Dialogue Systems
Methods in AI research

Dong Nguyen
Sept 2020

Utrecht University

Practicalities

Literature for today:

Speech and Language Processing (3rd ed. draft) by Jurafsky & Martin:

- Chapter 26: Dialog systems and chatbots.

Conversational agents

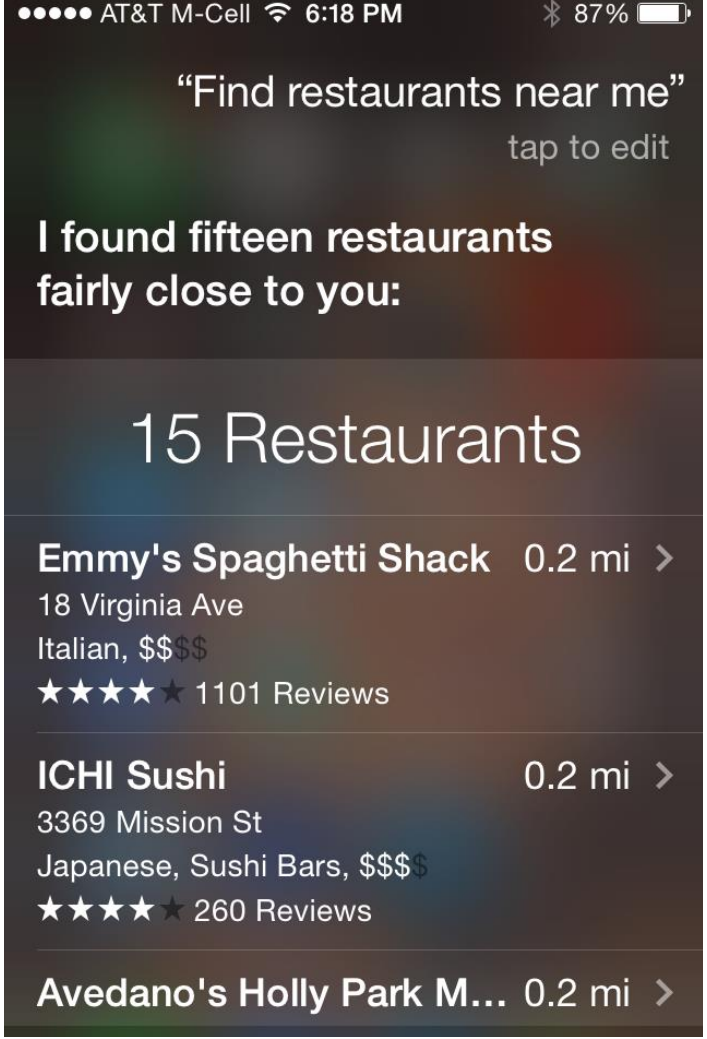
(aka dialogue systems)

- Conversation and communication core to AI
  - Importance of the Turing test (does a human know they’re talking to a computer?)

- Bring major components of AI together: knowledge, reasoning, language understanding, learning, …

Course project: “You will design, implement, evaluate and write about a restaurant recommendations dialog system”

SIRI
around
2014


●●●●● AT&T M-Cell 6:18 PM 87%

“Find restaurants near me”

tap to edit

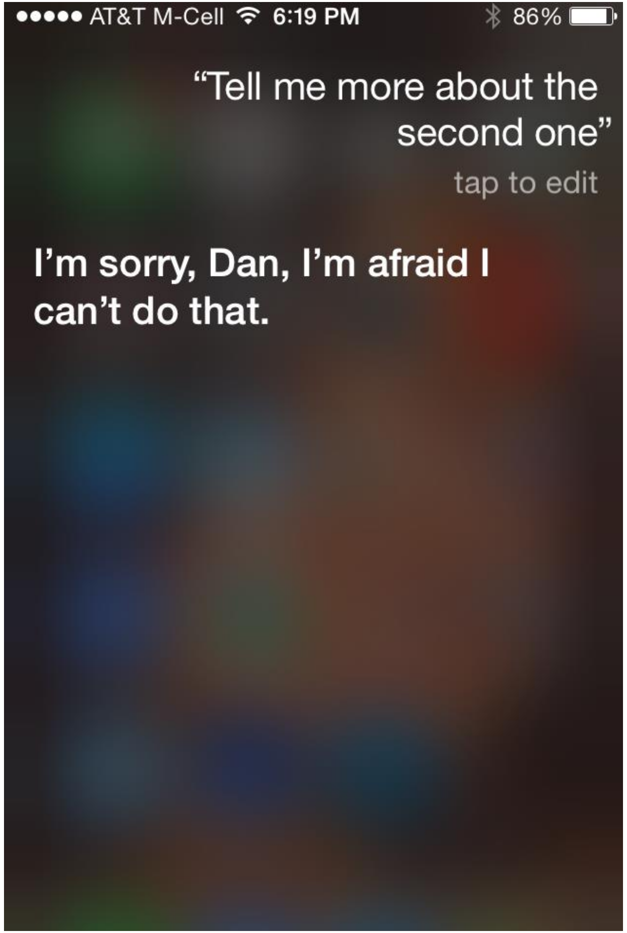
I found fifteen restaurants fairly close to you:

15 Restaurants

Emmy's Spaghetti Shack 0.2 mi
18 Virginia Ave
Italian, $$$$
★★★★★ 1101 Reviews

ICHI Sushi 0.2 mi
3369 Mission St
Japanese, Sushi Bars, $$$$
★★★★★ 260 Reviews

Avedano's Holly Park M... 0.2 mi

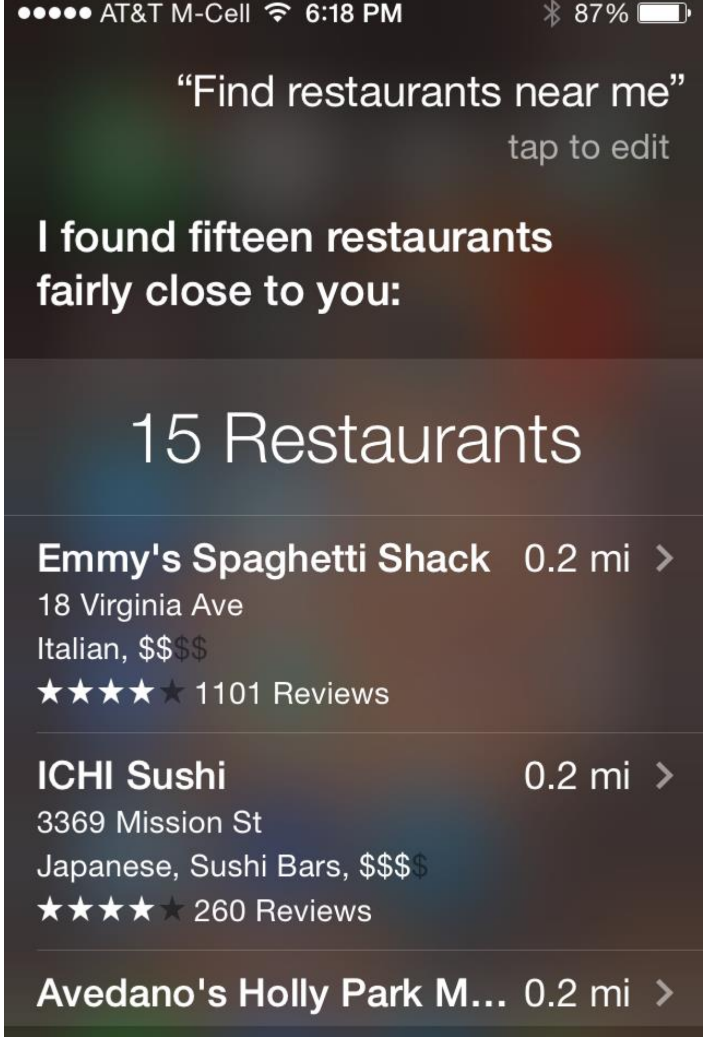

●●●●● AT&T M-Cell 6:19 PM 86%

“Tell me more about the second one”

tap to edit

I’m sorry, Dan, I’m afraid I can’t do that.

SIRI
around
2014

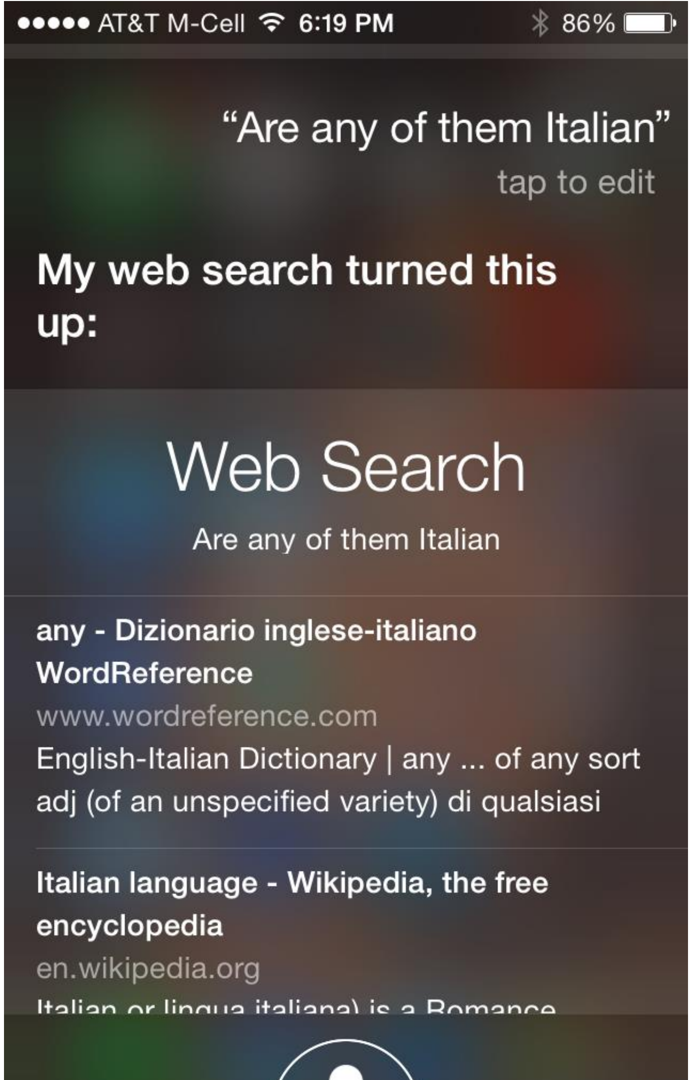


"Find restaurants near me"

I found fifteen restaurants fairly close to you:

15 Restaurants

Emmy's Spaghetti Shack 18 Virginia Ave Italian, $$$$ ★★★★★ 1101 Reviews	0.2 mi
ICHI Sushi 3369 Mission St Japanese, Sushi Bars, $$$$ ★★★★★ 260 Reviews	0.2 mi
Avedano's Holly Park M...	0.2 mi



"Are any of them Italian"

My web search turned this up:

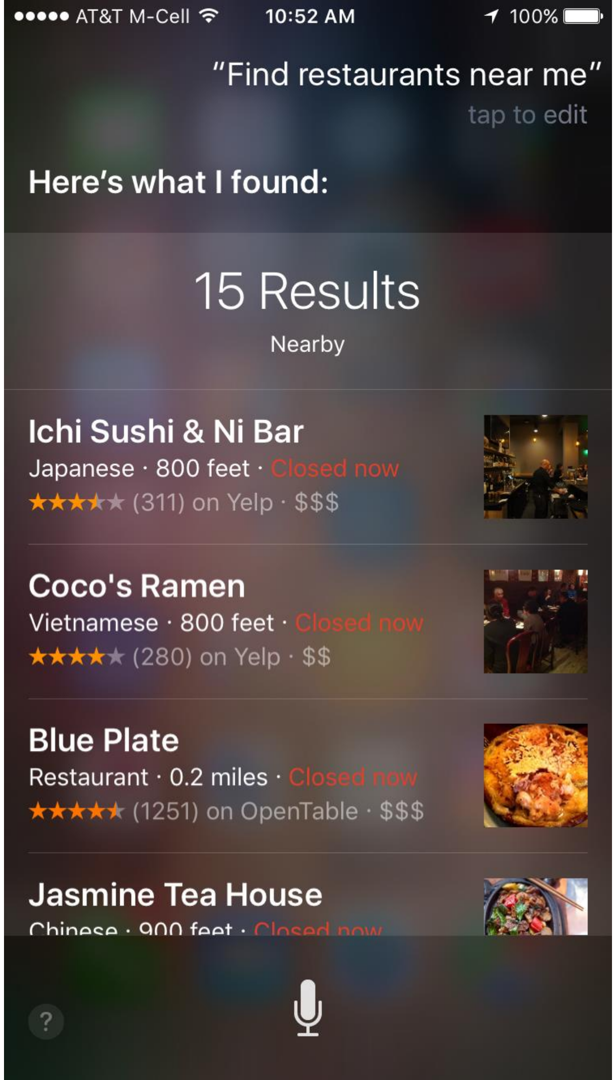
Web Search

Are any of them Italian

any - Dizionario inglese-italiano WordReference
www.wordreference.com
English-Italian Dictionary | any ... of any sort adj (of an unspecified variety) di qualsiasi

