CS 2731
Introduction to Natural Language Processing

Session 23: Dialogue systems

Michael Miller Yoder

November 18, 2024

University of Pittsburgh | School of Computing and Information

Course logistics: project

- Project progress report feedback and grades are on Canvas
- I will go through project peer reviews soon
- Final project presentations are on **Wed Dec 11**
- Project report is **due Thu Dec 12**

Learning objectives for this session

Students will be able to:

- Differentiate between dialogue systems and chatbots
- List properties of human conversation
- Explain how the notion of “frames” and “slot-filling” plays a part in task-based dialogue systems
- Identify operations in dialogue-state architectures
    - Including natural language understanding, dialogue state tracking, dialogue policies, natural language generation
- Give examples of dialogue acts
- Evaluate specific aspects of task-based dialogue systems
- Identify common ethical considerations with dialogue systems

Dialogue systems and chatbots

Conversational Systems

- Personal Assistants on phones or other devices
  - SIRI, Alexa, Cortana, Google Assistant
- Playing music, setting timers and clocks
- Chatting for fun
- Booking travel reservations
- Clinical uses for mental health

Slide credit: David Mortensen, Jurafsky & Martin

Two kinds of conversational systems

- Chatbots
    - mimic informal human chatting for fun, or even for therapy
- (Task-based) Dialogue Agents
    - interfaces to personal assistants
    - cars, robots, appliances
    - booking flights or restaurants

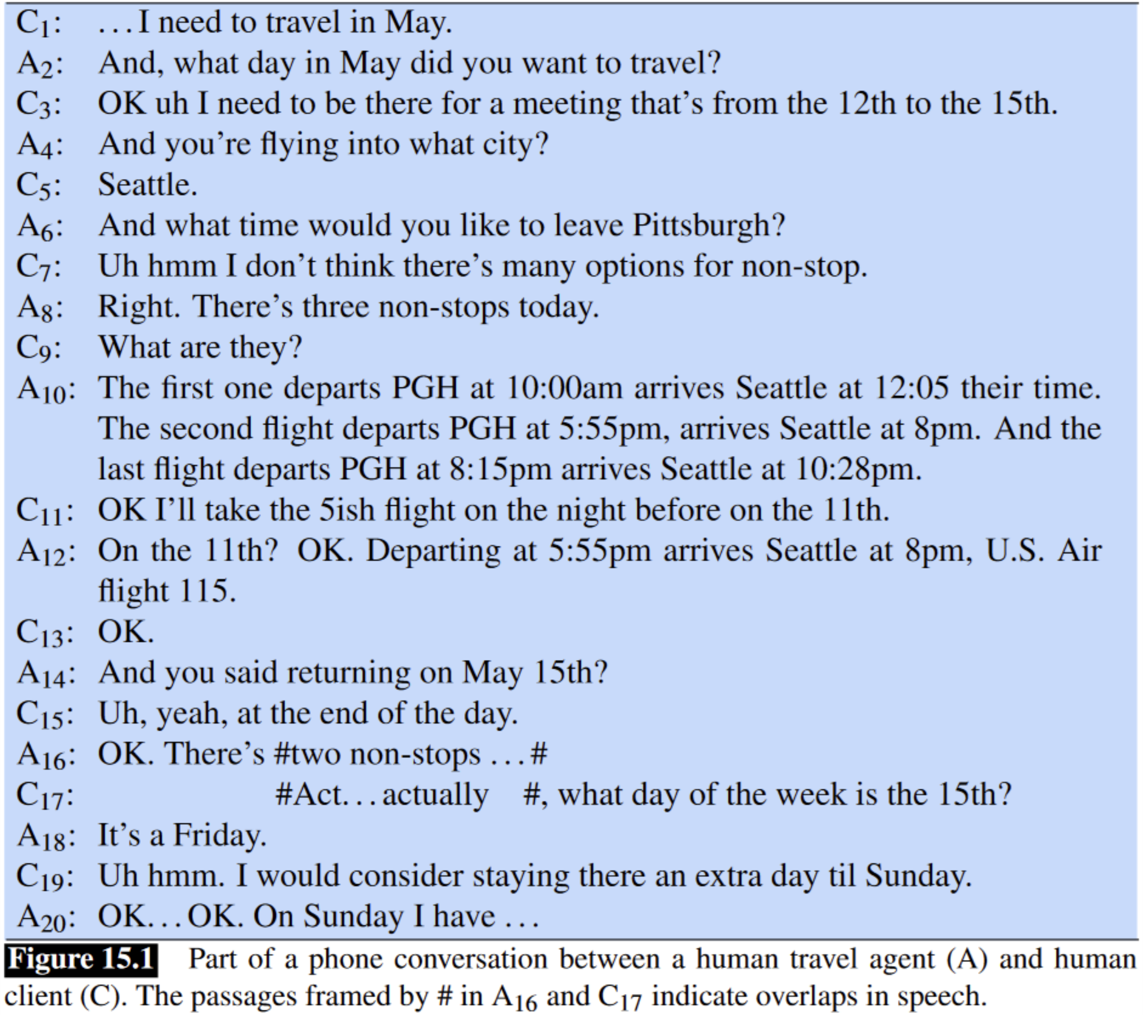
Slide credit: David Mortensen, Jurafsky & Martin

Spoken conversational systems

- Incorporates speech recognition and text-to-speech
    - Additional possible sources of error
- Benefits of speech as an interface
    - Highly intuitive
    - Eyes and hands-free
    - Small devices
    - Rich communication channel

Slide adapted from Diane Litman

Properties of human conversation

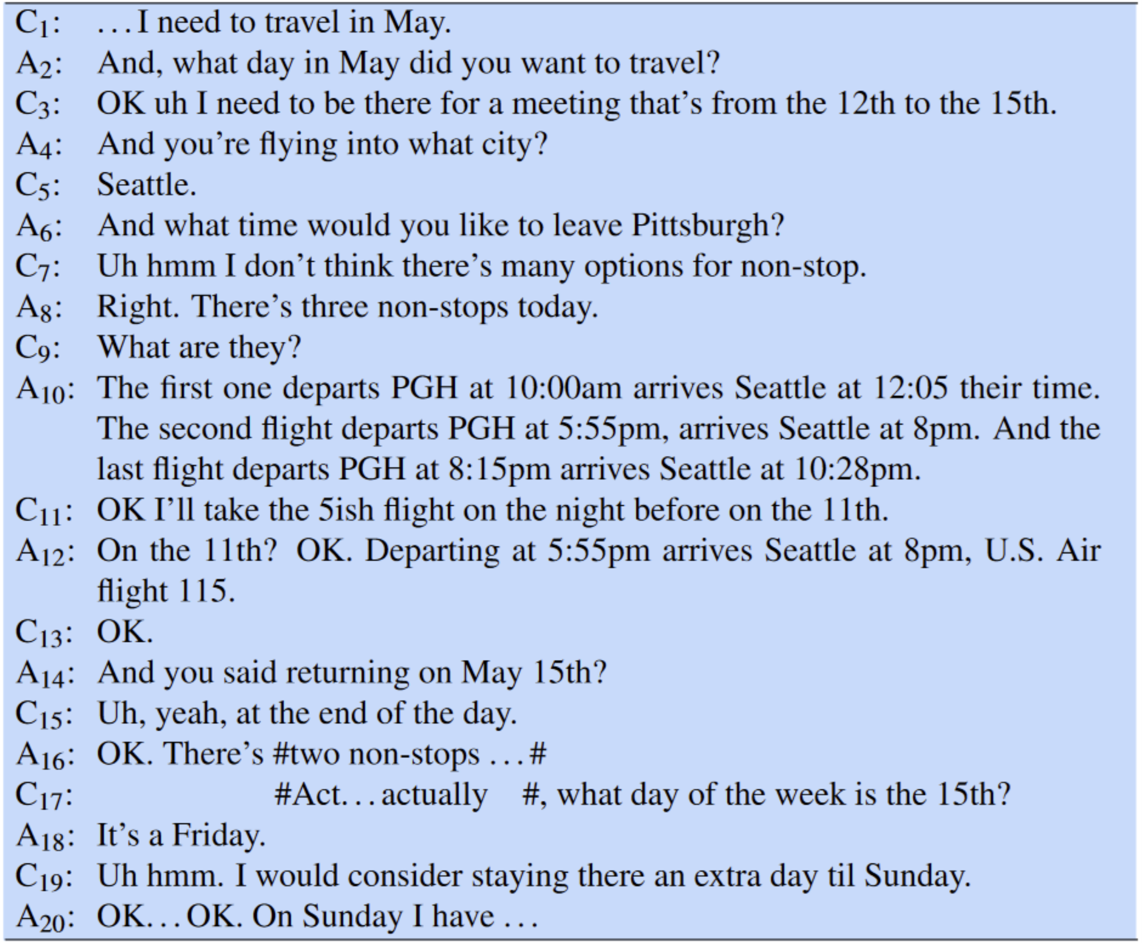


C_1: …I need to travel in May.
A_2: And, what day in May did you want to travel?
C_3: OK uh I need to be there for a meeting that's from the 12th to the 15th.
A_4: And you're flying into what city?
C_5: Seattle.
A_6: And what time would you like to leave Pittsburgh?
C_7: Uh hmm I don't think there's many options for non-stop.
A_8: Right. There's three non-stops today.
C_9: What are they?
A_{10}: 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_{11}: OK I'll take the 5ish flight on the night before on the 11th.
A_{12}: On the 11th? OK. Departing at 5:55pm arrives Seattle at 8pm, U.S. Air flight 115.
C_{13}: OK.
A_{14}: And you said returning on May 15th?
C_{15}: Uh, yeah, at the end of the day.
A_{16}: OK. There's #two non-stops …#
C_{17}: #Act…actually #, what day of the week is the 15th?
A_{18}: It's a Friday.
C_{19}: Uh hmm. I would consider staying there an extra day til Sunday.
A_{20}: OK…OK. On Sunday I have …

Figure 15.1 Part of a phone conversation between a human travel agent (A) and human client (C). The passages framed by # in A_{16} and C_{17} indicate overlaps in speech.

