify dialogue act tag(s) from surface form

- Challenge: Surface form can be ambiguous
    - “Can you X?” – yes/no question, or info-request
        - “Flying on the 11th, at what time?” – check, statement

- Requires interpretation by hearer
    - Strategies: cue recognition

Cue-based Interpretation

- Employs sets of features to identify
  - Words and collocations: Please -> request
  - Prosody: Rising pitch -> yes/no question
  - Conversational structure: prior act

Cue-based Interpretation

- Employs sets of features to identify
  - Words and collocations: Please -> request
  - Prosody: Rising pitch -> yes/no question
  - Conversational structure: prior act
- Example: Check:
  - Syntax: tag question ",right?"
  - Syntax + prosody: Fragment with rise
  - N-gram: argmax d P(d)P(W|d)
    - So you, sounds like, etc
- Details later ....