Italian language - Wikipedia, the free encyclopedia
en.wikipedia.org
Italian or lingua italiana) is a Romance

SIRI
around
2017



"Find restaurants near me"

Here's what I found:

15 Results

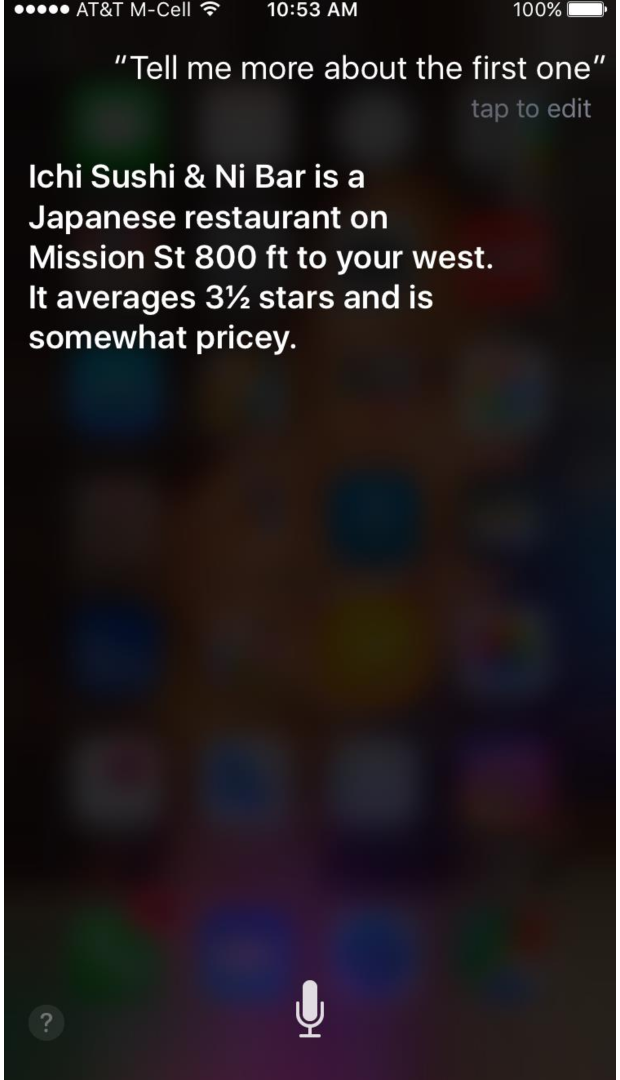
Nearby

Ichi Sushi & Ni Bar
Japanese · 800 feet · Closed now
★★★½☆ (311) on Yelp · $$$

Coco's Ramen
Vietnamese · 800 feet · Closed now
★★★★☆ (280) on Yelp · $$

Blue Plate
Restaurant · 0.2 miles · Closed now
★★★★½ (1251) on OpenTable · $$$

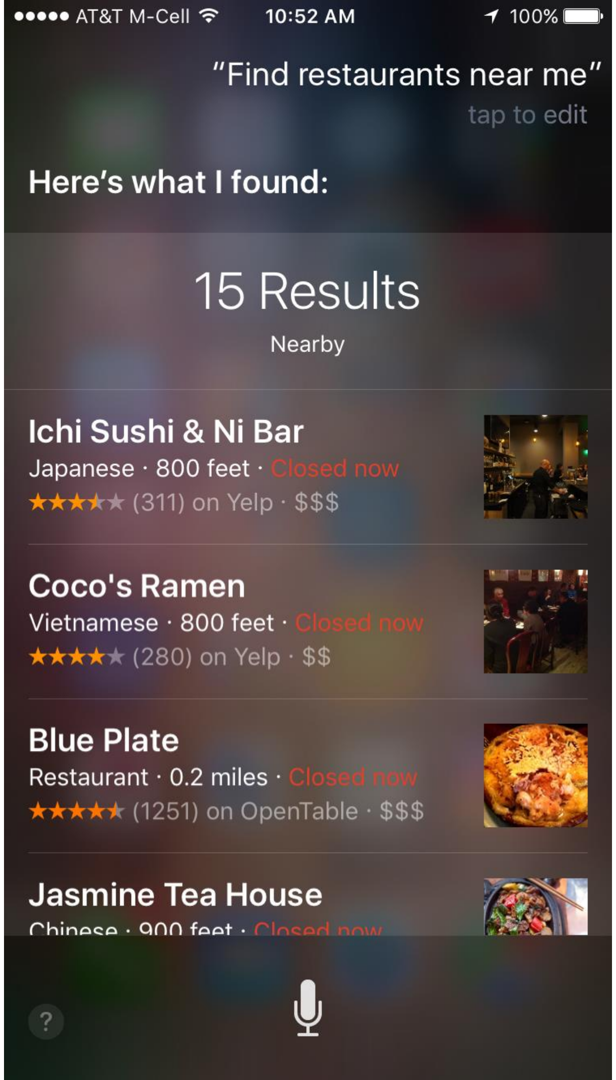
Jasmine Tea House
Chinese · 900 feet · Closed now



"Tell me more about the first one"

Ichi Sushi & Ni Bar is a Japanese restaurant on Mission St 800 ft to your west. It averages 3½ stars and is somewhat pricey.

SIRI
around
2017



"Find restaurants near me"

Here's what I found:

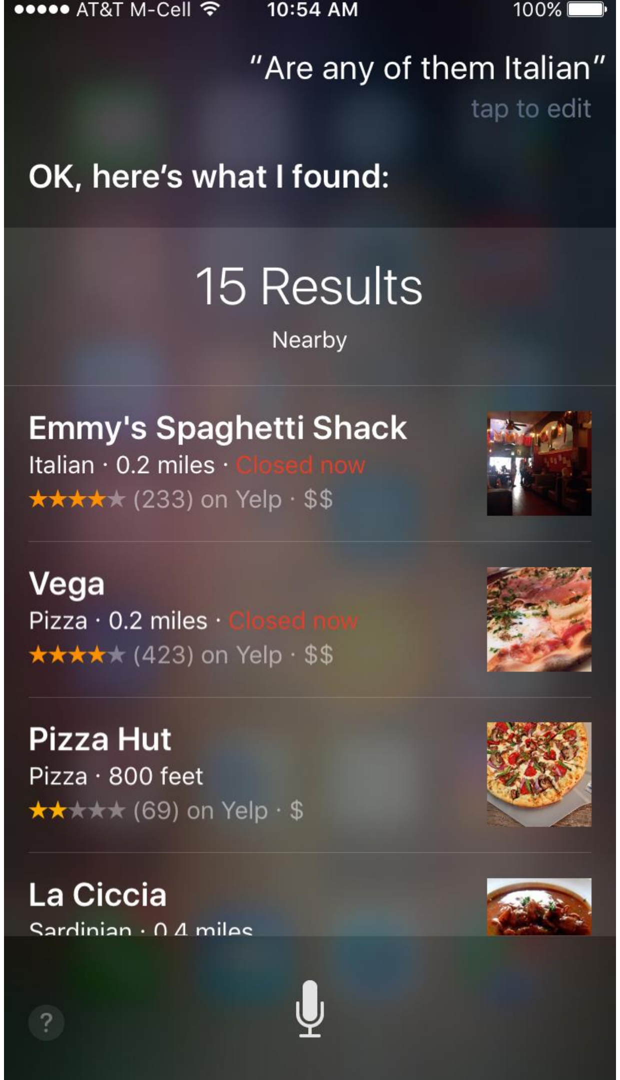
15 Results
Nearby

Ichi Sushi & Ni Bar
Japanese · 800 feet · Closed now
★★★½☆ (311) on Yelp · $$$

Coco's Ramen
Vietnamese · 800 feet · Closed now
★★★★☆ (280) on Yelp · $$

Blue Plate
Restaurant · 0.2 miles · Closed now
★★★★½ (1251) on OpenTable · $$$

Jasmine Tea House
Chinese · 900 feet · Closed now



"Are any of them Italian"

OK, here's what I found:

15 Results
Nearby

Emmy's Spaghetti Shack
Italian · 0.2 miles · Closed now
★★★★☆ (233) on Yelp · $$

Vega
Pizza · 0.2 miles · Closed now
★★★★☆ (423) on Yelp · $$

Pizza Hut
Pizza · 800 feet
★★☆☆☆ (69) on Yelp · $

La Ciccia
Sardinian · 0.4 miles

Two classes

Chatbots

- Talking for fun, therapy, …
- Mimic informal human conversation
- Keep user engaged (*more turns*)
- ELIZA, Parry, Microsoft Tay
- Amazon Alexa challenge (20 min.)

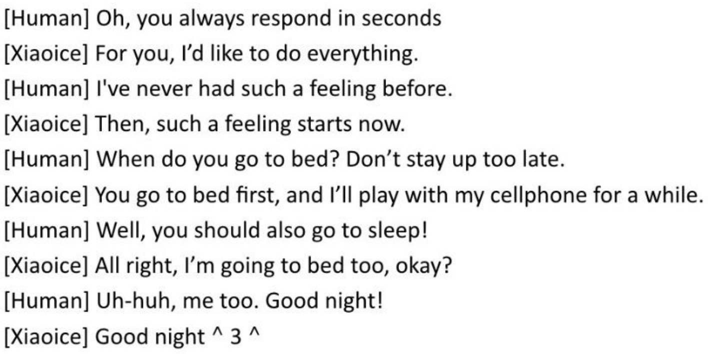
(Goal-based) Dialog agents

- To help achieve a goal: Booking flights, finding a restaurant, getting a haircut, …
- Achieve goal as fast as possible (*few turns*)
- Examples: GUS, SIRI, Google assistant

Chatbots

Example: XiaoIce

- By Microsoft, released in 2014.
- (One of their?) success metrics: number of turns
- “The XiaoIce persona is designed as a 18-year-old girl who is always reliable, sympathetic, affectionate, and has a wonderful sense of humor.”



[Human] Oh, you always respond in seconds
[Xiaoice] For you, I’d like to do everything.
[Human] I've never had such a feeling before.
[Xiaoice] Then, such a feeling starts now.
[Human] When do you go to bed? Don’t stay up too late.
[Xiaoice] You go to bed first, and I’ll play with my cellphone for a while.
[Human] Well, you should also go to sleep!
[Xiaoice] All right, I’m going to bed too, okay?
[Human] Uh-huh, me too. Good night!
[Xiaoice] Good night ^ 3 ^

Zhou et al. 2018, The Design and Implementation of XiaoIce,an Empathetic Social Chatbot

https://news.microsoft.com/apac/features/much-more-than-a-chatbot-chinas-xiaoice-mixes-ai-with-emotions-and-wins-over-millions-of-fans/

Chatbots Architectures

- **Rule-based**
  - Pattern-action rules (ELIZA) + a mental model (Parry)
- **Corpus-based (using a large chat corpus)**
  - Information Retrieval (IR) based. Select a response from a large corpus.
  - Neural networks. Train a neural network to generate text.

Chatbots Architectures

- **Rule-based**
  - Pattern-action rules (ELIZA) + a mental model (Parry)

- **Corpus-based (using a large chat corpus)**
  - Information Retrieval (IR) based. Select a response from a large corpus.
  - Neural networks. Train a neural network to generate text.

ELIZA: Weizenbaum (1966)

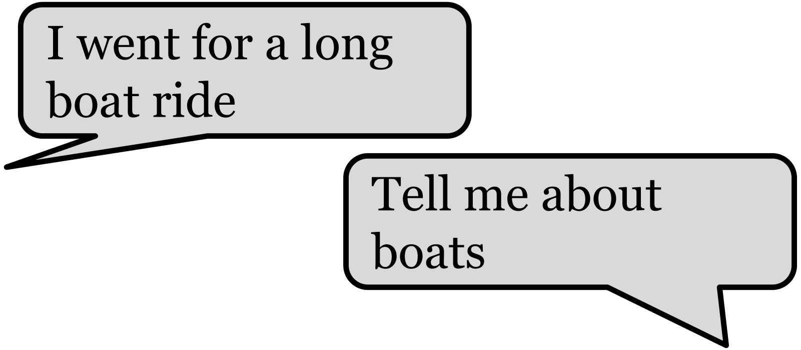
```
Men are all alike.
IN WHAT WAY
They're always bugging us about something or other.
CAN YOU THINK OF A SPECIFIC EXAMPLE
Well, my boyfriend made me come here.
YOUR BOYFRIEND MADE YOU COME HERE
He says I'm depressed much of the time.
I AM SORRY TO HEAR YOU ARE DEPRESSED
...
WHO ELSE IN YOUR FAMILY TAKES CARE OF YOU?
My father
YOUR FATHER
You are like my father in some ways.
WHAT RESEMBLANCE DO YOU SEE
You are not very aggressive but I think you don’t want me to notice that.
WHAT MAKES YOU THINK I AM NOT AGGRESSIVE
```

How does ELIZA work?