Figure from Jurafsky & Martin

Turn-taking

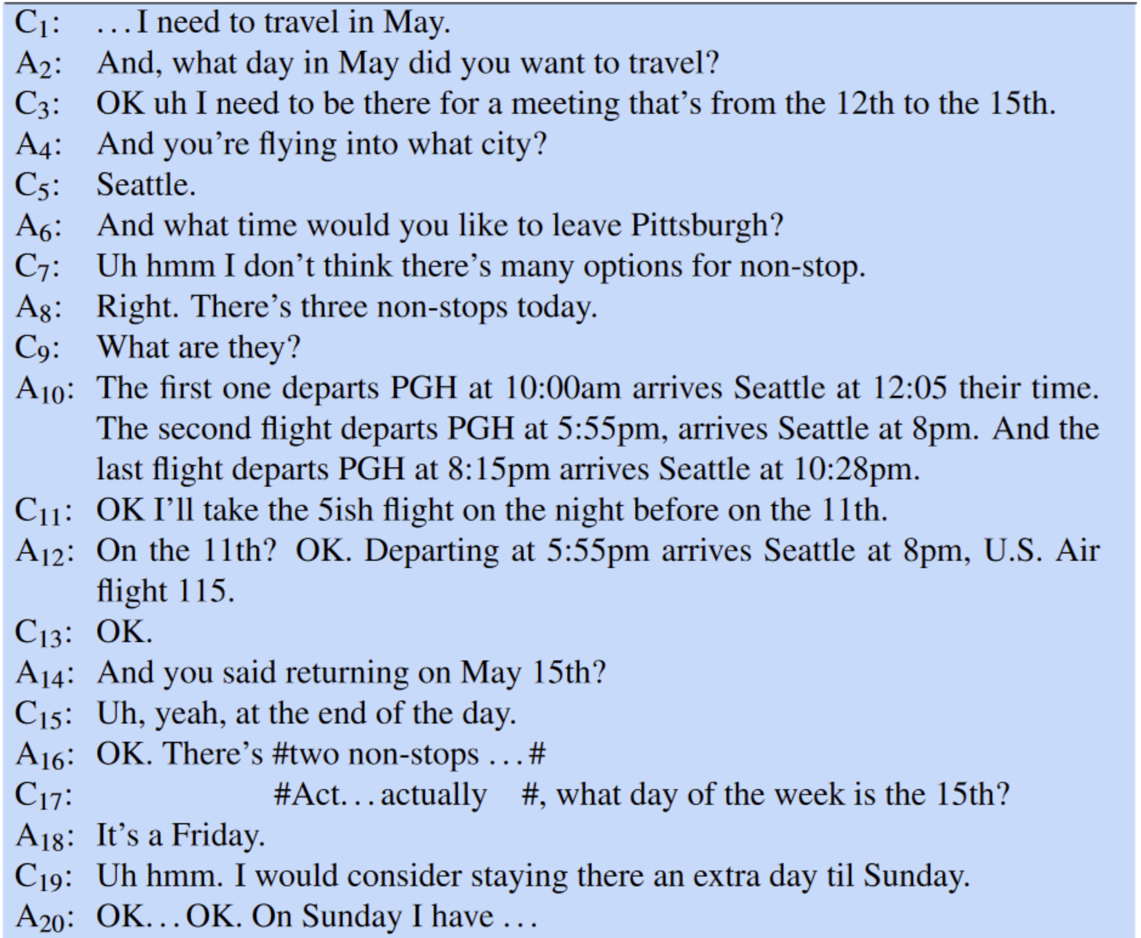


C_1: …I need to travel in May.
A_2: And, what day in May did you want to travel?
C_3: OK uh I need to be there for a meeting that's from the 12th to the 15th.
A_4: And you're flying into what city?
C_5: Seattle.
A_6: And what time would you like to leave Pittsburgh?
C_7: Uh hmm I don't think there's many options for non-stop.
A_8: Right. There's three non-stops today.
C_9: What are they?
A_{10}: 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_{11}: OK I'll take the 5ish flight on the night before on the 11th.
A_{12}: On the 11th? OK. Departing at 5:55pm arrives Seattle at 8pm, U.S. Air flight 115.
C_{13}: OK.
A_{14}: And you said returning on May 15th?
C_{15}: Uh, yeah, at the end of the day.
A_{16}: OK. There's #two non-stops …#
C_{17}: #Act…actually #, what day of the week is the 15th?
A_{18}: It's a Friday.
C_{19}: Uh hmm. I would consider staying there an extra day til Sunday.
A_{20}: OK…OK. On Sunday I have …

Figure 15.1 Part of a phone conversation between a human travel agent (A) and human client (C). The passages framed by # in A_{16} and C_{17} indicate overlaps in speech.

- A turn is a single contribution from one speaker
- Turn-taking is complex
- When to take/yield the floor?
- People can detect when their conversation partner is about to stop talking
- People interrupt each other, resulting in overlapping speech

Slide adapted from David Mortensen, Jurafsky & Martin

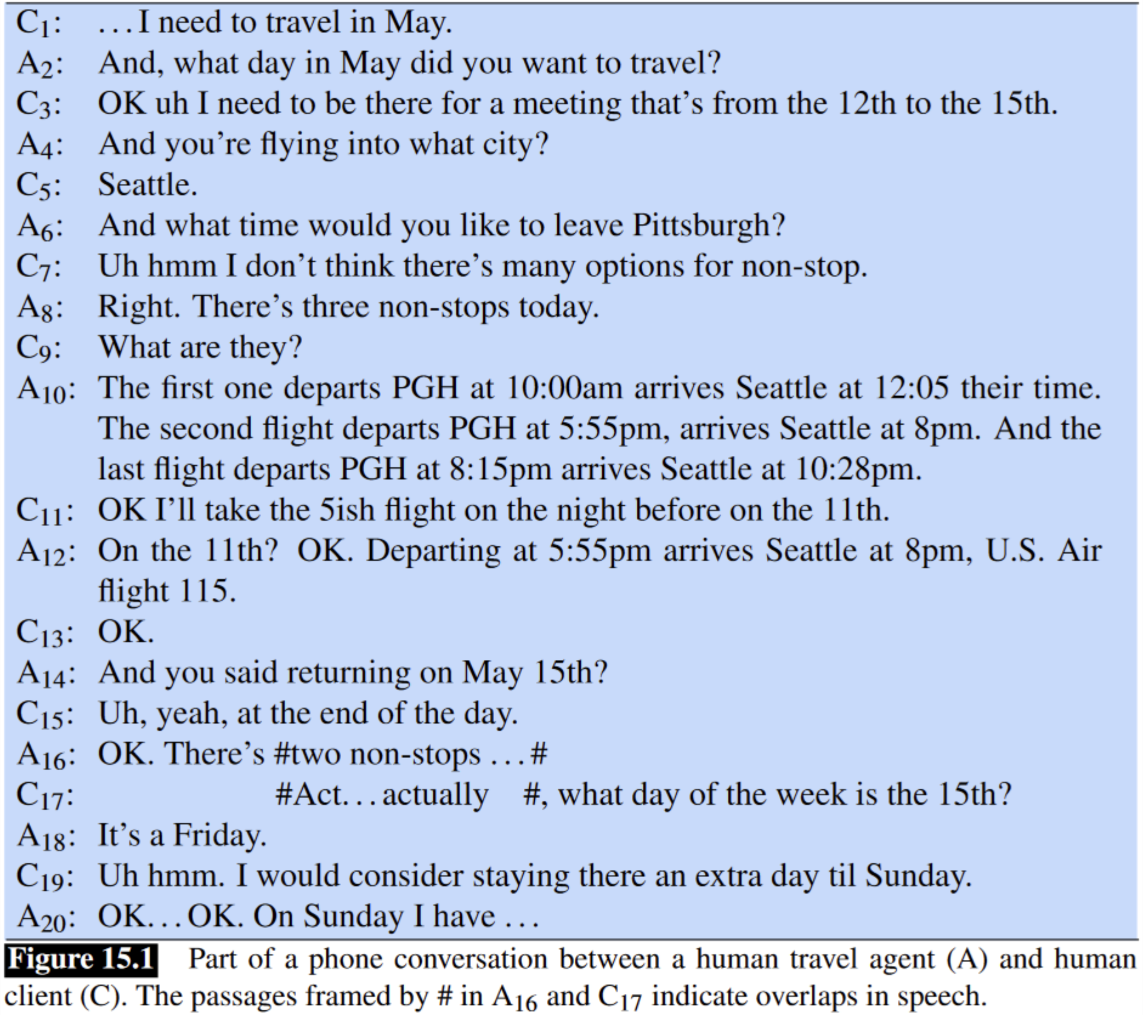


C_1: …I need to travel in May.
A_2: And, what day in May did you want to travel?
C_3: OK uh I need to be there for a meeting that’s from the 12th to the 15th.
A_4: And you’re flying into what city?
C_5: Seattle.
A_6: And what time would you like to leave Pittsburgh?
C_7: Uh hmm I don’t think there’s many options for non-stop.
A_8: Right. There’s three non-stops today.
C_9: What are they?
A_{10}: 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_{11}: OK I’ll take the 5ish flight on the night before on the 11th.
A_{12}: On the 11th? OK. Departing at 5:55pm arrives Seattle at 8pm, U.S. Air flight 115.
C_{13}: OK.
A_{14}: And you said returning on May 15th?
C_{15}: Uh, yeah, at the end of the day.
A_{16}: OK. There’s #two non-stops …#
C_{17}: #Act…actually #, what day of the week is the 15th?
A_{18}: It’s a Friday.
C_{19}: Uh hmm. I would consider staying there an extra day til Sunday.
A_{20}: OK…OK. On Sunday I have …

Figure 15.1 Part of a phone conversation between a human travel agent (A) and human client (C). The passages framed by # in A_{16} and C_{17} indicate overlaps in speech.

There are *vocal pauses* such as “uh”.

Slide adapted from David Mortensen, Jurafsky & Martin



C_1: …I need to travel in May.

A_2: And, what day in May did you want to travel?

C_3: OK uh I need to be there for a meeting that's from the 12th to the 15th.

A_4: And you're flying into what city?

C_5: Seattle.

A_6: And what time would you like to leave Pittsburgh?

C_7: Uh hmm I don't think there's many options for non-stop.

A_8: Right. There's three non-stops today.

C_9: What are they?

A_{10}: 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_{11}: OK I'll take the 5ish flight on the night before on the 11th.

A_{12}: On the 11th? OK. Departing at 5:55pm arrives Seattle at 8pm, U.S. Air flight 115.

C_{13}: OK.

A_{14}: And you said returning on May 15th?

C_{15}: Uh, yeah, at the end of the day.

A_{16}: OK. There's #two non-stops …#

C_{17}: #Act…actually #, what day of the week is the 15th?

A_{18}: It's a Friday.

C_{19}: Uh hmm. I would consider staying there an extra day til Sunday.

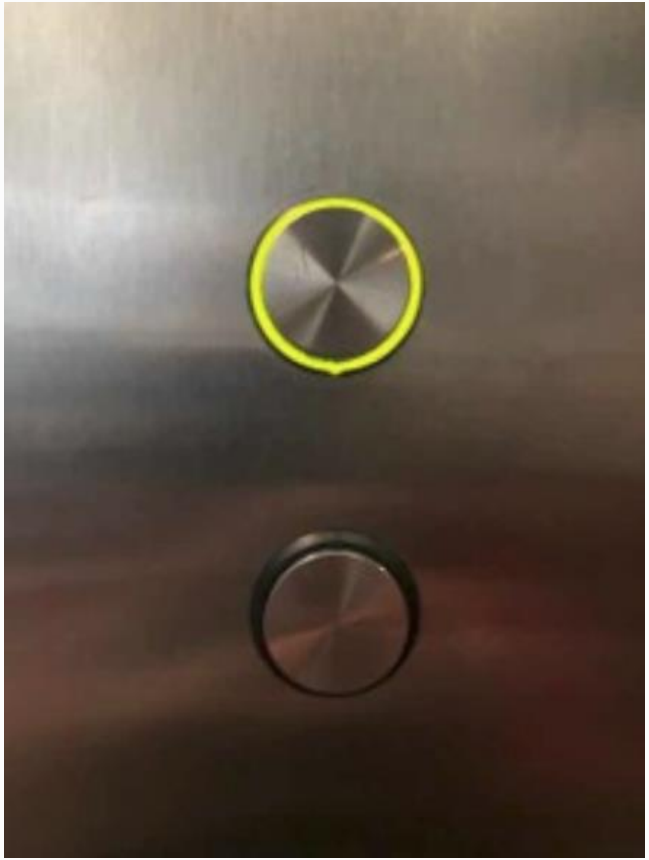
A_{20}: OK…OK. On Sunday I have …

Figure 15.1 Part of a phone conversation between a human travel agent (A) and human client (C). The passages framed by # in A_{16} and C_{17} indicate overlaps in speech.

There are *discourse markers* like "OK" and "Right".

Slide adapted from David Mortensen, Jurafsky & Martin

Grounding



Why do elevator buttons light up?

And what happens when pedestrian crosswalk buttons don't?