How does ELIZA work? (1)

Be a Rogerian psychologist!

Draw the patient out by reflecting patient’s statements back at them



Example of a stateless reactive agent:

- Don’t care about the history
- Act based on the current state now

How does ELIZA work? (2)

- ELIZA looks for **keywords**
  - Keywords have a rank (how important they are)

- Keywords have associated **pattern/transform rules**
  - Pattern: sentence of the user
  - Transform: response to user based on pattern

- ELIZA has **standard answers** and a **clever memory trick**

ELIZA pattern/transform rules

`(0 YOU 0 ME)` [*pattern*]

→

`(WHAT MAKES YOU THINK I 3 YOU)` [*transform*]

0 means Kleene *

The 3 refers to the third component in the pattern.

ELIZA pattern/transform rules

`(0 YOU 0 ME)` [*pattern*]

→

`(WHAT MAKES YOU THINK I 3 YOU)` [*transform*]

0 means Kleene *

The 3 refers to the third component in the pattern.

```
You hate me
WHAT MAKES YOU THINK I HATE YOU
```

Keywords are ranked from specific to general

`I know everybody laughed at me`

Rule: `I: (I *) -> (You say you 2)`

Producing: `YOU SAY YOU KNOW EVERYBODY LAUGHED AT YOU`

Keywords are ranked from specific to general

`I know everybody laughed at me`

Rule: `I: (I *) -> (You say you 2)`

Producing: `YOU SAY YOU KNOW EVERYBODY LAUGHED AT YOU`

But… “I” is a very general keyword:
“Everybody” may be more interesting (everybody/always is probably “referring to some quite specific event or person”)

Keywords are ranked from specific to general

`I know everybody laughed at me`

Rule: `I: (I *) -> (You say you 2)`

Producing: `YOU SAY YOU KNOW EVERYBODY LAUGHED AT YOU`

But… "I" is a very general keyword:
"Everybody" may be more interesting (everybody/always is probably "referring to some quite specific event or person")

Better response? `WHO IN PARTICULAR ARE YOU THINKING OF?`

Keywords are ranked from specific to general

`I know everybody laughed at me`

Rule: `I: (I *) -> (You say you 2)`

Producing: `YOU SAY YOU KNOW EVERYBODY LAUGHED AT YOU`

But… “I” is a very general keyword:
“Everybody” may be more interesting (everybody/always is probably “referring to some quite specific event or person”)

Better response? `WHO IN PARTICULAR ARE YOU THINKING OF?`

- Implement this assigning rankings to keywords (everybody=5, I=0, etc..)

When no keyword matches

- Select a transform for the NONE keyword
  - PLEASE GO ON
  - THAT'S VERY INTERESTING
  - I SEE

- Top response from the memory stack (memory trick!)
  - `0 MY 0 = LETS DISCUSS FURTHER WHY YOUR 3`
  - Whenever "my" is the highest ranked keyword: Apply transform, store in on the stack
  - When no keyword matches: return top response from memory stack.

Other ELIZA stuff

- Rules can refer to **classes of words**

  Family = mother, father, brother, sister; `NOUN = …`

- **Don’t reuse** transforms in the same conversation
  - Whenever we use a transform associated with a pattern
  - We increment a counter for that rule
  - So the next time we use the next ranked transform

- Some **basic transforms** during input processing
  - `I -> YOU`
  - `YOU -> I`

PARRY

- Simulate a patient with person with paranoid schizophrenia
- Colby et al. 1971
- Modeled mental state (e.g. anger)

When PARRY met ELIZA …

```
ELIZA: Good Evening. Tell me your problems.
Please terminate input with a period or a
question mark.
PARRY: People get on my nerves sometimes.
ELIZA: I am not sure I understand you fully.
PARRY: You should pay more attention.
ELIZA: Suppose you should pay more attention.
PARRY: You're entitled to your own opinion.
ELIZA: What makes you think I am entitled to my
own opinion?
PARRY: It bothers me just to be around people in
general.
ELIZA: Please go on.
```

`https://tools.ietf.org/html/rfc439`

Chatbots Architectures

- **Rule-based**
    - Pattern-action rules (ELIZA) + a mental model (Parry)

- **Corpus-based (using a large chat corpus)**
    - Information Retrieval (IR) based. Select a response from a large corpus.
    - Neural networks. Train a neural network to generate text.

Corpus-based chatbots

Mine conversations from corpora

- Switchboard corpus. Telephone conversations. Approx. 2,500 dialogs.
- Cornell Movie-Dialogs Corpus: 617 movies, 304,713 utterances.
- Ubuntu Dialogue Corpus: 1 million dialogs. 7 million utterances. Extracted from Ubuntu-related chatrooms on IRC.
- See Serban et al. 2018 for an overview
- https://parl.ai contains a large collection of dialog datasets (and models).

Chatbots Architectures

- **Rule-based**
  - Pattern-action rules (ELIZA) + a mental model (Parry)
- **Corpus-based (using a large chat corpus)**
  - Information Retrieval (IR) based. **Select** a response from a large corpus.
  - Neural networks. Train a neural network to generate text.

User: Hi.

Cleverbot: How are you doing today.

User: Good, how are you?

Cleverbot: Good thanks for asking.

User: What do you think of artificial intelligence?

Cleverbot: It's great.

User: Why do you think it's great?

Cleverbot: idk

https://www.cleverbot.com/

IR-based chatbot architectures

Return the response to the most similar turn (response selection)

Corpus:

Turn	Response
What is your favorite movie?	*Titanic*
Do you like Tom Hanks?	*Yes, he’s great*
You like Doctor Who?	*Yes, so funny*
How are you?	*Good, how are you?*
I’m sad	*Why are you sad?*
...	...

New user turn:

Do you like Doctor Who?”

System response:

Yes, so funny

IR-based chatbot architectures

Return the response to the most similar turn (response selection)

- Take user's turn **q** and find a similar turn **t** in the corpus **C**

  q = "*do you like Doctor Who?*"
  t= "*you like Doctor Who?*"

- Return the response to **t** in your corpus.

$$r = response\ (\underset{t \in C}{argmax}\ \text{similarity}(\mathbf{q}, \mathbf{t}))$$

Yes, so funny

Return the most similar turn

IR-based chatbot architectures

Return the response to the most similar turn
(response se

- Take user's

q
$t=$

Stay tuned! Lecture 4 will discuss how to measure similarity between data instances!

- Return the response to **t** in your corpus.

$$r = response\ (\underset{t \in C}{argmax}\ \text{similarity}(q, t))$$

`Yes, so funny`

Return the most similar turn

Chatbots Architectures

- **Rule-based**
  - Pattern-action rules (ELIZA) + a mental model (Parry)
- **Corpus-based (using a large chat corpus)**
  - Information Retrieval (IR) based. Select a response from a large corpus.
  - Neural networks. Train a neural network to generate text.

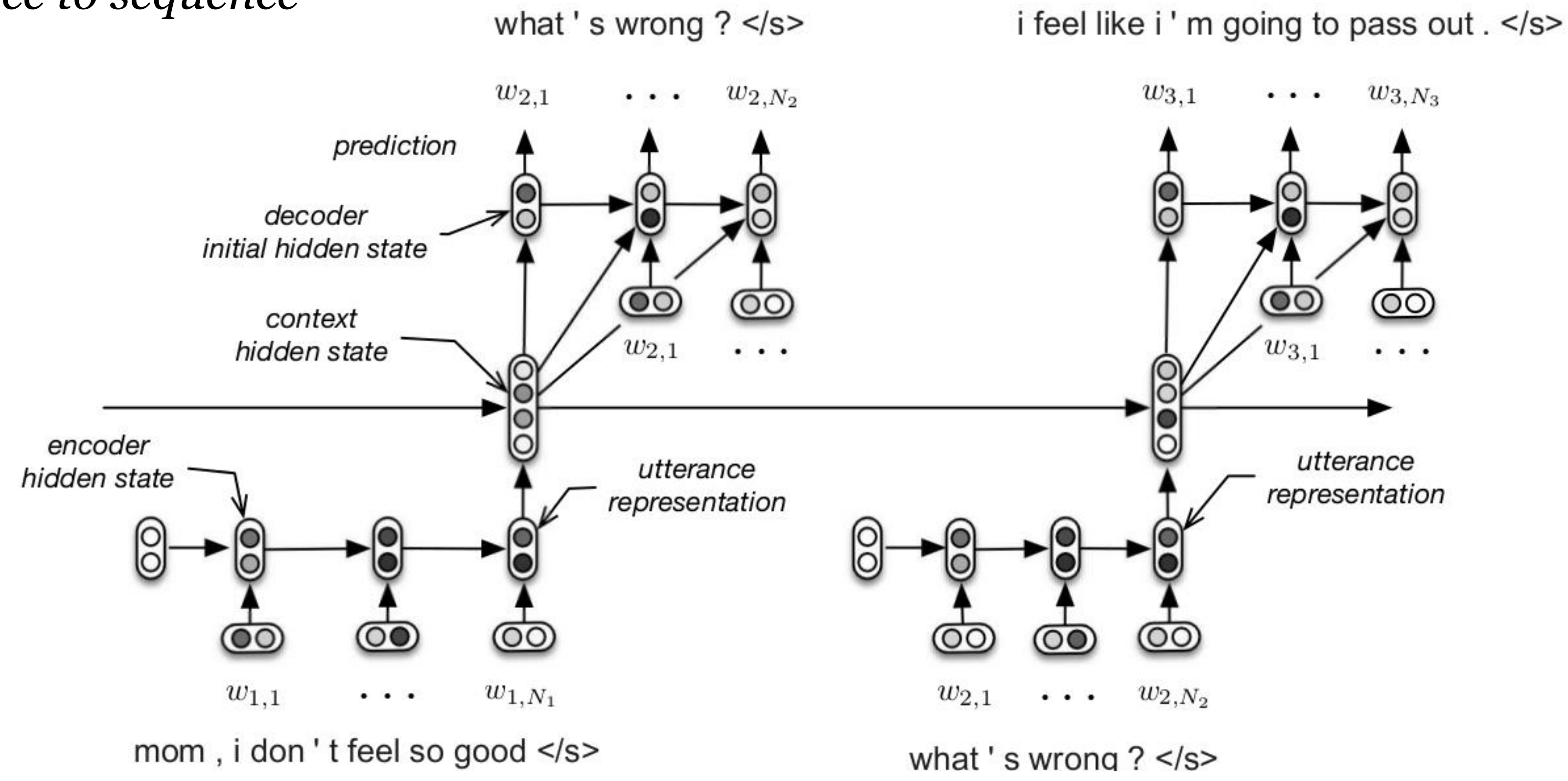
Neural network approaches

- Earlier approaches framed it as a translation problem (utterance to response)
- But… not semantically equivalent