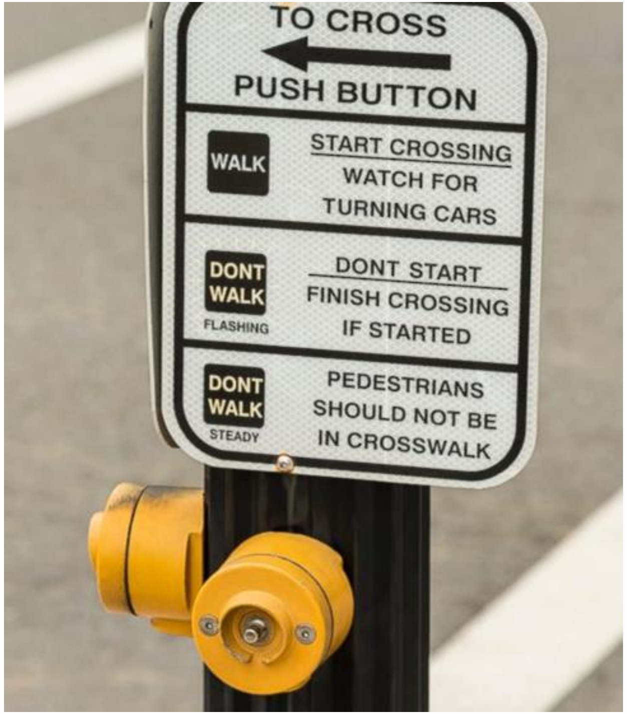


Image: ABC News

Slide adapted from David Mortensen

Grounding with Discourse Markers

A: And you said returning on May 15th?
C: Uh, yeah, at the end of the day.
A: OK

C: OK I'll take the 5ish flight on the night before on the 11th.
A: On the 11th? OK.

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

Slide credit: David Mortensen, Jurafsky & Martin

Grounding = acknowledgment

- Conversation participants need *common ground*: set of things mutually believed by both speaker and hearer
- Principle of closure: Agents performing an action require evidence, sufficient for current purposes, that they have succeeded in performing it (Clark 1996, Norman 1988)
- Speech is an action too! So speakers need to ground each other's utterances.
- Grounding: acknowledging that the hearer has understood

Slide adapted from Jurafsky & Martin

Grounding is important for computers too!

System: Did you want to review more of your profile?

User: No.

System: What’s next? AWKWARD

System: Did you want to review more of your profile?

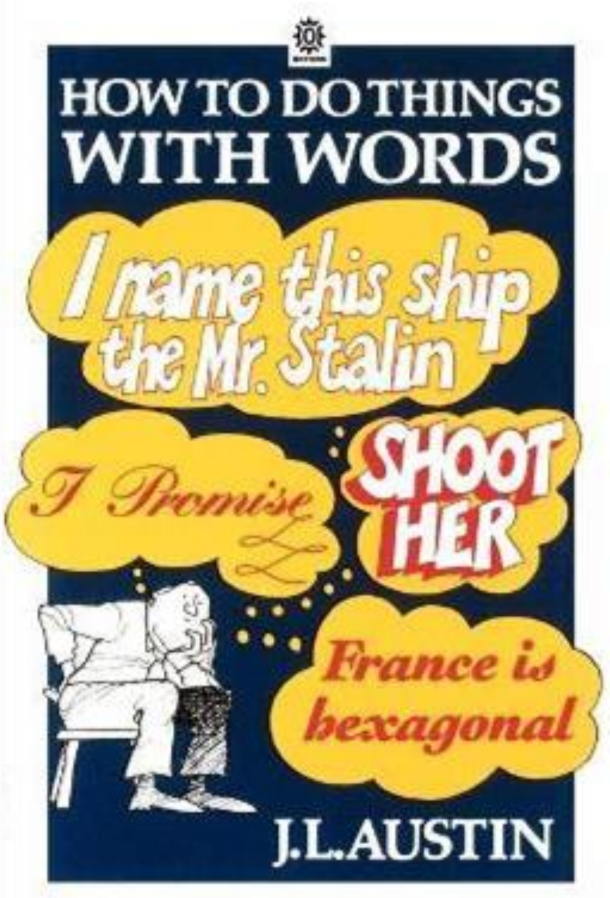
User: No.

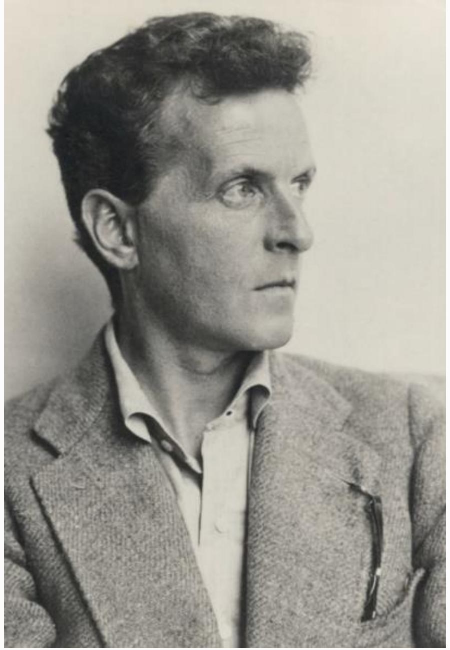
System: Okay, what’s next? LESS AWKWARD!

Slide credit: Jurafsky & Martin

Speech acts: sentences that do things

Utterances as actions




Ludwig Wittgenstein

Each turn in a dialogue is a kind of action [Wittgenstein 1953, Austin 1962]

Slide adapted from David Mortensen, Jurafsky & Martin

Speech Acts: sentences that do things

Some sentences inform: *Today is Thursday*

Here are some *performative* sentences that change the state of the world:

- *I hereby name this ship the QE2.*
  The ship now has a name.
- *I hereby bequeath this cell phone to my son.*
  My son will now inherit the cell phone.
- *I hereby take this person to be my spouse.*
  I am now married.
- *I hereby declare war.*
  There is war.
- *I hereby excommunicate you.*
  You are excommunicated.

Slide credit: David Mortensen

Intent vs form

The following three sentences all have the same intent (to get the agent to book a flight to Seattle), but they take the form of a question, a command, and a statement.

- Can you book me a flight to Seattle?
- Book me a flight to Seattle.
- I'd like a flight to Seattle.

To respond appropriately, an automated conversational agent has to understand the user's intent. Task-oriented dialogue systems include classifiers for identifying the user's intent.

Slide credit: David Mortensen

Conversations have structure

Local structure between adjacent speech acts, from the field of conversation analysis [Sacks et al. 1974]

Called adjacency pairs:

- Question > Answer
- Proposal > Acceptance/Rejection
- Compliments ("Nice jacket!") > Downplayer ("Oh, this old thing?")

Slide adapted from Jurafsky & Martin

Subdialogues

Correction subdialogue

Agent: OK. There's #two non-stops#

Client: #Act- actually#, what day of the week is the 15th?

Agent: It's a Friday.

Client: Uh hmm. I would consider staying there an extra day til Sunday.

Agent: OK...OK. On Sunday I have ...

Slide adapted from Jurafsky & Martin

Conversational initiative

- Some conversations are controlled by one person
    - A reporter interviewing a chef asks questions, and the chef responds.
    - This reporter has the **conversational initiative** (Walker and Whittaker 1990)

- Most human conversations have **mixed initiative**:
    - I lead, then you lead, then I lead.

- Mixed initiative is very hard for NLP systems, which often default to simpler styles that can be frustrating for humans:
    - **User initiative** (user asks or commands, system responds)
    - **System initiative** (system asks user questions to fill out a form, user can't change the direction)

Slide adapted from Jurafsky & Martin

Conversational implicature

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

Client: OK, uh, I need to be there for a meeting that’s from the 12th to the 15th.

Slide adapted from Jurafsky & Martin

Task-based dialogue systems

Task-based dialogue agents

"**Task-based**" or "**goal-based**" dialogue agents

- Systems that have the goal of helping a user solve a task
  - Setting a timer
  - Making a travel reservation
  - Playing a song
  - Buying a product

Slide adapted from Jurafsky & Martin

Frame-based dialogue systems: two basic architectures

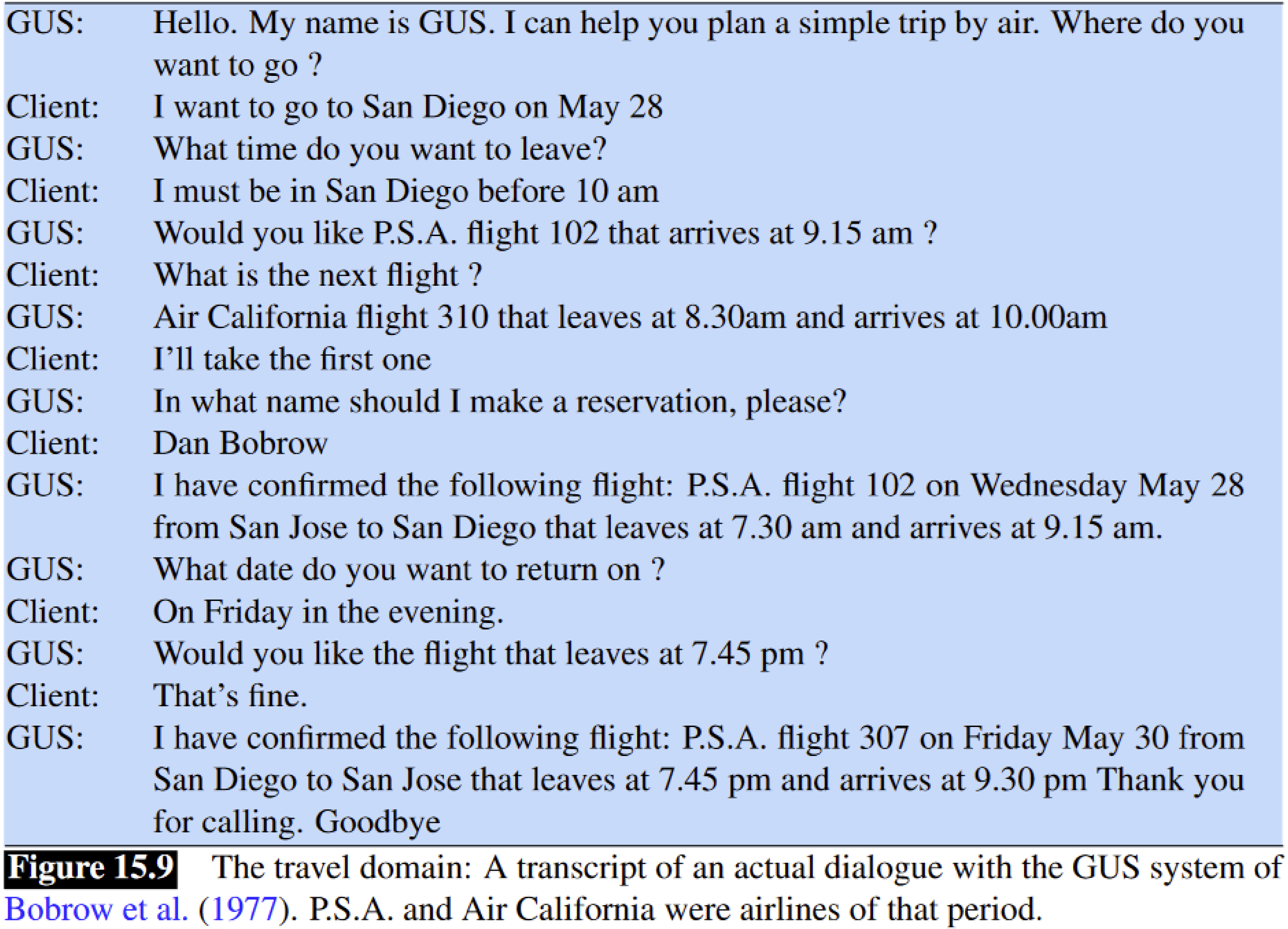
The **GUS** architecture [Bobrow et al. 1977]

- Sometimes just called "**frame-based**" architecture
- Over 40 years old, but still used in most industrial task-based dialogue agents

The **dialogue-state** architecture

- Extension of GUS
- More common in research systems
- Some aspects making their way into industrial systems

Frame-based task-oriented dialogue systems (GUS)



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

Figure 15.9 The travel domain: A transcript of an actual dialogue with the GUS system of Bobrow et al. (1977). P.S.A. and Air California were airlines of that period.

Figure credit: Jurafsky & Martin

Frames

A set of **slots**, to be filled with information of a given **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?

Slide adapted from Jurafsky & Martin

Filling a Frame

Show me morning flights from Boston to SF on Tuesday

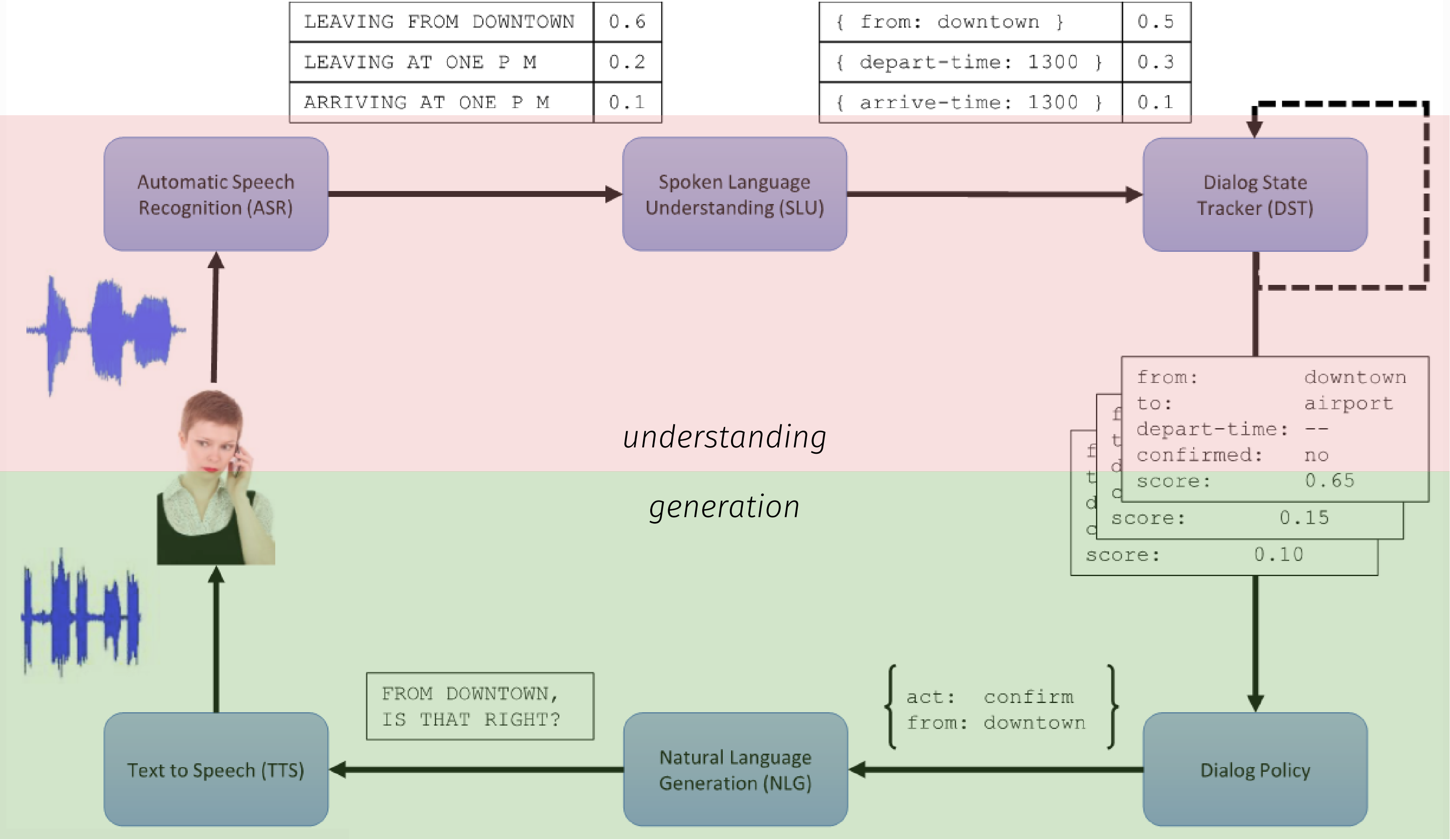
Wake me up tomorrow at six

DOMAIN:	air-travel
INTENT:	show-flights
ORIGIN-CITY:	Boston
ORIGIN-DATE:	Tuesday
ORIGIN-TIME:	morning
DEST-CITY:	San Francisco

DOMAIN:	alarm
INTENT:	set-alarm
TIME:	2022 12 1 0600

Slide credit: David Mortensen, Jurafsky & Martin

Dialogue-state architecture



LEAVING FROM DOWNTOWN	0.6
LEAVING AT ONE P M	0.2
ARRIVING AT ONE P M	0.1

{ from: downtown }	0.5
{ depart-time: 1300 }	0.3
{ arrive-time: 1300 }	0.1

Automatic Speech Recognition (ASR)

Spoken Language Understanding (SLU)

Dialog State Tracker (DST)

understanding

generation

```
from:         downtown
to:           airport
depart-time:  --
confirmed:    no
score:        0.65
```

```
score:        0.15
```

```
score:        0.10
```

Text to Speech (TTS)

```
FROM DOWNTOWN,
IS THAT RIGHT?
```

Natural Language Generation (NLG)

```
{ act:   confirm
  from:  downtown }
```

Dialog Policy

Figure from Williams et al. 2016

Components in a dialogue-state architecture

NLU: extracts slot fillers from the user's utterance using machine learning

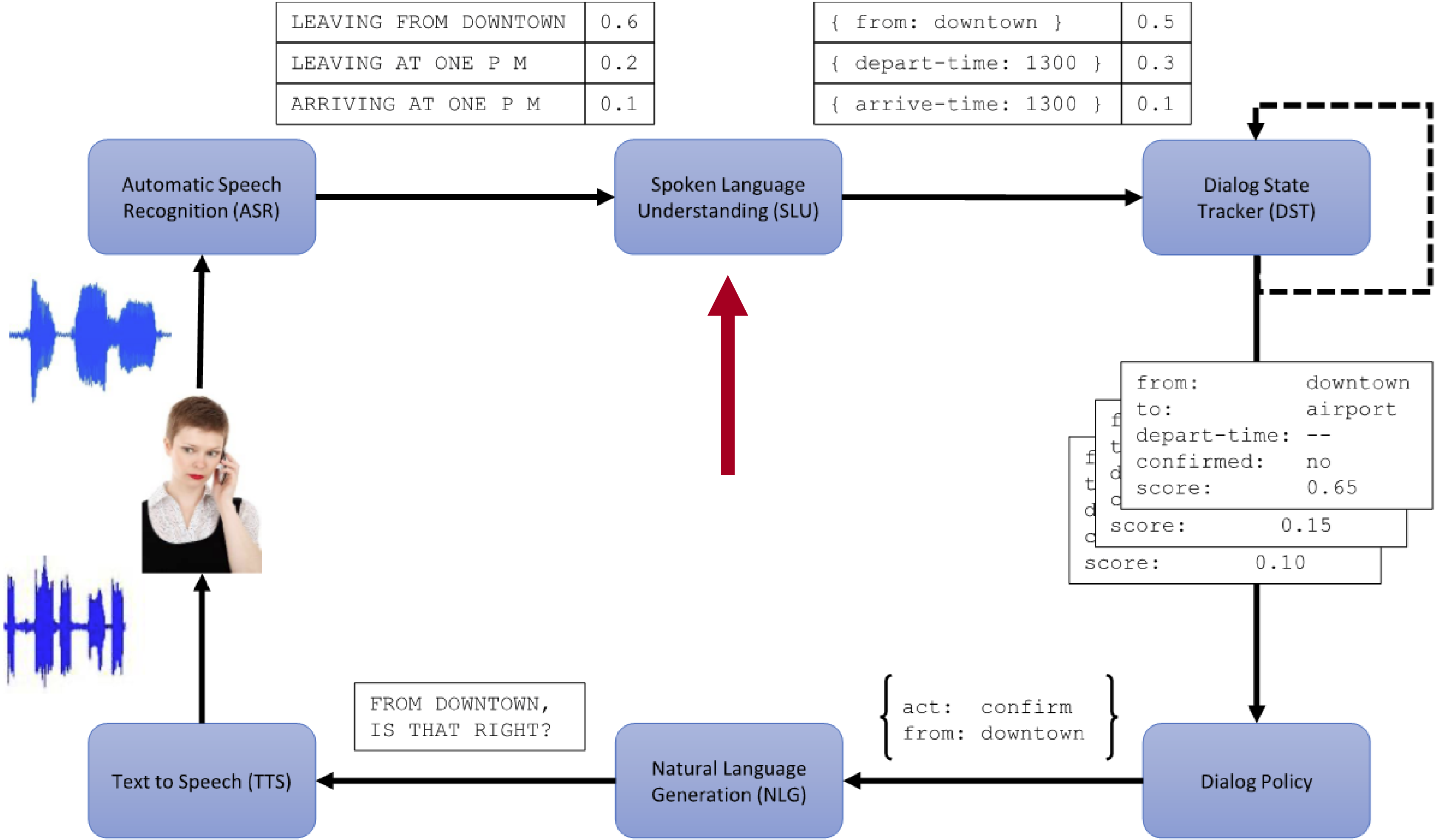
Dialogue state tracker: maintains the current state of the dialogue (user's most recent dialogue act, set of slot-filler constraints from user)

Dialogue policy: decides what the system should do or say next

- GUS policy: ask questions until the frame was full then report back
- More sophisticated: know when to answer questions, when to ask a clarification question, etc.