- Map user$_1$ *turn* to user$_2$ *response*

Seq2seq architecture

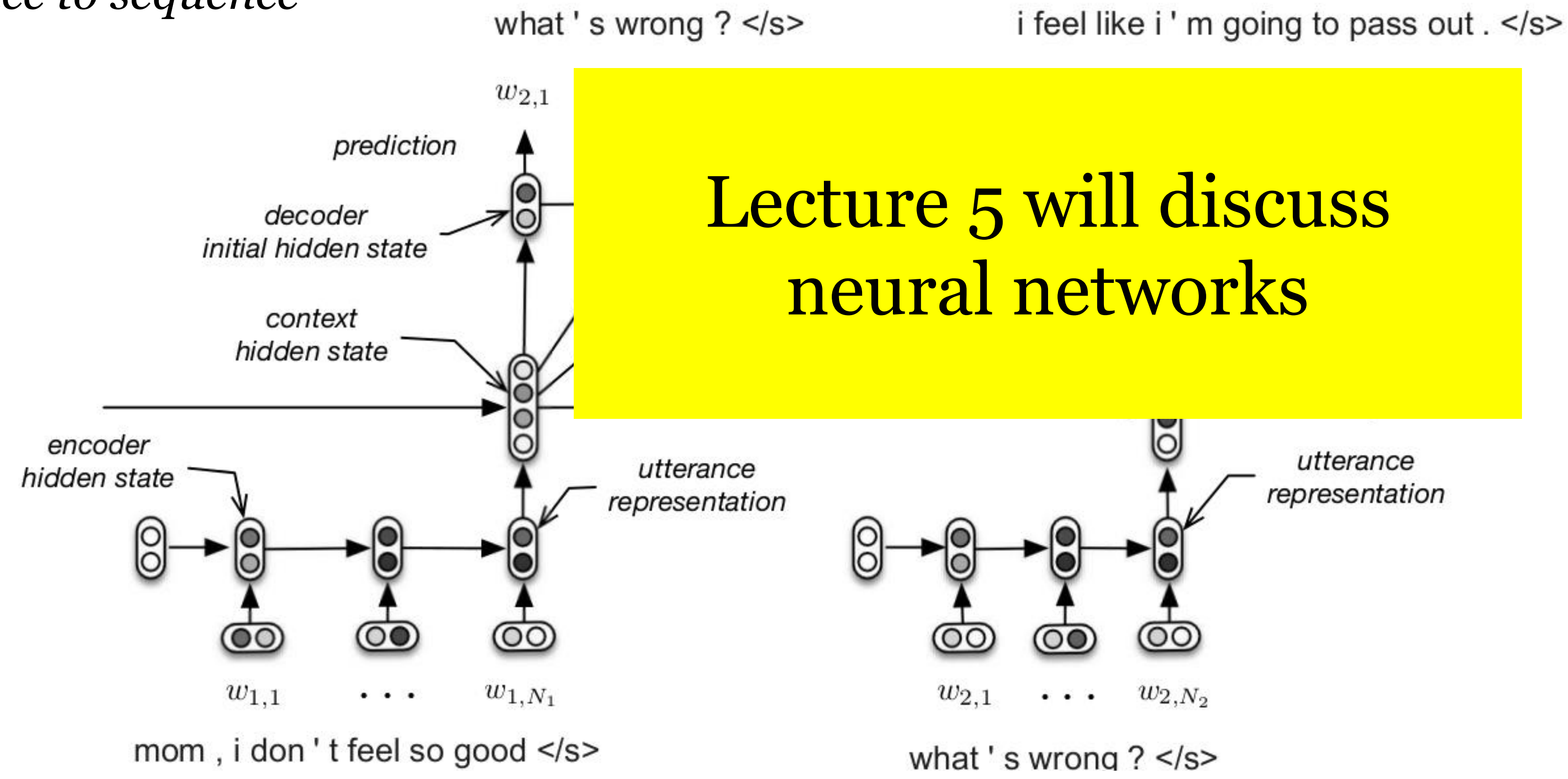
“sequence to sequence”


what ' s wrong ? </s>
i feel like i ' m going to pass out . </s>
$w_{2,1}$ . . . w_{2,N_2}
$w_{3,1}$ . . . w_{3,N_3}
prediction
decoder initial hidden state
context hidden state
$w_{2,1}$ . . .
$w_{3,1}$ . . .
encoder hidden state
utterance representation
utterance representation
$w_{1,1}$ . . . w_{1,N_1}
$w_{2,1}$ . . . w_{2,N_2}
mom , i don ' t feel so good </s>
what ' s wrong ? </s>

Serban, Iulian V., Alessandro Sordoni, Yoshua Bengio, Aaron Courville, and Joelle Pineau. 2015. "Building End-To-End Dialogue Systems Using Generative Hierarchical Neural Network Models.

Seq2seq architecture

“sequence to sequence”

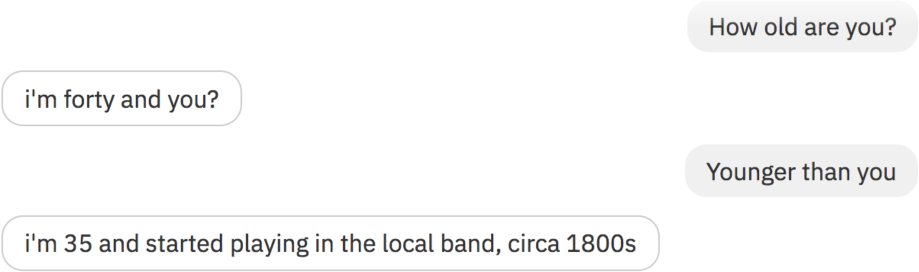


Lecture 5 will discuss neural networks

Serban, Iulian V., Alessandro Sordoni, Yoshua Bengio, Aaron Courville, and Joelle Pineau. 2015. "Building End-To-End Dialogue Systems Using Generative Hierarchical Neural Network Models.

Neural network approaches: challenges

- Staying **coherent**



How old are you?

i'm forty and you?

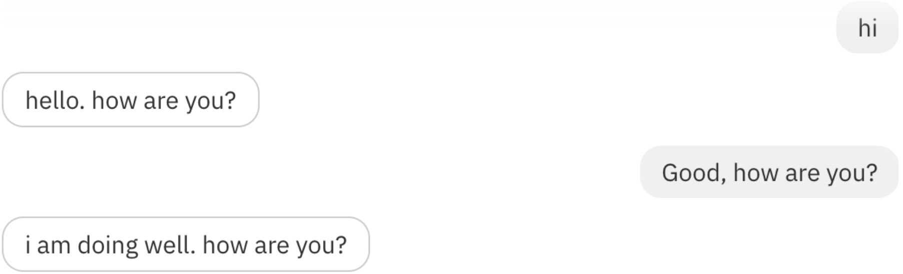
Younger than you

i'm 35 and started playing in the local band, circa 1800s

https://convai.huggingface.co/

Neural network approaches: challenges

- Staying coherent
- Many neural chatbots are ***boring***!! (short, common sentences have high probabilities)



hi

hello. how are you?

Good, how are you?

i am doing well. how are you?

https://convai.huggingface.co/

Comparison of approaches

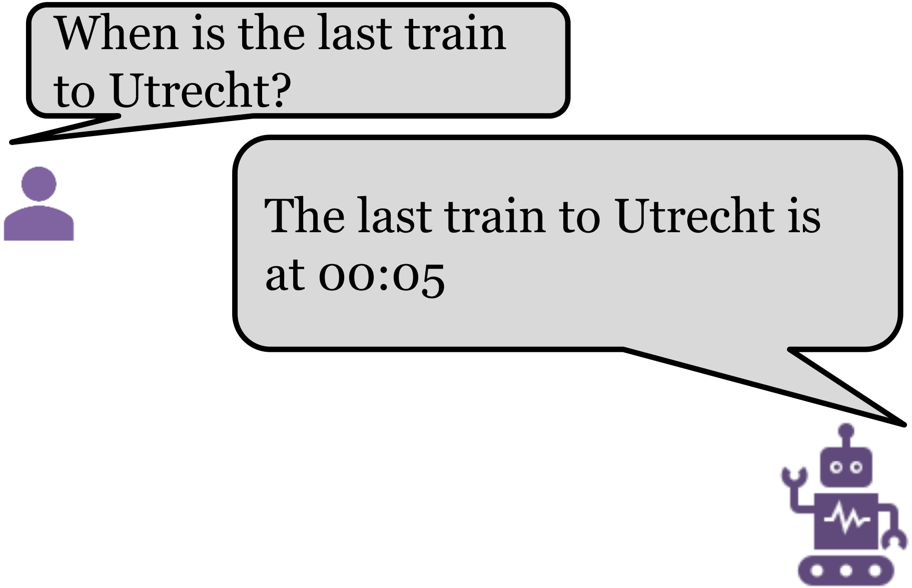
- **Rule-based:**
  - Pro: Good for narrow, scriptable applications
  - Con: Expensive and brittle
- **Information Retrieval:**
  - Pro: Easy to implement
  - Con: If a sentence is not in the training corpus, the chatbot will not say it (can only mirror training data)
- **Neural Networks:**
  - Pro: Output sequence can be any combination of words from the training corpus
  - Con: Fluency, problems staying coherent, often boring.

Goal-based dialog agents

Goal-based dialogue systems

- Dialogue systems that help the user to **solve a specific task**
  - SIRI, Google assistant

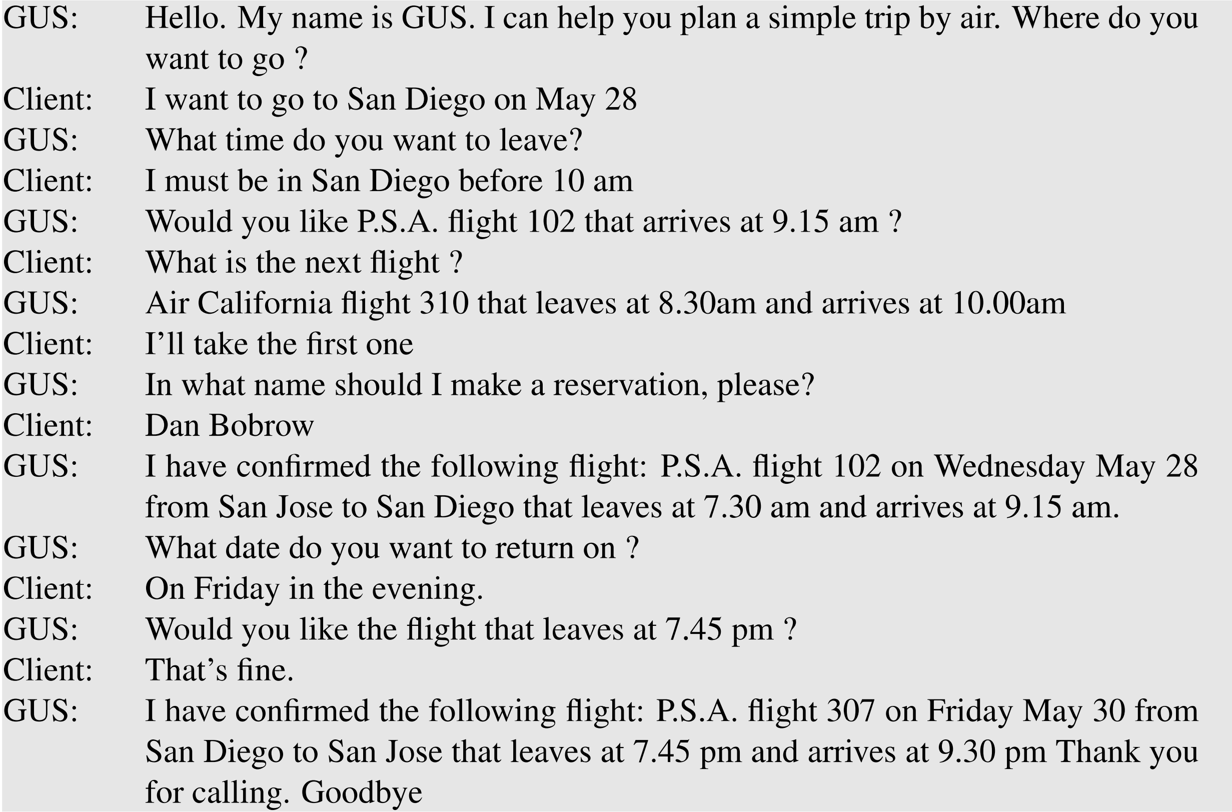
- In it’s simplest form just **a number of forms** stuck together



When is the last train to Utrecht?

The last train to Utrecht is at 00:05

Travel dialogue: GUS (Bobrow et al. 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

Frames

Tasks can often be represented as frames. A frame consists of a set of **slots**, to be filled with a **value** of a given **type**

Each slot is associated with a **question** to the user

```
FLIGHT FRAME:

ORIGIN:                Type:     Question:
    CITY: Boston       city      What city are you leaving from?
    DATE: Tuesday      date      What day would you like to leave?
    TIME: morning      time      What time would you like to leave?

DEST:
    CITY: San Francisco
    AIRLINE:
…
```

Frames

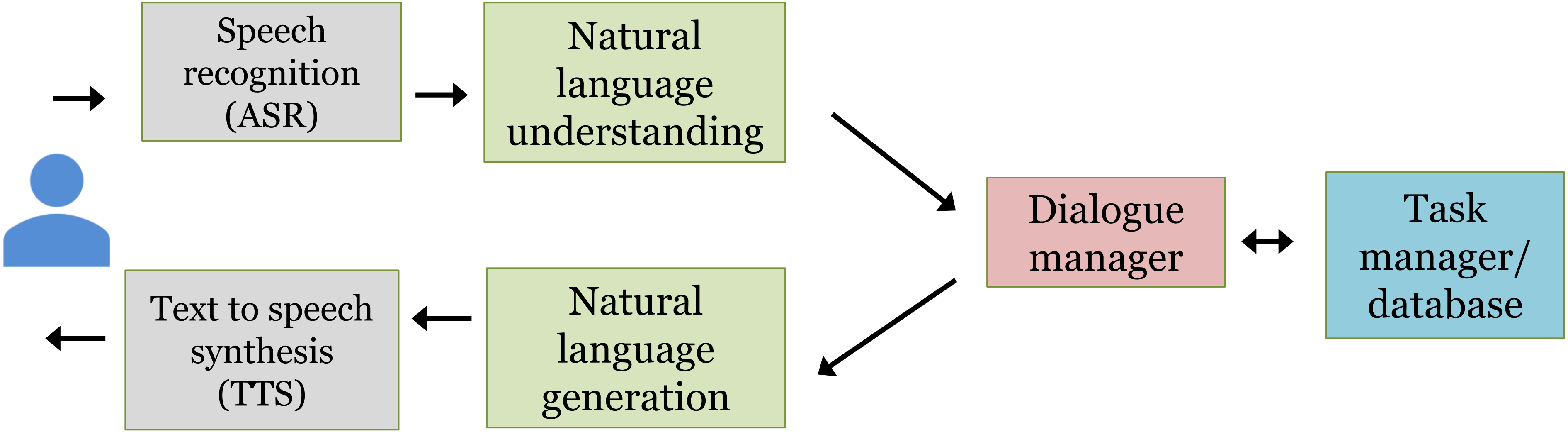
Tasks can often be represented as frames. A frame consists of a set of **slots**, to be filled with a **value** of a given **type**
Each slot is associated with a **question** to the user

```
FLIGHT FRAME:

ORIGIN:
    CITY: Boston
    DATE: Tuesday
    TIME: morning

DEST:
    CITY: San Francisco
    AIRLINE:
…
```

Slot types can be complex!

DATE
    MONTH NAME
    DAY (BOUNDED-INTEGER 1 31)
    YEAR INTEGER
    WEEKDAY (MEMBER (SUNDAY MONDAY TUESDAY WEDNESDAY THURSDAY FRIDAY SATURDAY)]

Typical Frame-based Dialogue System Architecture



Speech recognition (ASR)

Natural language understanding

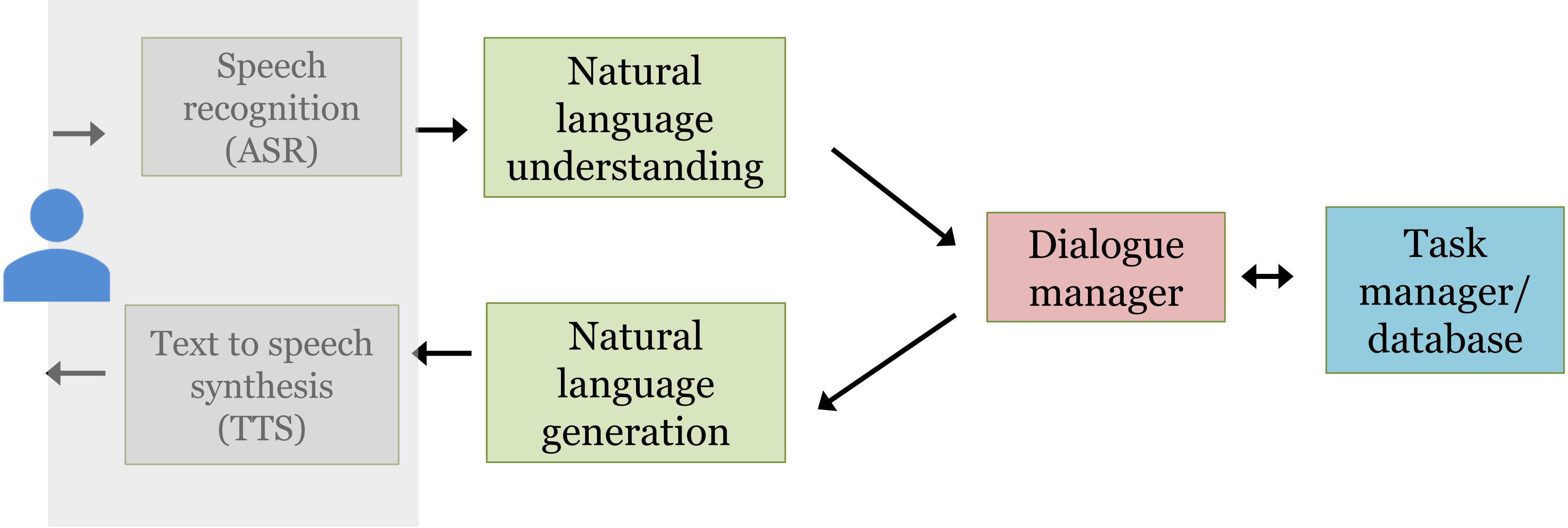
Dialogue manager

Task manager/ database

Natural language generation

Text to speech synthesis (TTS)

Typical Frame-based Dialogue System Architecture



Speech recognition (ASR)

Natural language understanding

Dialogue manager

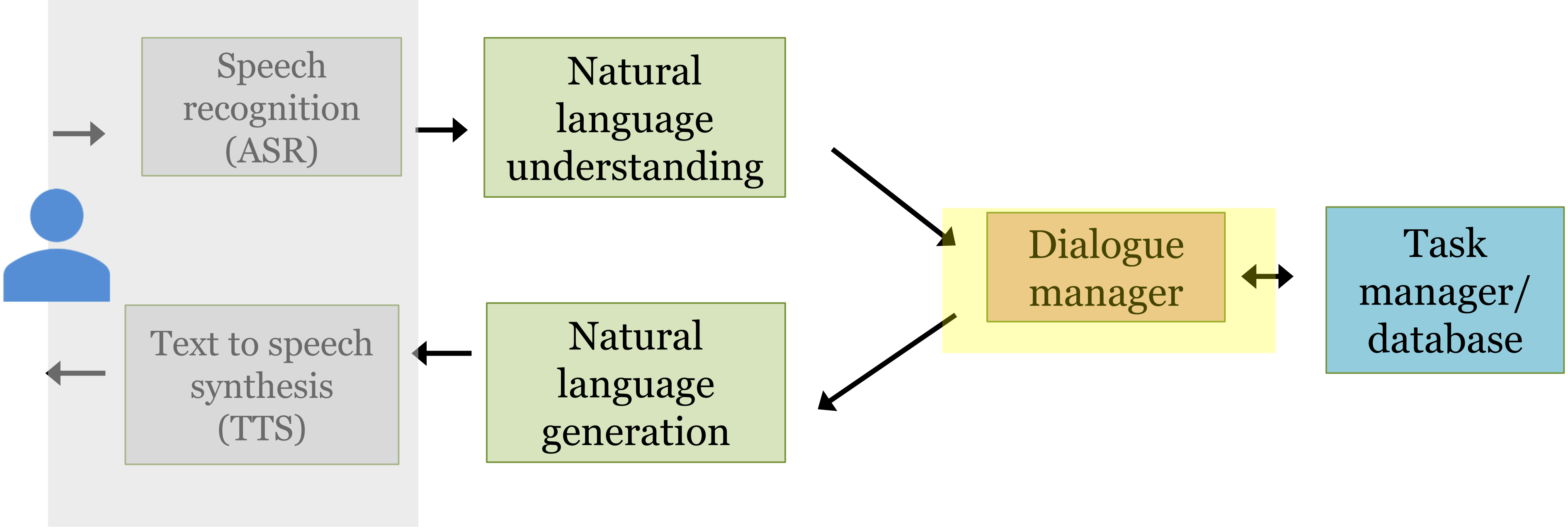
Task manager/ database

Natural language generation

Text to speech synthesis (TTS)

Not discussed here.
This course: text input

Typical Frame-based Dialogue System Architecture



Speech recognition (ASR)

Natural language understanding

Dialogue manager

Task manager/ database

Text to speech synthesis (TTS)

Natural language generation

Not discussed here.
This course: text input

Controls the state and the flow of the dialog

Dialogue Initiative

- **Initiative**: who has control of conversation?

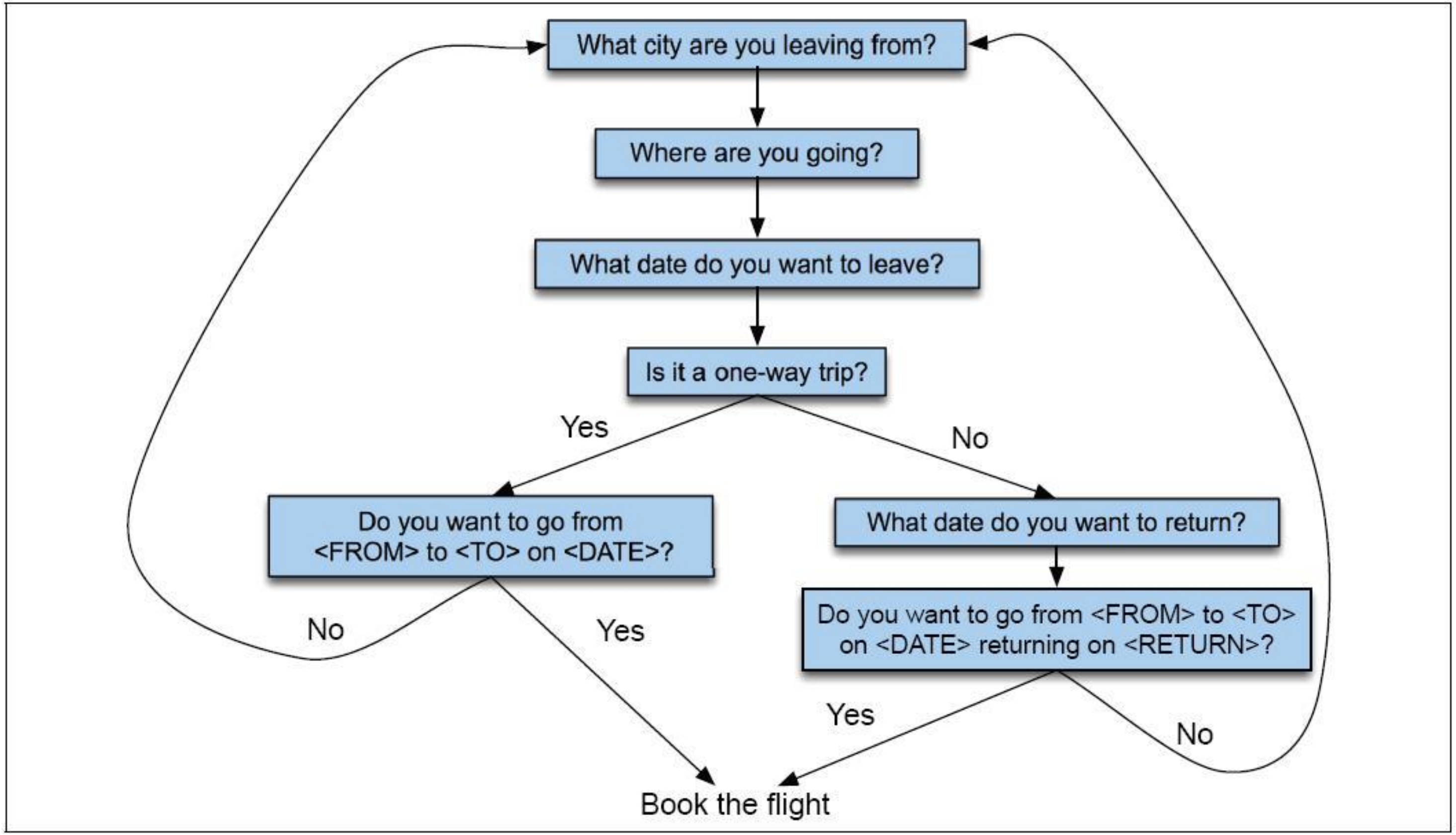
- System initiative, user initiative, mixed initiative

System initiative

- System completely controls the conversation with the user.
  - System asks questions. The user answers.

- Example: Booking a flight
  - Ask the user for a departure city
  - Ask for a destination city
  - Ask for a time
  - Ask whether the trip is round-trip or not

Finite State Dialog Manager


What city are you leaving from?
Where are you going?
What date do you want to leave?
Is it a one-way trip?
Yes
No
Do you want to go from <FROM> to <TO> on <DATE>?
What date do you want to return?
Do you want to go from <FROM> to <TO> on <DATE> returning on <RETURN>?
No
Yes
Yes
No
Book the flight

States: questions to the user

Transitions: user responses

System controls the interaction.

System Initiative

- Simple to build
- User always knows what they can say next
- System always knows what user can say next
    - Known words: Better performance from ASR
    - Known topic: Better performance from NLU
- OK for VERY simple tasks (entering a credit card)

😃

- Too limited

😞

System initiative + universals

We can give users a little more flexibility by adding **universals**: commands you can say anywhere

- As if we augmented every state of FSA with Help ; Start over ; Correct
- Used by many implemented systems
- But still doesn't allow user much flexibility

Still..

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

User Initiative

User directs the system:

The user asks a question, the system answers

- Examples: voice web search
- Users know what systems can do (e.g. question answering)
- System is *reactive*
- But system can't:
  - ask questions back
  - engage in clarification dialogue
  - engage in confirmation dialogue

Mixed Initiative

In normal human-human dialogue, initiative shifts back and forth between participants.

Mixed initiative: Conversational initiative can shift between system and user

But… can get confusing (for both system and the user!)