NLG: produce more natural, less templated utterances than GUS

Slide adapted from Jurafsky & Martin



LEAVING FROM DOWNTOWN	0.6
LEAVING AT ONE P M	0.2
ARRIVING AT ONE P M	0.1

{ from: downtown }	0.5
{ depart-time: 1300 }	0.3
{ arrive-time: 1300 }	0.1

Automatic Speech Recognition (ASR)
Spoken Language Understanding (SLU)
Dialog State Tracker (DST)
Dialog Policy
Natural Language Generation (NLG)
Text to Speech (TTS)

from: downtown
to: airport
depart-time: --
confirmed: no
score: 0.65

score: 0.15

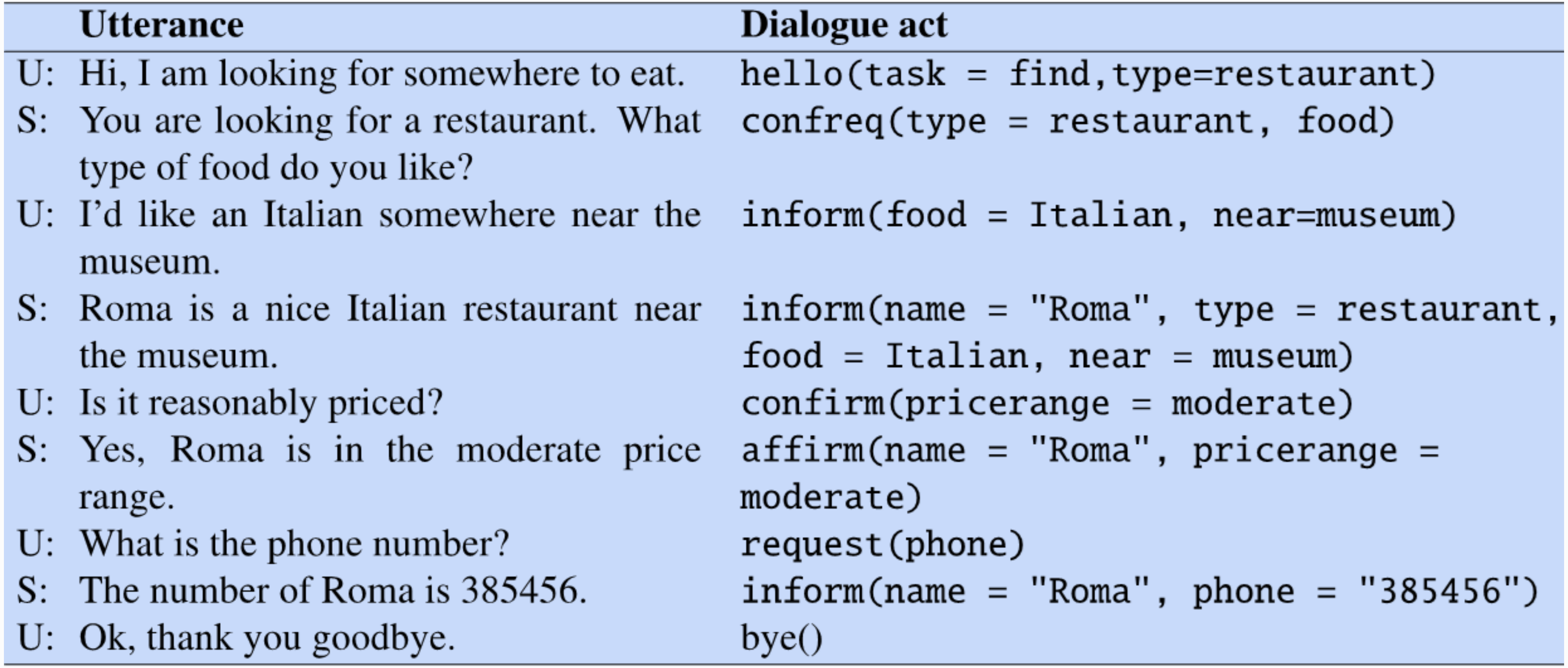
score: 0.10

{ act: confirm
from: downtown }

FROM DOWNTOWN,
IS THAT RIGHT?

Dialogue Acts

Combine the ideas of **speech acts** and **grounding** into a single representation



	Utterance	Dialogue act
U:	Hi, I am looking for somewhere to eat.	`hello(task = find,type=restaurant)`
S:	You are looking for a restaurant. What type of food do you like?	`confreq(type = restaurant, food)`
U:	I'd like an Italian somewhere near the museum.	`inform(food = Italian, near=museum)`
S:	Roma is a nice Italian restaurant near the museum.	`inform(name = "Roma", type = restaurant, food = Italian, near = museum)`
U:	Is it reasonably priced?	`confirm(pricerange = moderate)`
S:	Yes, Roma is in the moderate price range.	`affirm(name = "Roma", pricerange = moderate)`
U:	What is the phone number?	`request(phone)`
S:	The number of Roma is 385456.	`inform(name = "Roma", phone = "385456")`
U:	Ok, thank you goodbye.	bye()

Slide adapted from Jurafsky & Martin. Figure credit: Young et al. 2010

NLU: slot filling with machine learning

Machine learning classifiers to map words to semantic frame-fillers:

Input: `"I want to fly to San Francisco on Monday please"`

Output: `Destination: SF`

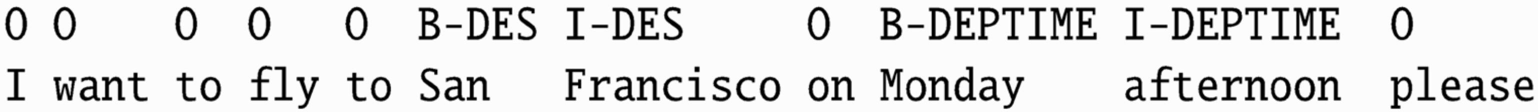
`Depart-time: Monday`

Requirements: Lots of labeled data

Slide adapted from Jurafsky & Martin

Slot filling as sequence labeling: BIO tagging

Train a classifier to label each input word with a tag that tells us what slot (if any) it fills



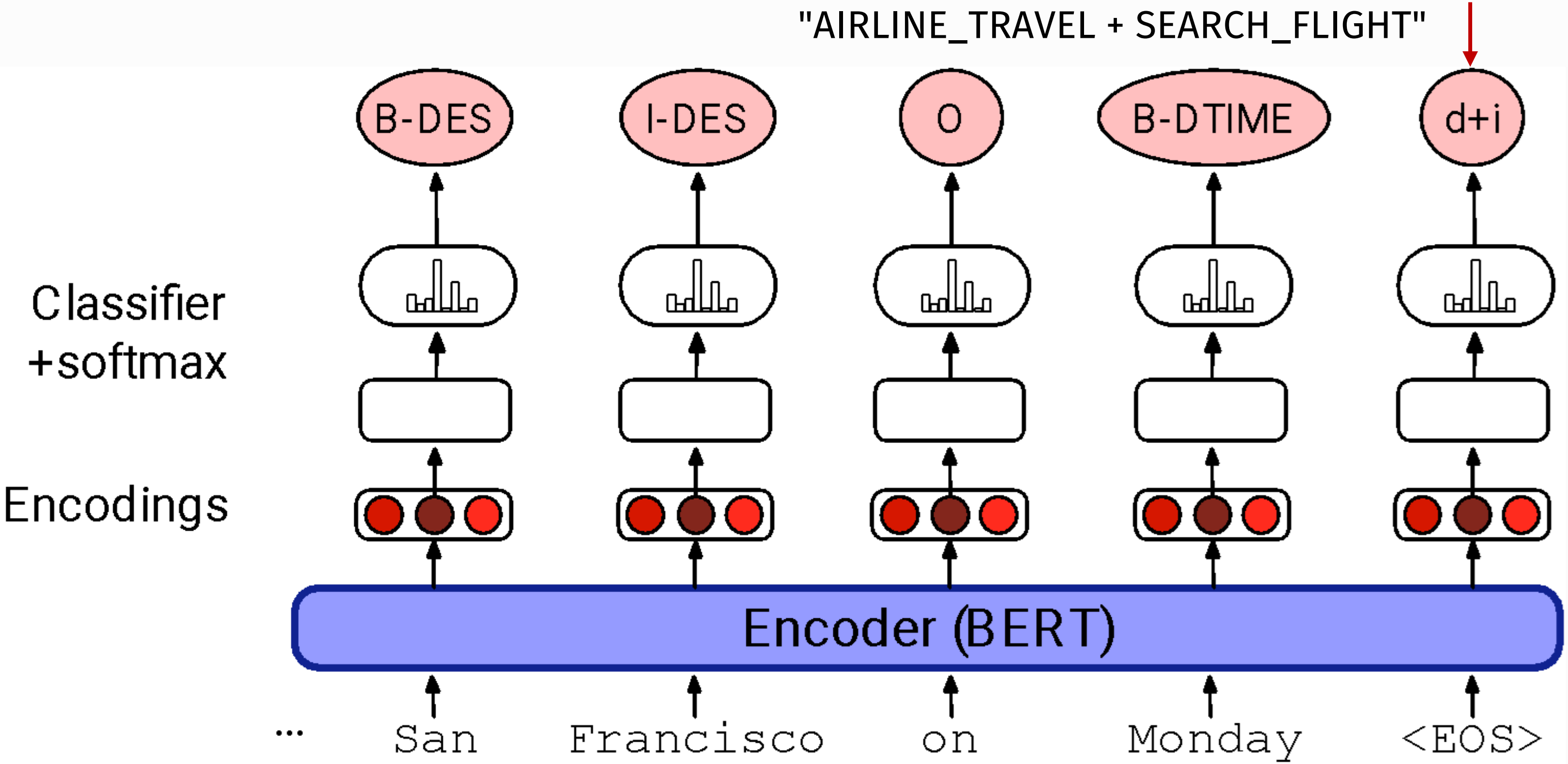
O	O	O	O	O	B-DES	I-DES	O	B-DEPTIME	I-DEPTIME	O
I	want	to	fly	to	San	Francisco	on	Monday	afternoon	please

Convert the training data to this format

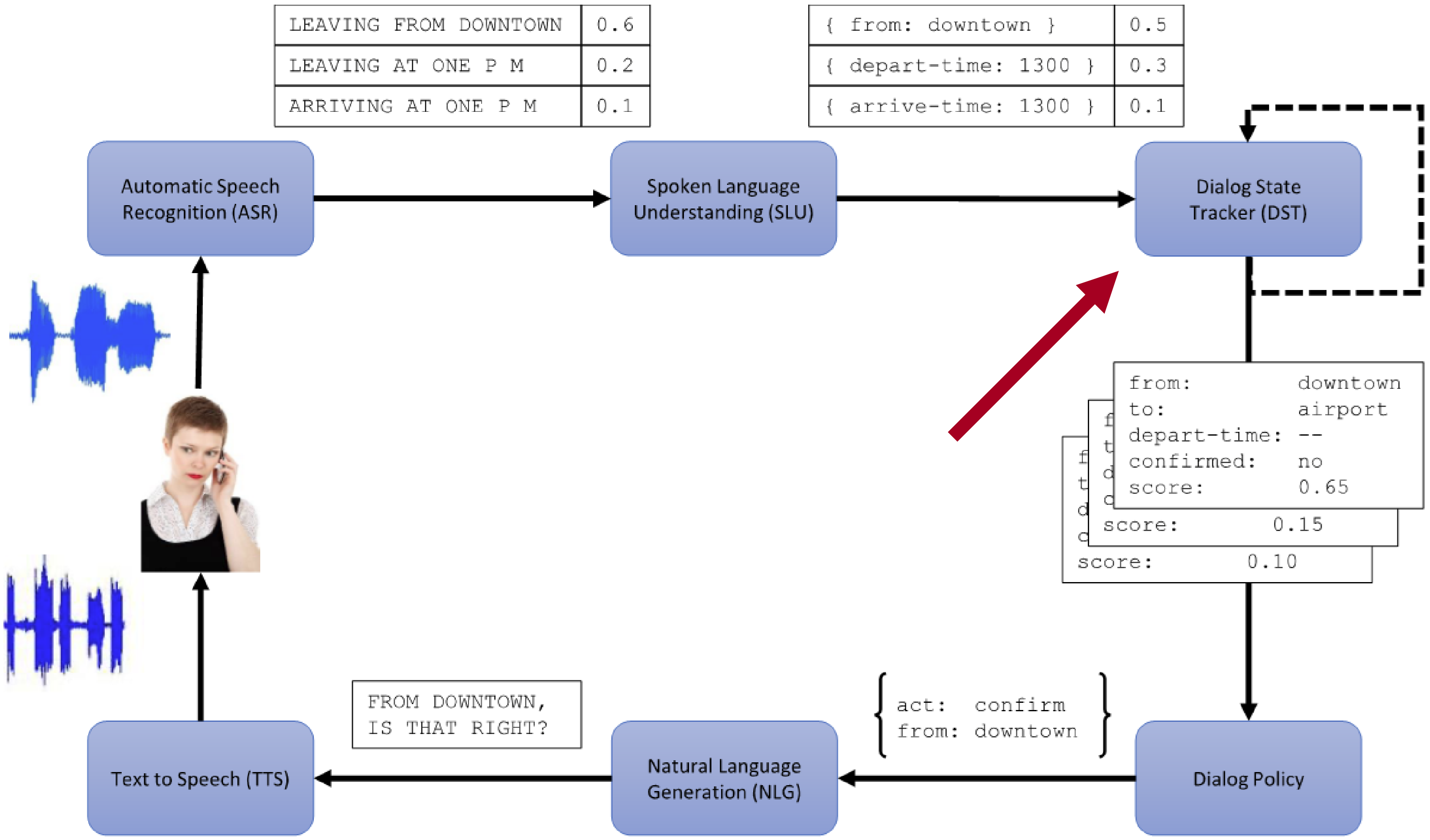
Slide adapted from Jurafsky & Martin

Slot filling using contextual embeddings

Can do domain and intent too: e.g., generate the label "AIRLINE_TRAVEL + SEARCH_FLIGHT"



Slide adapted from Jurafsky & Martin



LEAVING FROM DOWNTOWN	0.6
LEAVING AT ONE P M	0.2
ARRIVING AT ONE P M	0.1

{ from: downtown }	0.5
{ depart-time: 1300 }	0.3
{ arrive-time: 1300 }	0.1

Automatic Speech Recognition (ASR)

Spoken Language Understanding (SLU)

Dialog State Tracker (DST)

from: downtown
to: airport
depart-time: --
confirmed: no
score: 0.65

score: 0.15

score: 0.10

Dialog Policy

{ act: confirm
from: downtown }

Natural Language Generation (NLG)

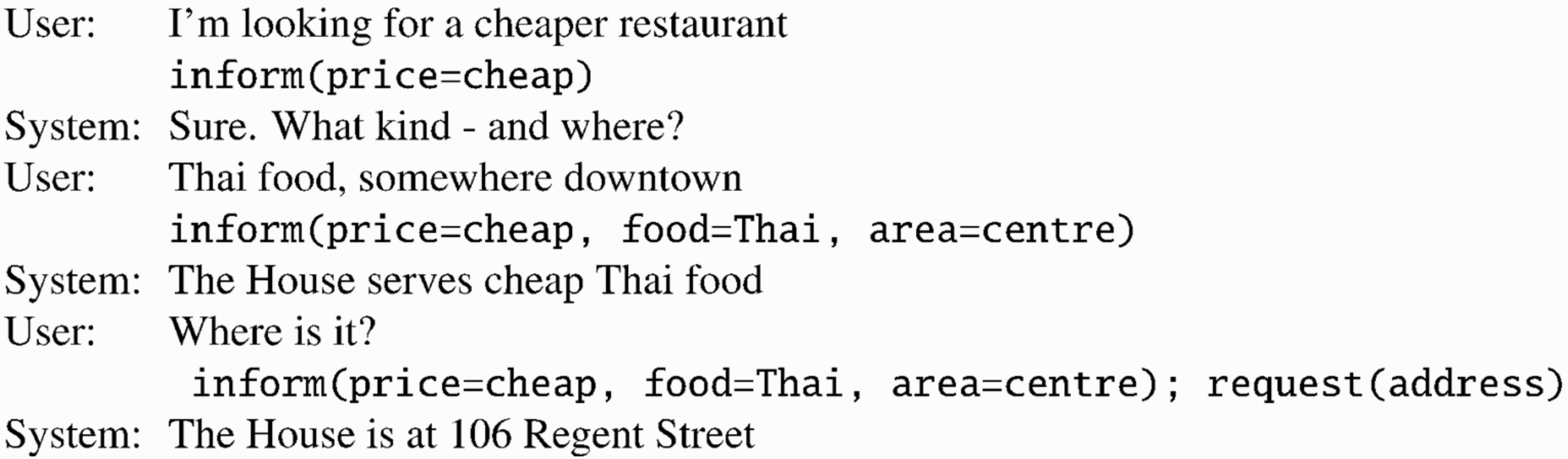
FROM DOWNTOWN,
IS THAT RIGHT?

Text to Speech (TTS)

The task of dialogue state tracking

Dialogue state:

1. Current state of the frame (slots)
2. User’s most recent dialogue act
    a. Classify based on encodings of current sentence + prior dialogue acts



User: I’m looking for a cheaper restaurant
`inform(price=cheap)`
System: Sure. What kind - and where?
User: Thai food, somewhere downtown
`inform(price=cheap, food=Thai, area=centre)`
System: The House serves cheap Thai food
User: Where is it?
`inform(price=cheap, food=Thai, area=centre); request(address)`
System: The House is at 106 Regent Street

Slide adapted from Jurafsky & Martin. Example from Mrkšić et al. 2017

A special case of dialogue act detection: correction acts

- If system misrecognizes an utterance
- User might make a correction
    - Repeat themselves
    - Rephrasing
    - Saying “no” to a confirmation question

Slide adapted from Jurafsky & Martin

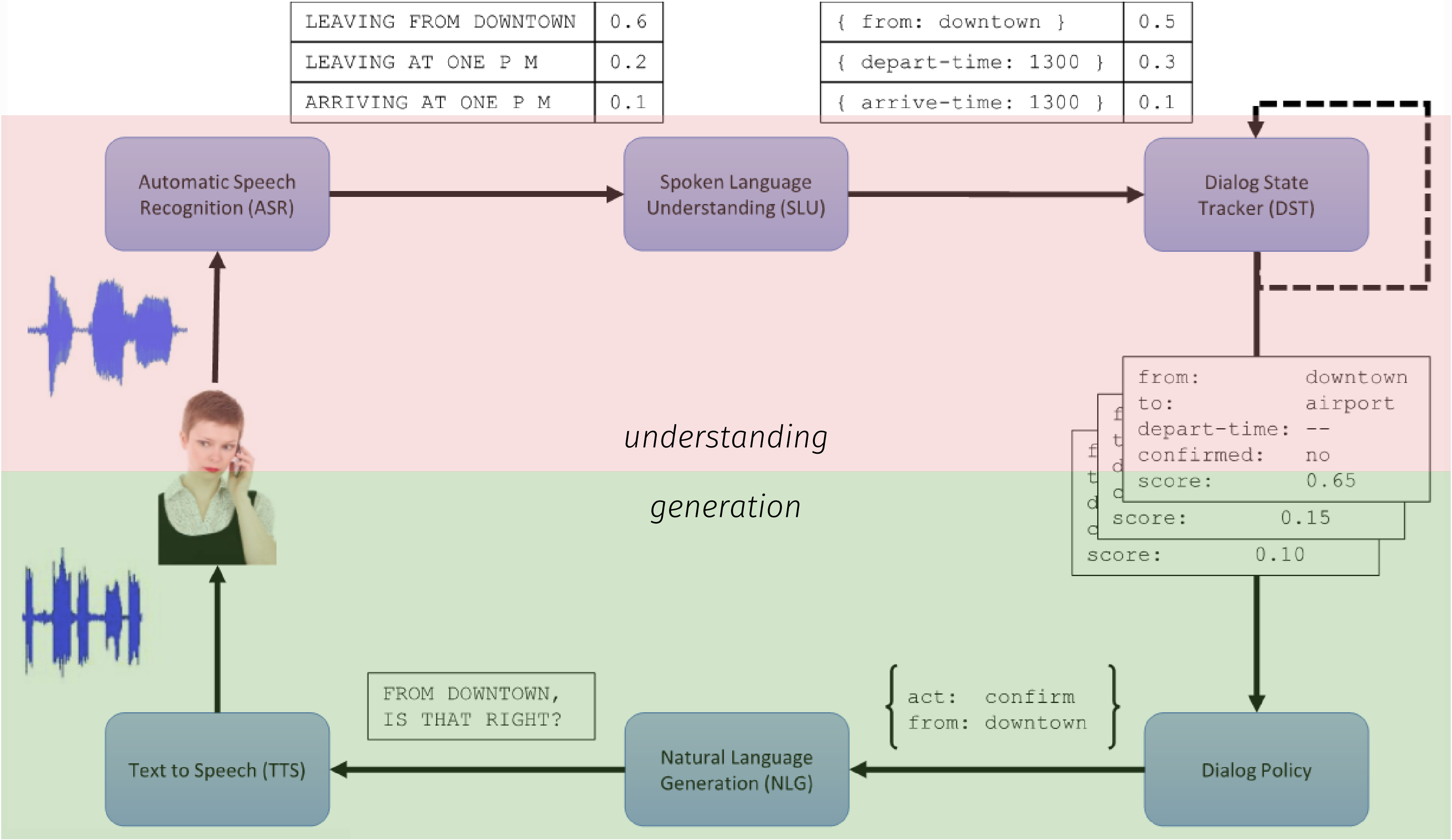
Corrections are harder to recognize!

- From speech, corrections are misrecognized twice as often (in terms of word error rate) as non-corrections! [Swerts et al. 2000]
- Hyperarticulation (exaggerated prosody) is a large factor [Shriberg et al. 1992]

"I said BAL-TI-MORE, not Boston"

- Features for detecting corrections:
  - Lexical: “no”, “correction”, “I don’t”, swear words, utterance length
  - Repeating things: high similarity between candidate correction act and user’s prior utterance (word overlap or embedding dot product)
  - Hyperarticulation, ASR confidence, language model probability

Slide adapted from Jurafsky & Martin



LEAVING FROM DOWNTOWN	0.6
LEAVING AT ONE P M	0.2
ARRIVING AT ONE P M	0.1

{ from: downtown }	0.5
{ depart-time: 1300 }	0.3
{ arrive-time: 1300 }	0.1

Automatic Speech Recognition (ASR)

Spoken Language Understanding (SLU)

Dialog State Tracker (DST)

```
from:         downtown
to:           airport
depart-time:  --
confirmed:    no
score:        0.65
```

```
score:        0.15
```

```
score:        0.10
```

understanding

generation

Text to Speech (TTS)

FROM DOWNTOWN,
IS THAT RIGHT?

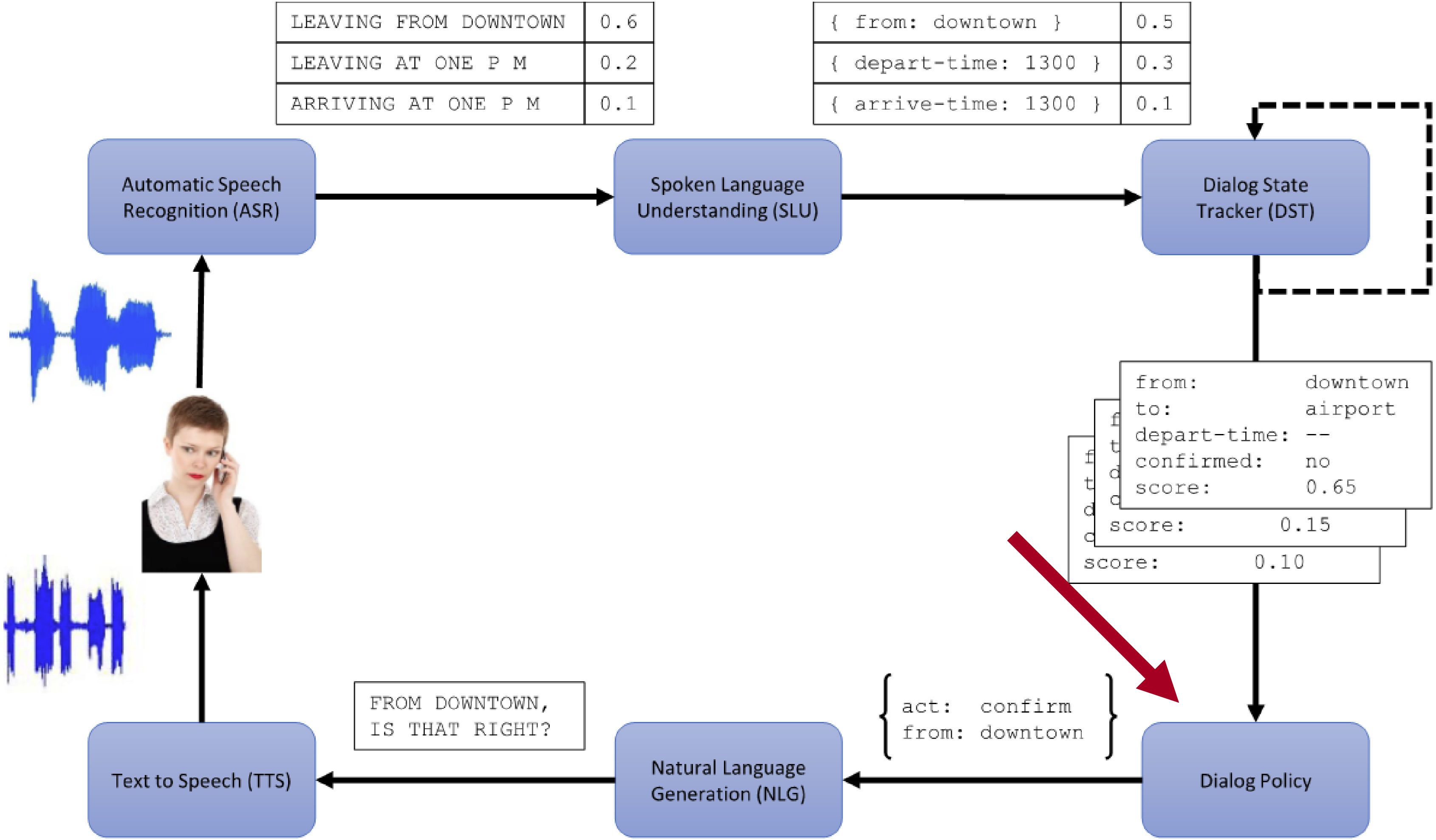
Natural Language Generation (NLG)

```
{ act:   confirm  }
{ from:  downtown }
```