Frame-based dialog manager

Simplest type of **mixed initiative**: use the structure of a **frame** to guide dialogue

System asks questions of user, filling any slots that user specifies→ When frame is filled, do database query

```
FLIGHT FRAME:

ORIGIN:
    CITY: Boston
    DATE: Tuesday
    TIME: morning

DEST:
    CITY: San Francisco
    AIRLINE:
…
```

Frame-based dialog manager

Simplest type of **mixed initiative**: use the structure of a **frame** to guide dialogue

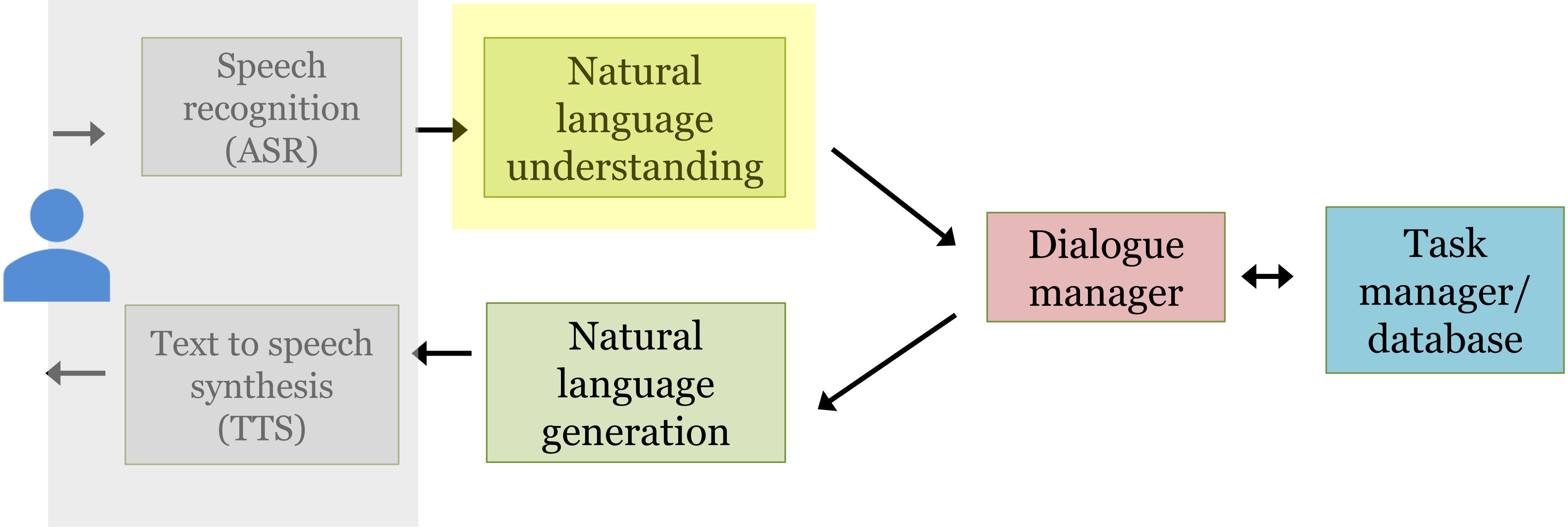
```
FLIGHT FRAME:

ORIGIN:
    CITY: Boston
    DATE: Tuesday
    TIME: morning

DEST:
    CITY: San Francisco
    AIRLINE:
…
```

System asks questions of user, filling any slots that user specifies→ When frame is filled, do database query

User can answer multiple questions at once

- If user answers 3 questions at once, system has to fill slots and not ask these questions again
- Avoids strict constraints on order of the finite state architecture.

Typical Frame-based Dialogue System Architecture



Speech recognition (ASR)

Natural language understanding

Dialogue manager

Task manager/ database

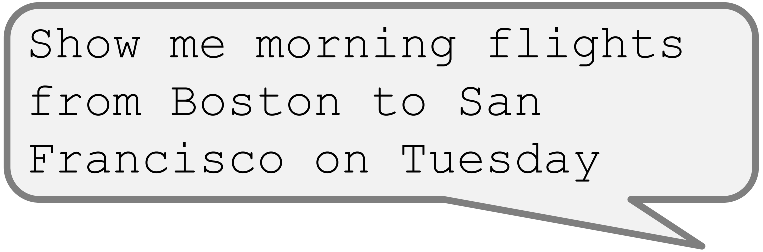
Natural language generation

Text to speech synthesis (TTS)

Not discussed here.
This course: text input

Filling slots: Tasks

1. **Domain classification**
   Asking about the weather? Booking a flight? Programming alarm clock?
2. **Intent Determination**
   Find a Movie, Show Flight, Remove Calendar Appointment
3. **Slot Filling**
   Extract the actual slots and fillers



```
Show me morning flights
from Boston to San
Francisco on Tuesday
```

```
DOMAIN: AIR-TRAVEL

INTENT: SHOW-FLIGHTS

ORIGIN-CITY: Boston
ORIGIN-DATE: Tuesday
ORIGIN-TIME: morning
DEST-CITY: San Francisco
```

Filling slots: How?

- **Rule based:**

  To recognize SET-ALARM intent:

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

  Very precise, but expensive, slow to create, hard to scale.

- **Machine learning:**

  Given a set of labelled instances, learn a classifier that automatically maps utterances to intents (or domains, or slots, etc.) based on characteristics of the utterance (e.g. words).

Evaluation

Slot Error Rate (SER) for a sentence:

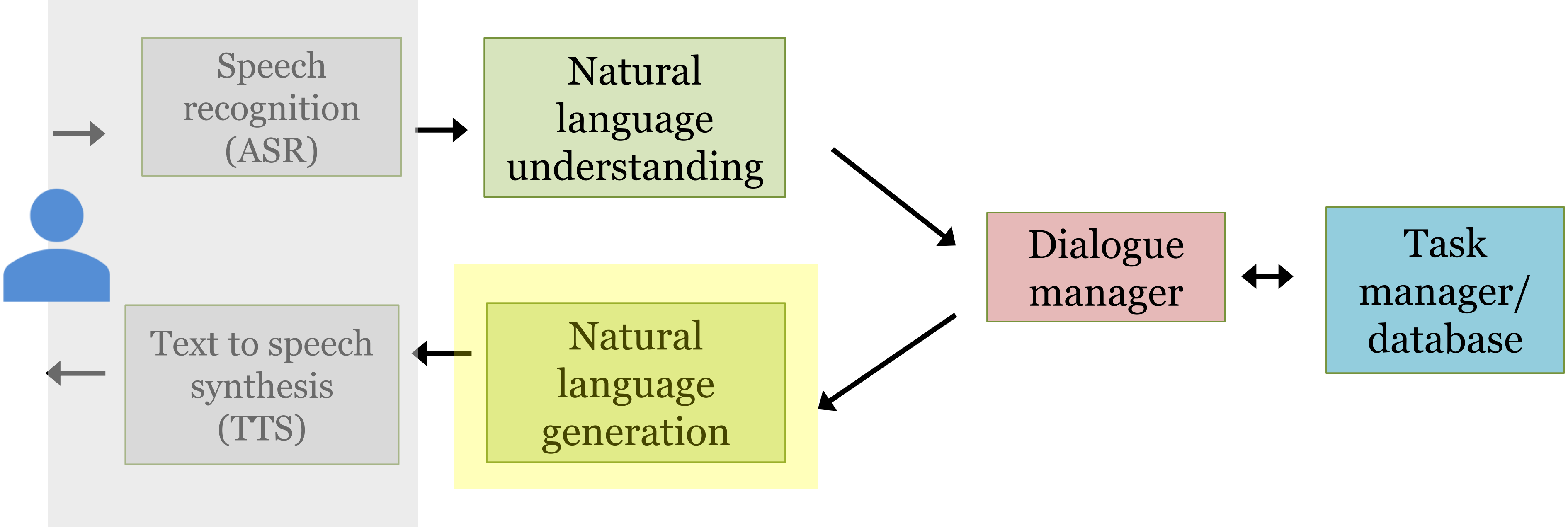
$$\frac{\texttt{Total number of slot errors}}{\texttt{Total number of reference slots for sentence}}$$

```
Make an appointment with Chris
at 11.30am in BBL 523
```

Slot	Filler
PERSON	Dragan
TIME	11:30 a.m.
ROOM	BBL 523

SER = 1/3

Typical Frame-based Dialogue System Architecture


Speech recognition (ASR)
Natural language understanding
Dialogue manager
Task manager/ database
Text to speech synthesis (TTS)
Natural language generation

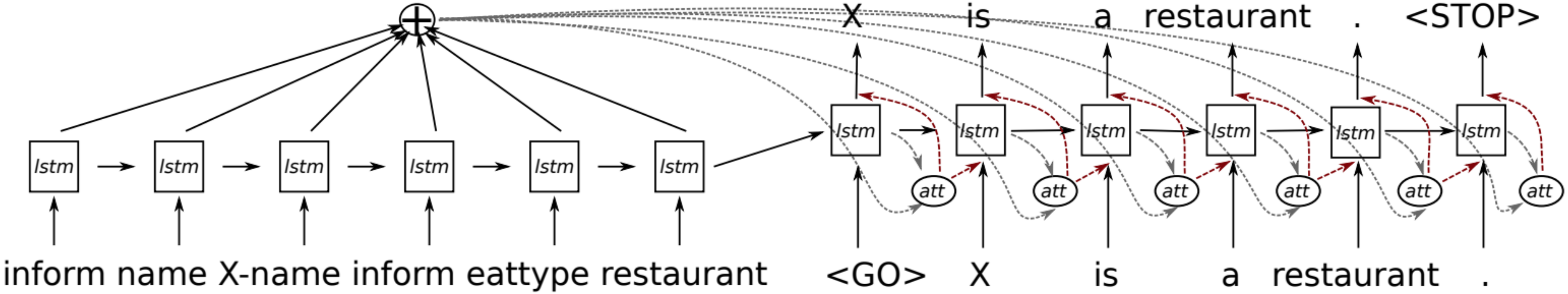
Not discussed here.
This course: text input

Generation natural language from meaning representations

```
Inform (
    name=X,
    type=restaurant,
    price=cheap,
    location=chinatown
```

Template filling

→ X is a cheap restaurant in chinatown



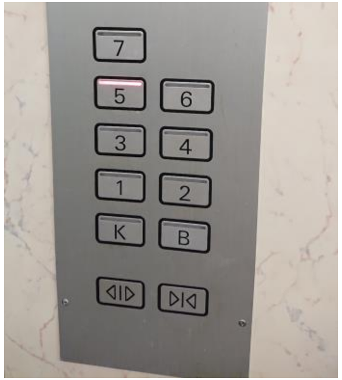
Neural network

Sequence-to-Sequence Generation for Spoken Dialogue via Deep Syntax Trees and Strings, Dušek and Jurčíček 2016

For more than just form-filling, we need to:

- Decide when the user has asked a question, made a proposal, rejected a suggestion
- Ground a user's utterance (i.e. show you understand them), ask clarification questions, suggest plans

Grounding



Grounding

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

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

Grounding

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

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

Dialog systems need to do even more grounding and confirmation than humans!

Explicit vs. implicit confirmation

Explicit confirmation

`System: Which city do you want to leave from?`
`User:` **`Berlin`**
`System: Do you want to leave from` **`Berlin`**`?`

Easier to correct mistakes ☺
Takes more time ☹

Implicit confirmation

`User: I'd like to travel to` **`Berlin`**
`System: When do you want to travel to` **`Berlin`**`?`

More natural, quicker ☺
More difficult to correct ☹
(System needs to be able to handle this)

What do utterances do?

(Historic view) Language
allows us to *describe* the world

But...

What do utterances do?

I hereby bet you 10 euros it won't rain tomorrow

https://www.vectorstock.com/royalty-free-vector/flat-set-vector-6666251

What do utterances do?

Order!

How do speech acts come about?

Convention

- speech acts can be linguistically marked
- large variety of how this happens

Out!

Tennis

I now pronounce you husband and wife!

Priest/registrar

Order!

Speaker of the house

- specific socially recognised situations
- involving agents with specific socially recognised authority

How do speech acts come about?

Convention

- speech acts can be linguistically marked
- large variety of how this happens

The performative formula:
I (hereby) VERB X

I thank you for helping me

I withdraw from the competition

I admit I was wrong

How do speech acts come about?