Dialog Policy

Figure from Williams et al. 2016

Dialogue policies and generation



LEAVING FROM DOWNTOWN	0.6
LEAVING AT ONE P M	0.2
ARRIVING AT ONE P M	0.1

{ from: downtown }	0.5
{ depart-time: 1300 }	0.3
{ arrive-time: 1300 }	0.1

Automatic Speech Recognition (ASR)

Spoken Language Understanding (SLU)

Dialog State Tracker (DST)

from: downtown
to: airport
depart-time: --
confirmed: no
score: 0.65

score: 0.15

score: 0.10

{ act: confirm
from: downtown }

FROM DOWNTOWN,
IS THAT RIGHT?

Text to Speech (TTS)

Natural Language Generation (NLG)

Dialog Policy

Dialogue policy

- At turn *i* predict action A_i to take, given entire history.
- Simplify by just conditioning on the current dialogue state (filled frame slots) and the last turn and turn by system and user:

$$\hat{A}_i = \underset{A_i \in A}{\operatorname{argmax}} P(A_i | \text{Frame}_{i-1}, A_{i-1}, U_{i-1})$$

- Estimate probabilities by a neural classifier using neural representations of the slot fillers and utterances

Slide adapted from Jurafsky & Martin

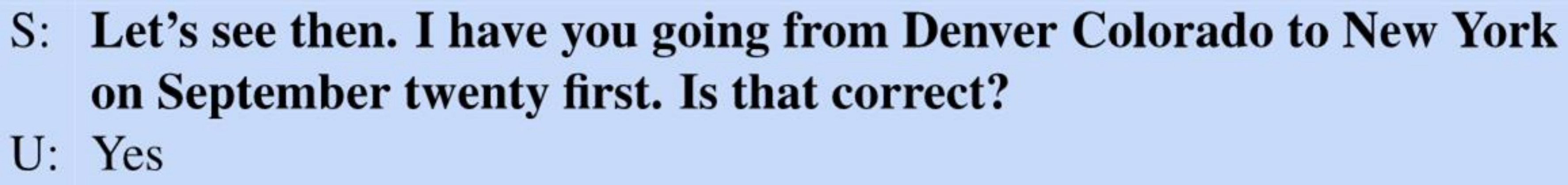
Policy example: Confirmation and rejection

- Two important mechanisms to make sure the system has understood the user:
    - **confirming** understandings with the user
    - **rejecting** utterances that the system is likely to have misunderstood.

Slide adapted from Jurafsky & Martin

Explicit vs implicit confirmation

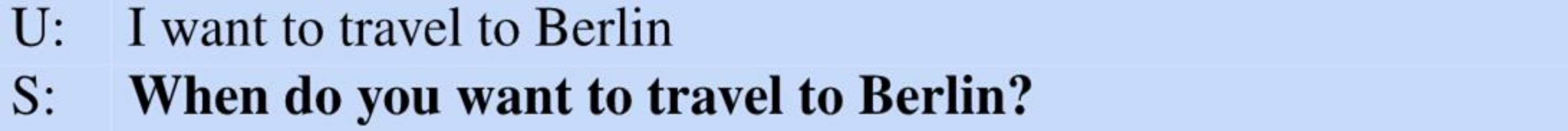
Explicit



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

U: Yes

Implicit:



U: I want to travel to Berlin

S: **When do you want to travel to Berlin?**

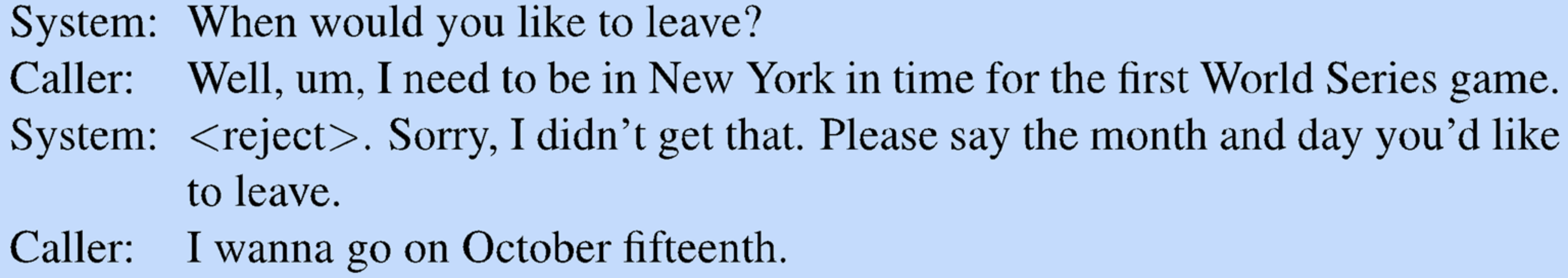
Explicit confirmation makes it easier for the user to correct issues, but implicit is more natural [Danieli and Gerbino 1995, Walker et al. 1998].

Slide adapted from Jurafsky & Martin

Rejection

I’m sorry, I didn’t understand that.

- Progressive prompting for rejection: give the user guidance on how to respond



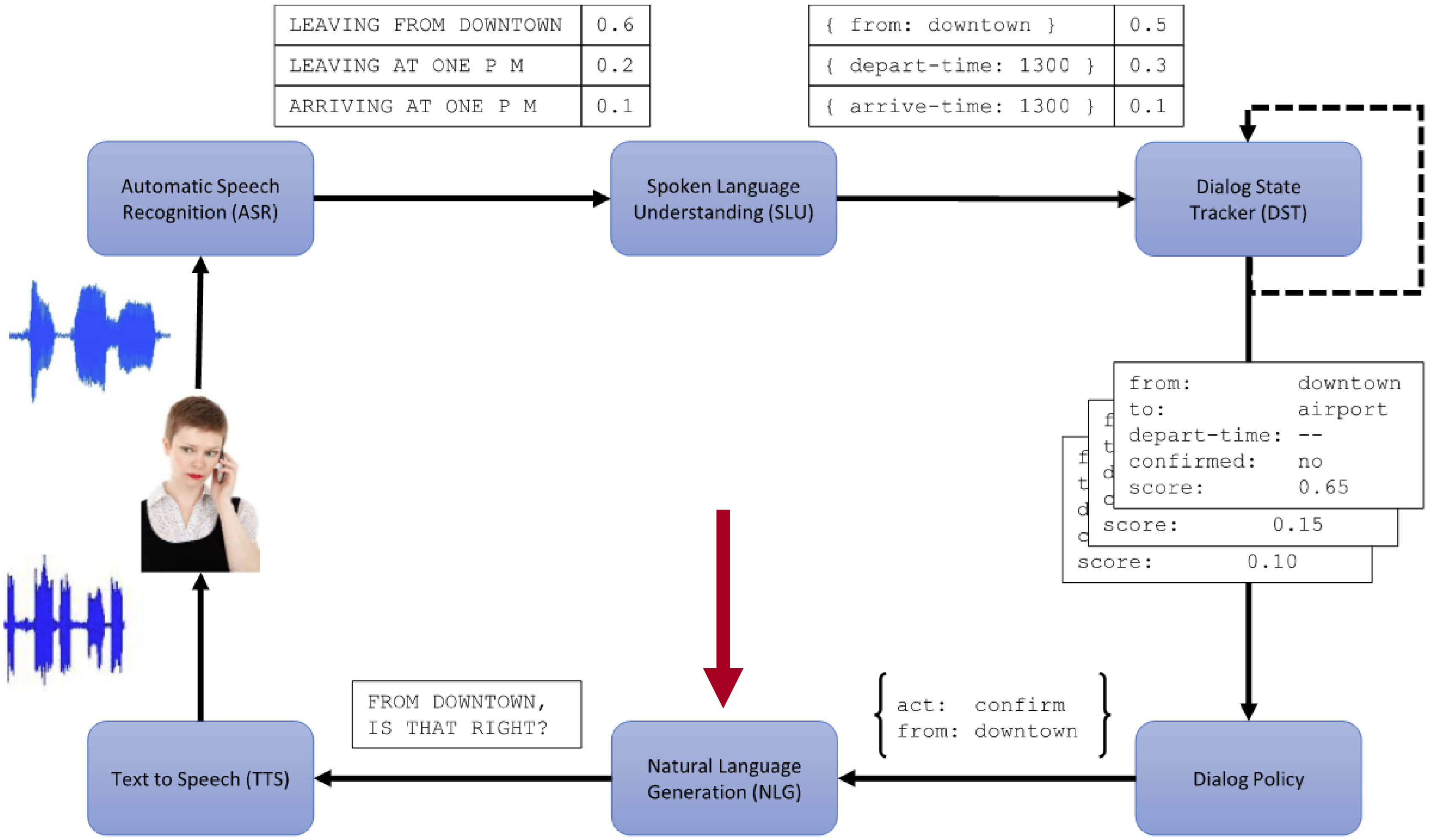
System: When would you like to leave?

Caller: Well, um, I need to be in New York in time for the first World Series game.

System: <reject>. Sorry, I didn’t get that. Please say the month and day you’d like to leave.

Caller: I wanna go on October fifteenth.