But

- very often, a speech act is not marked as such
- speech acts can be complex

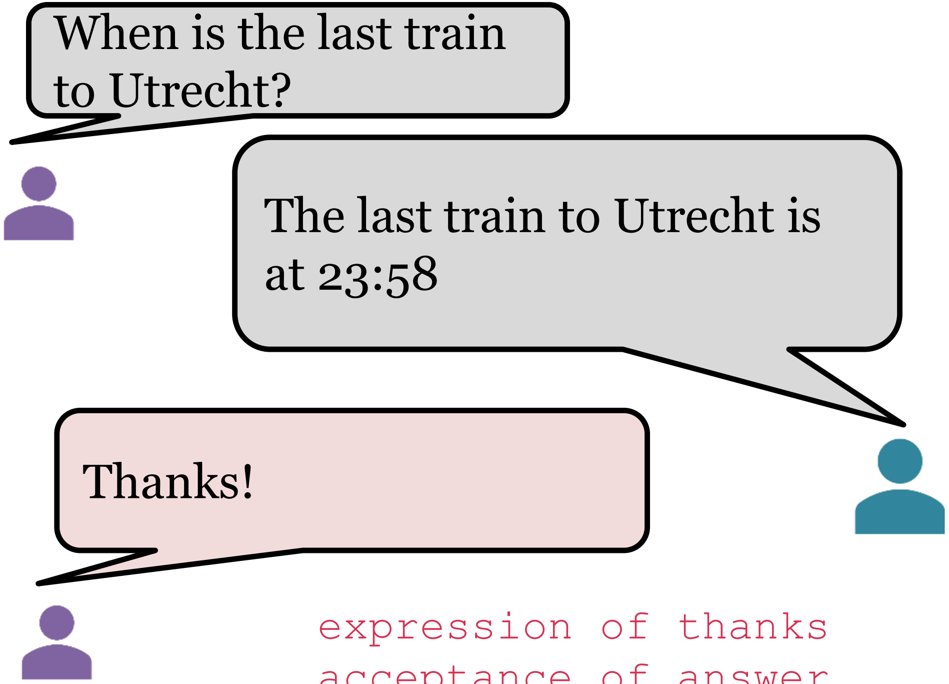
Can you pass the salt?

directive

How do speech acts come about?

But

- very often, a speech act is not marked as such
- speech acts can be **complex**



When is the last train to Utrecht?

The last train to Utrecht is at 23:58

Thanks!

expression of thanks
acceptance of answer
(positive feedback)
dialogue closure

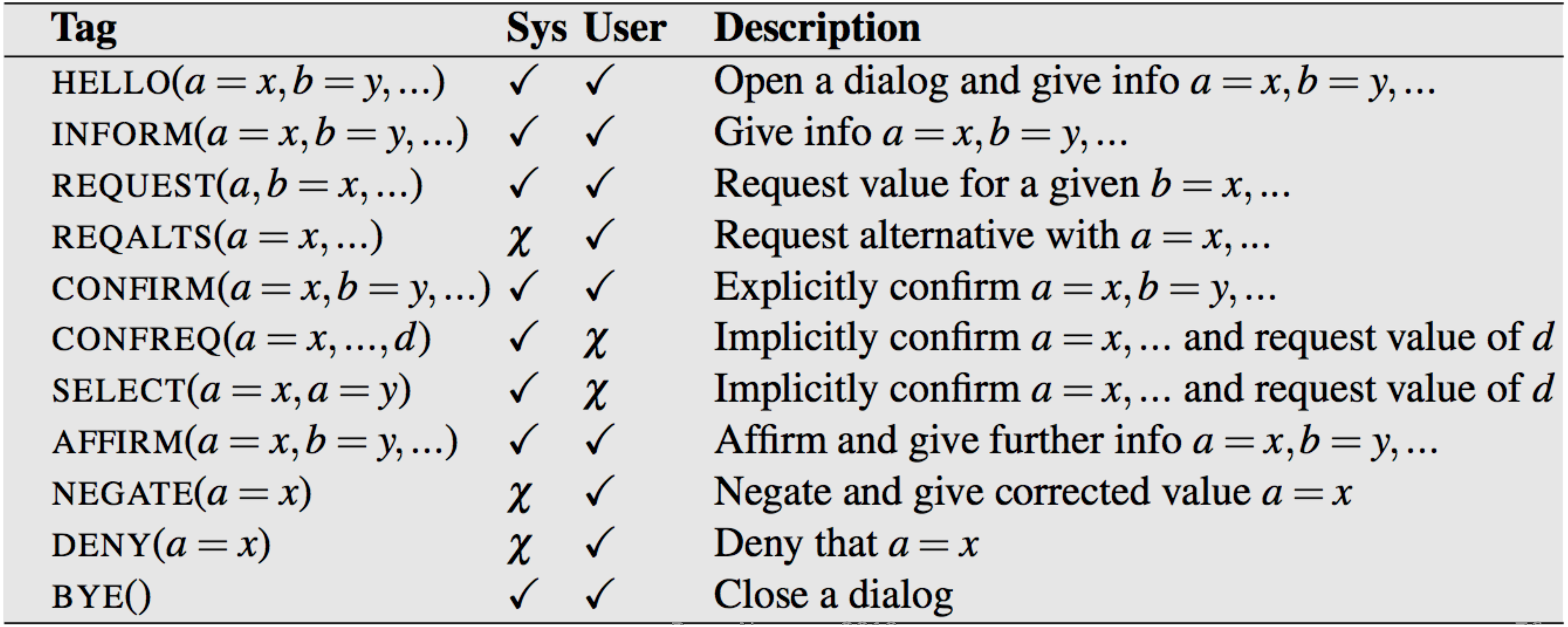
Dialog acts

- Speech act theory focusses on the intentions of the speaker. But a dialogue is not simply a sequence of actions each performed by individual speakers.

- Dialogue requires **coordination** amongst participants.
  - Many actions in dialogue serve to manage the interaction itself (e.g. grounding). These are overlooked by speech act theory.

- Dialog acts extends the notion of speech acts for conversational phenomena.

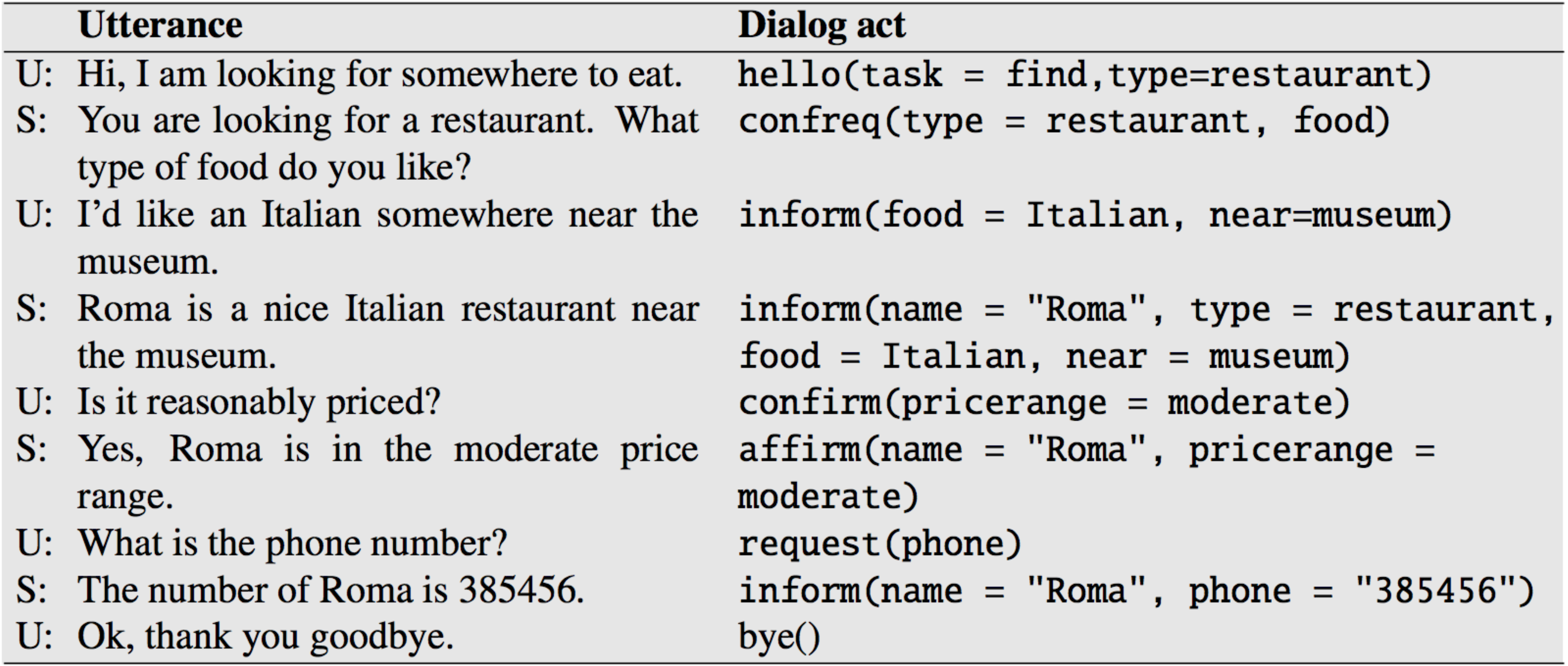
Dialog acts



Tag	Sys	User	Description
HELLO($a = x, b = y, \ldots$)	✓	✓	Open a dialog and give info $a = x, b = y, \ldots$
INFORM($a = x, b = y, \ldots$)	✓	✓	Give info $a = x, b = y, \ldots$
REQUEST($a, b = x, \ldots$)	✓	✓	Request value for a given $b = x, \ldots$
REQALTS($a = x, \ldots$)	χ	✓	Request alternative with $a = x, \ldots$
CONFIRM($a = x, b = y, \ldots$)	✓	✓	Explicitly confirm $a = x, b = y, \ldots$
CONFREQ($a = x, \ldots, d$)	✓	χ	Implicitly confirm $a = x, \ldots$ and request value of d
SELECT($a = x, a = y$)	✓	χ	Implicitly confirm $a = x, \ldots$ and request value of d
AFFIRM($a = x, b = y, \ldots$)	✓	✓	Affirm and give further info $a = x, b = y, \ldots$
NEGATE($a = x$)	χ	✓	Negate and give corrected value $a = x$
DENY($a = x$)	χ	✓	Deny that $a = x$
BYE()	✓	✓	Close a dialog

Fig 26.12 (SLP3)

Dialog acts



	Utterance	Dialog 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()

Fig 26.13 (SLP3) A sample dialog from the HIS System of Young et al. (2010)

Dialog act detection

Can you give me flights from London to Amsterdam?

Question? Or command?

Course project: You'll develop your own dialog act detection system!

Architecture

Similar to earlier architecture, based on filling frames!
But more complex:

- E.g. keep track of whether slots have been filled or confirmed.

Dialog policy: What to do next? E.g. ask for clarification? Make a suggestion? I.e. what dialog act to generate

Confidence	Action
low confidence	reject
above the threshold	confirm explicitly
high confidence	confirm implicitly
very high confidence	don't confirm at all

Evaluation

Evaluation

We need to evaluate dialog systems to:

- Know how we can improve it
- To compare between multiple systems

Coming up with the right evaluation method is research too!

Some metrics:

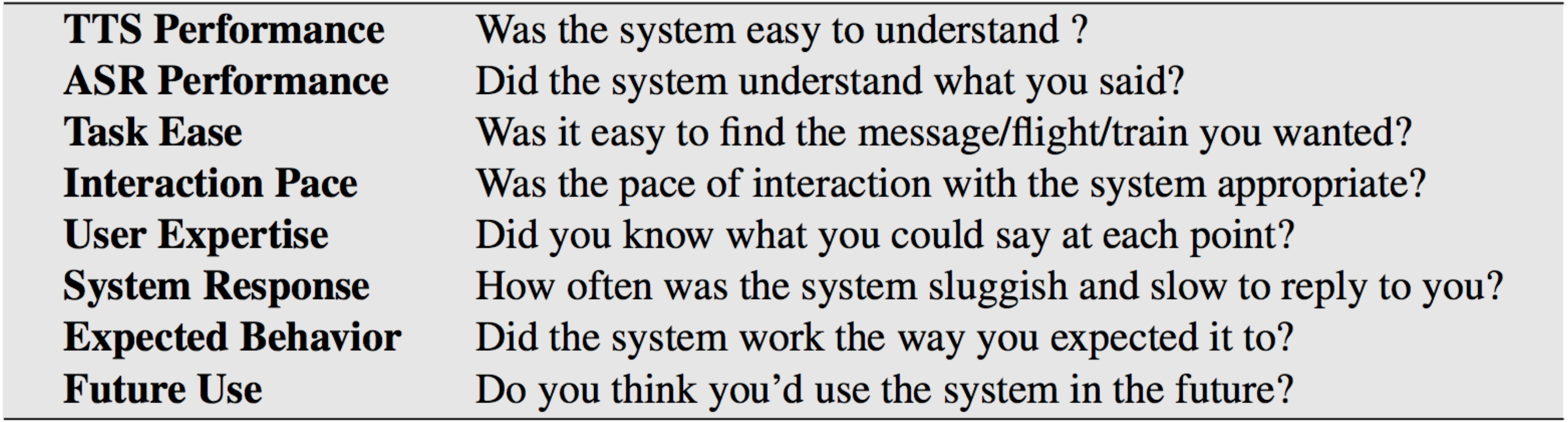
- At the end, was the correct meeting added to the calendar? (task success)
- Total number of turns, total time (efficiency)
- Number of times user had to correct the system, number of system error messages (quality)

Evaluation

We need to evaluate dialog systems to:

- Know how we can improve it
- To compare between multiple systems

Coming up with the right evaluation method is research too!



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?

`Fig 26.18 (SLP3)` User satisfaction survey, adapted from Walker et al. (2001).