Slide adapted from Jurafsky & Martin



LEAVING FROM DOWNTOWN	0.6
LEAVING AT ONE P M	0.2
ARRIVING AT ONE P M	0.1

{ from: downtown }	0.5
{ depart-time: 1300 }	0.3
{ arrive-time: 1300 }	0.1

Automatic Speech Recognition (ASR)
Spoken Language Understanding (SLU)
Dialog State Tracker (DST)
Dialog Policy
Natural Language Generation (NLG)
Text to Speech (TTS)

```
from:          downtown
to:            airport
depart-time:   --
confirmed:     no
score:         0.65
```

```
score:         0.15
```

```
score:         0.10
```

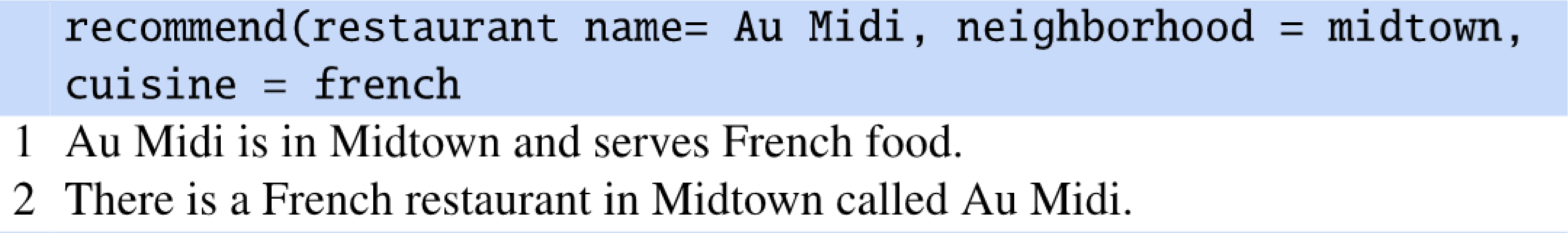
```
{ act:   confirm
  from:  downtown }
```

```
FROM DOWNTOWN,
IS THAT RIGHT?
```

NLG: sentence realization

Input: content from the dialogue policy prediction

Output: fully formed sentences



	`recommend(restaurant name= Au Midi, neighborhood = midtown, cuisine = french`
1	Au Midi is in Midtown and serves French food.
2	There is a French restaurant in Midtown called Au Midi.

Training data is hard to come by

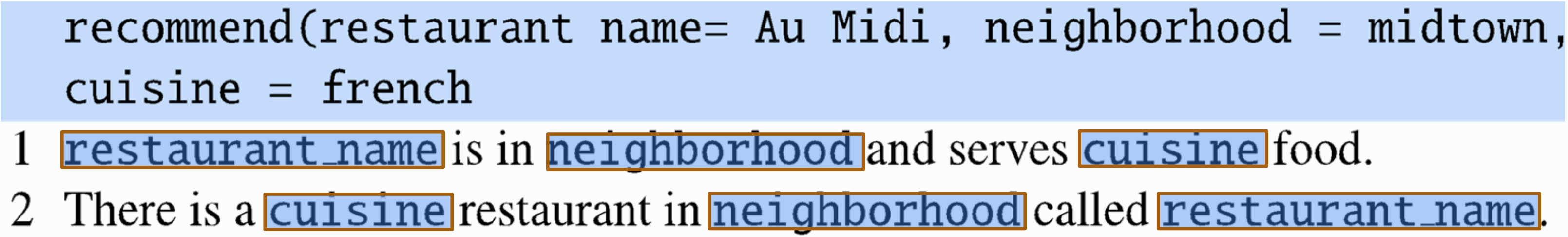
- Don't see each restaurant in each situation

Slide adapted from Jurafsky & Martin

NLG: sentence realization

Common way to improve generalization:

- **Delexicalization**: replacing words in the training set that represent slot values with a generic placeholder token



```
recommend(restaurant name= Au Midi, neighborhood = midtown,
cuisine = french
```

1 `restaurant_name` is in `neighborhood` and serves `cuisine` food.

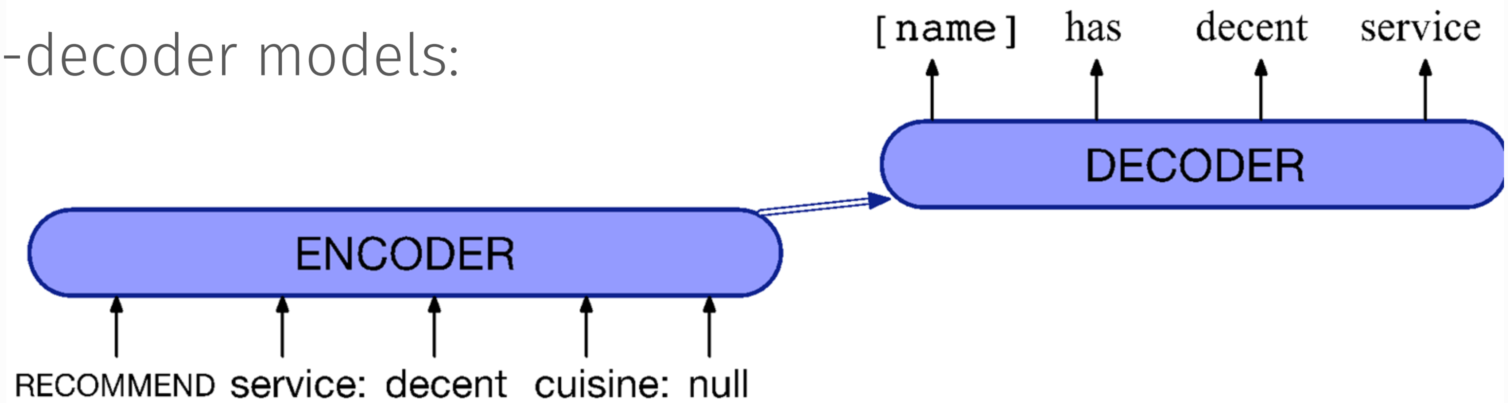
2 There is a `cuisine` restaurant in `neighborhood` called `restaurant_name`.

Slide adapted from Jurafsky & Martin

NLG: sentence realization

Mapping from frames to delexicalized sentences

Encoder-decoder models:



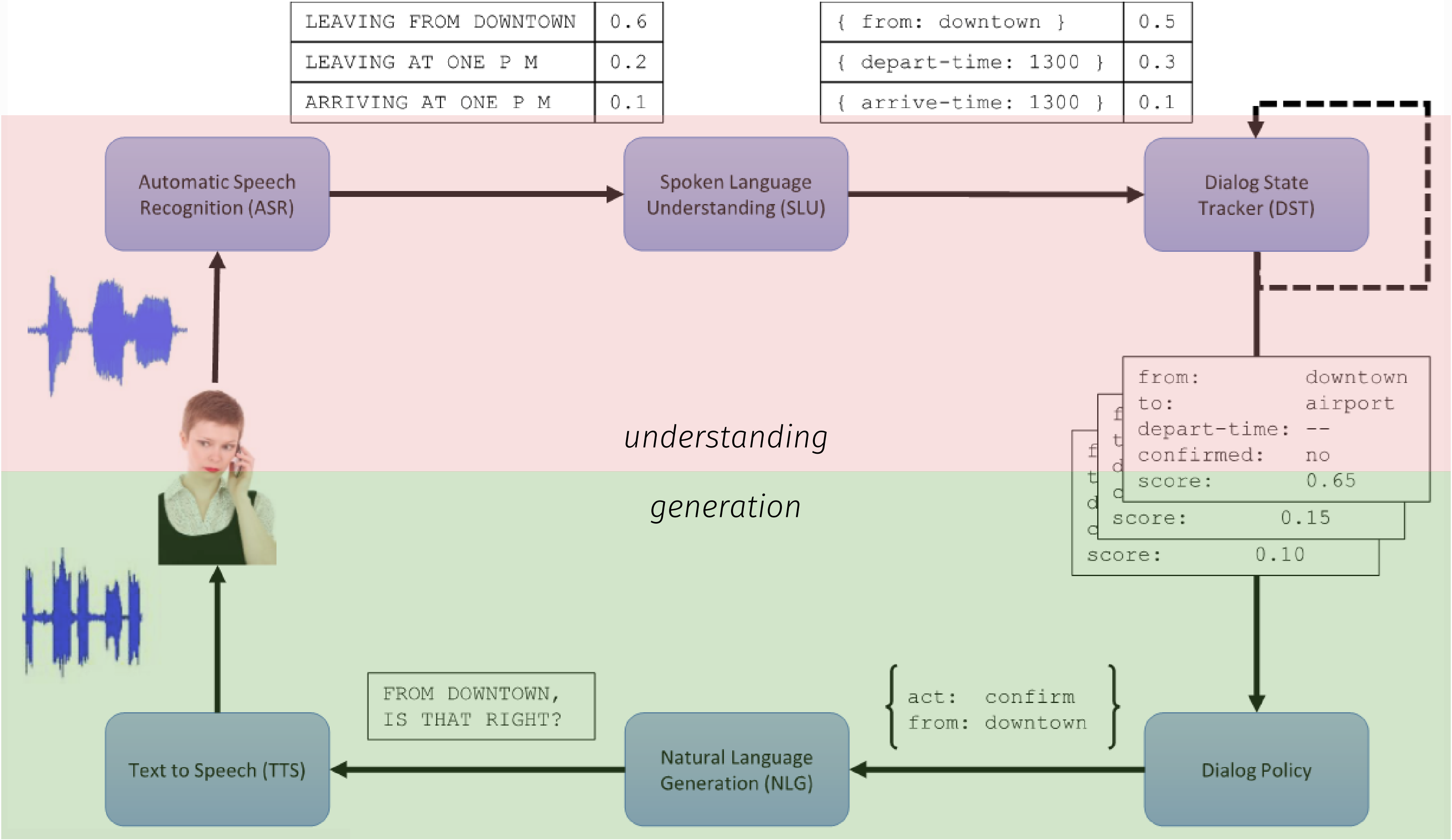
Output:

`restaurant_name has decent service`

Relexicalize to:

`Au Midi has decent service`

Slide adapted from Jurafsky & Martin



LEAVING FROM DOWNTOWN	0.6
LEAVING AT ONE P M	0.2
ARRIVING AT ONE P M	0.1

{ from: downtown }	0.5
{ depart-time: 1300 }	0.3
{ arrive-time: 1300 }	0.1

Automatic Speech Recognition (ASR)

Spoken Language Understanding (SLU)

Dialog State Tracker (DST)

```
from:         downtown
to:           airport
depart-time:  --
confirmed:    no
score:        0.65
```

```
score:        0.15
```

```
score:        0.10
```

understanding

generation

```
FROM DOWNTOWN,
IS THAT RIGHT?
```

```
{ act:   confirm
  from:  downtown }
```

Text to Speech (TTS)

Natural Language Generation (NLG)

Dialog Policy

Figure from Williams et al. 2016

Evaluating dialogue systems

Task-based systems are evaluated by task success!

“Make an appointment with Chris at 10:30 in Gates 104”

Slot	Filler
PERSON	Chris
TIME	11:30 a.m.
ROOM	Gates 104

Slot error rate: 1/3
Task success: At end, was the correct meeting added to the calendar?
Efficiency/quality: how many turns total? how many turns to correct errors?

Slide adapted from Jurafsky & Martin

Evaluate a task-based dialogue system

Options:

- United Airlines https://www.united.com/en/us/fly/help-center.html
  - Click “Chat with us”
- Amtrak’s Julie
  - https://www.amtrak.com/contact-us
- Ben: PA Health and Human Services COMPASS chat
  - https://www.compass.dhs.pa.gov
  - Click the chat robot icon in the bottom right corner next to “Need help?”
- Another automated chat service from a company you know of

Chat with the system for a few turns. Consider these questions:

- How do they seem to determine user intent? (dialogue acts)
- Can you tell what slots they're trying to fill? How do they prompt the user about those slots?
- How do they handle input that is unexpected?
- Does any of its responses seem “unnatural”?
- Anything else you notice

Wrapping up

- Automated conversational systems can be divided into 2 types:
    - Open-domain “chatbots”
    - Task-oriented dialogue systems
- Conversation is a complex joint interaction between participants
    - Turn-taking and grounding are example issues that dialogue systems must address
- Task-based dialogue systems are often filling “frames” of needed information from the user to complete a task
- Dialogue-state architecture includes NLU, NLG, slot-filling, dialogue act and dialogue policy classification
- Evaluation of task-based dialogue systems includes measuring task success and efficiency

Questions?