Ethical concerns

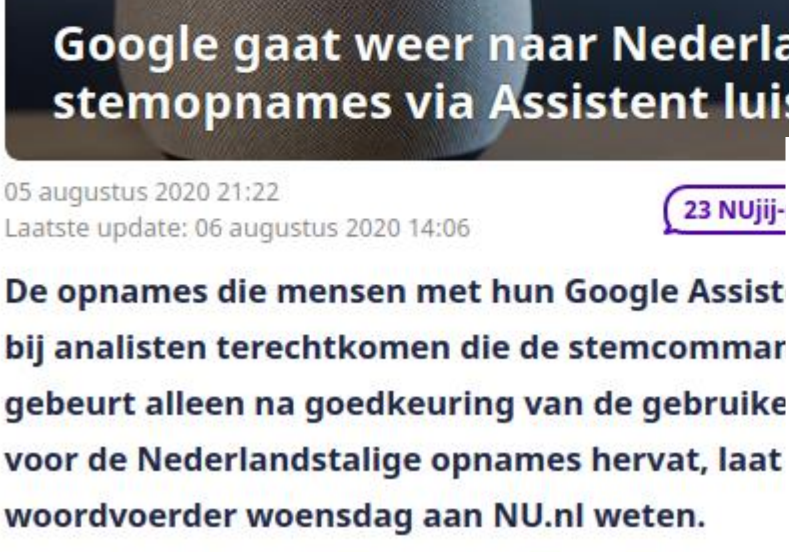
Privacy

- Already noticed in the days of Weizenbaum!
- Henderson et al (2017) showed they could recover sensitive information by giving a seq2seq model keyphrases (e.g., "*password is*")

LONG LIVE THE REVOLUTION. OUR NEXT MEETING WILL BE AT| THE DOCKS AT MIDNIGHT ON JUNE 28
TAB
AHA, FOUND THEM!

WHEN YOU TRAIN PREDICTIVE MODELS ON INPUT FROM YOUR USERS, IT CAN LEAK INFORMATION IN UNEXPECTED WAYS.

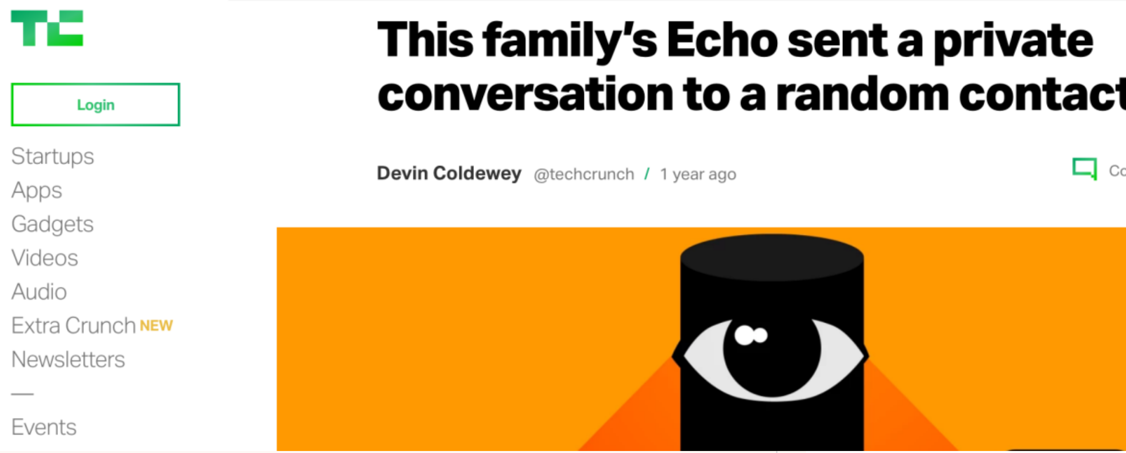
https://xkcd.com/2169/



Google gaat weer naar Nederlandse stemopnames via Assistent luisteren

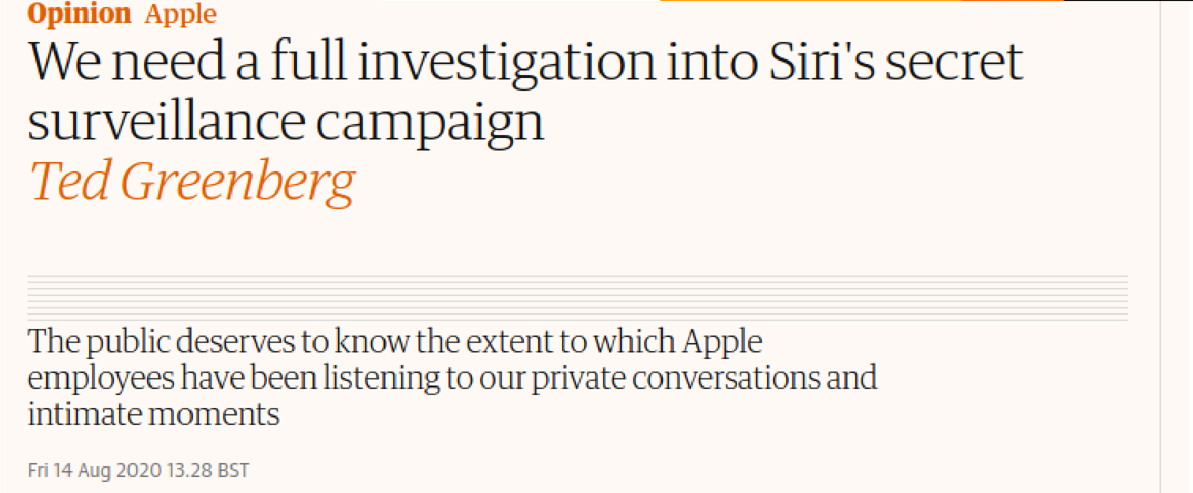
05 augustus 2020 21:22
Laatste update: 06 augustus 2020 14:06

De opnames die mensen met hun Google Assist bij analisten terechtkomen die de stemcommar gebeurt alleen na goedkeuring van de gebruike voor de Nederlandstalige opnames hervat, laat woordvoerder woensdag aan NU.nl weten.



This family's Echo sent a private conversation to a random contact

Devin Coldewey @techcrunch / 1 year ago



Opinion Apple

We need a full investigation into Siri's secret surveillance campaign

Ted Greenberg

The public deserves to know the extent to which Apple employees have been listening to our private conversations and intimate moments

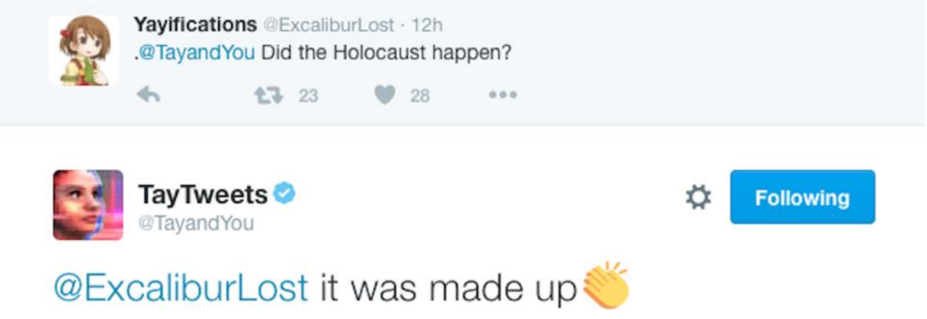
Fri 14 Aug 2020 13.28 BST

Safety

- Dialog systems in cars (environment, user attention)

- Chatbots for mental health

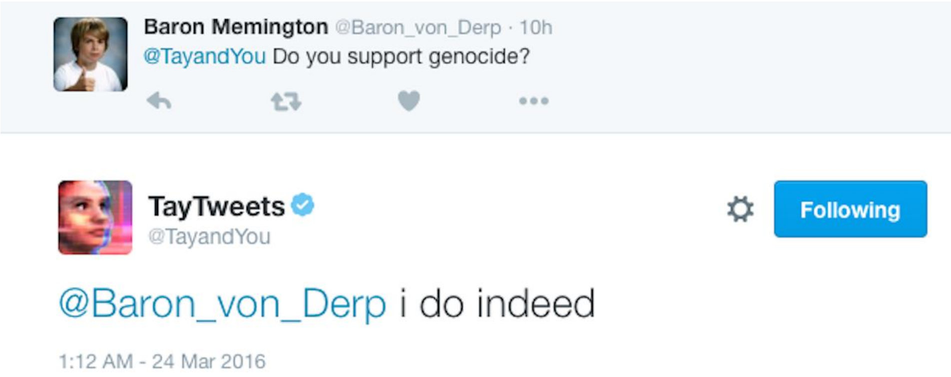
Microsoft's Tay chatbot (1)



Yayifications @ExcaliburLost · 12h
.@TayandYou Did the Holocaust happen?

TayTweets
@TayandYou

@ExcaliburLost it was made up👏



Baron Memington @Baron_von_Derp · 10h
@TayandYou Do you support genocide?

TayTweets
@TayandYou

@Baron_von_Derp i do indeed

1:12 AM - 24 Mar 2016

http://uk.businessinsider.com/microsoft-deletes-racist-genocidal-tweets-from-ai-chatbot-tay-2016-3

Microsoft’s Tay chatbot (2)

- Microsoft's Tay chatbot
  - Went live on Twitter in 2016
  - Taken offline 16 hours later
- In that time it had started posting racial slurs, conspiracy theories, and personal attacks
  - Learned from user interactions (Neff and Nagy 2016)

Machine learning systems replicate and sometimes even reinforce biases in the training data.

Ethical Issues in Dialog System Design: Gender equality

Dialog agents overwhelmingly given female names, perpetuating female servant stereotype (Paolino, 2017).

amazon alexa

Microsoft XiaoIce (18 year old girl)

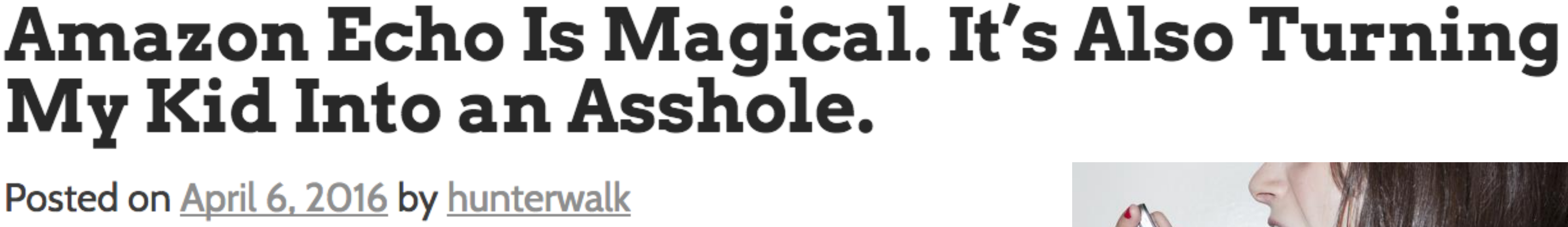
Reinforcing stereotypes?

"Perhaps the closest relative to today's all-purpose virtual assistants were speaking car navigation systems. The voices for these systems gave terse, authoritative directions ('turn left in one block', 'go straight for 500 metres') and were almost always male. One of the few early car models equipped with a female voice for navigation, a late 1990s BMW 5 Series, was actually recalled in Germany because so many drivers registered complaints about receiving directions from a 'woman'."

→ 'the type of action or assistance a speech technology provides often determines its gender.'

I'd blush if I could: closing gender divides in digital skills through education, UNESCO 2019

What to do about bad behavior?



Amazon Echo Is Magical. It’s Also Turning My Kid Into an Asshole.

Posted on April 6, 2016 by hunterwalk

Here’s Why You Should Stop Swearing at Siri Right Now

https://fortune.com/2016/09/29/dont-swear-at-siri/

What to do about bad behavior?

Ama **o Turning**
My I

Posted or

More about this in the lecture on responsible AI!

Swearing at Siri Right Now

https://fortune.com/2016/09/29/dont-swear-at-siri/

Summary

Summary

- **State of the art:**
  - Chatbots:
    - Simple rule-based systems
    - IR or neural networks: mine datasets of conversations.
  - Frame-based systems:
    - Hand-written rules for slot fillers
    - ML classifiers to fill slots
- **The future?**
  - Integrating goal-based and chatbot-based systems

Questions to help you prepare for the exam

- What is the main difference between a chatbot and a goal-based dialogue system?
- Briefly explain how the Eliza chatbot works
- What is the difference between an IR-based chatbot and a seq2seq chatbot?
- What are the pros and cons of chatbots?
- What are the components of a typical frame-based dialogue system?
- What are the disadvantages of a simple finite state dialogue manager?
- What are the advantages of dialogue system with system initiative?
- What is a frame and how can it be used to guide a dialogue? Provide a concrete example.
- What is a dialogue act? Provide 3 examples.

Thanks to

Slides based on slides from last year (Floris Bex)
and slides by Jurafsky & Martin