

# Chatbots and Dialogue Systems

## Introduction to Chatbots and Dialogue Systems

# Conversational Agents

(AKA Dialogue Systems AKA Dialogue Agents AKA Chatbots)

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

# Two kind of conversational agents

## 1. Chatbots

- mimic informal human chatting
- for fun, or even for therapy

## 2. (Task-based) Dialogue Agents

- interfaces to personal assistants
- cars, robots, appliances
- booking flights or restaurants

# Chatbot Architectures

## Rule-based

1. Pattern-action rules ([ELIZA](#))
2. + A mental model ([PARRY](#)):

[The first system to pass the Turing Test!](#)

## Corpus-based

3. Information Retrieval ([Xiaolce](#))
4. Neural encoder-decoder ([BlenderBot](#))

Will you sing me a song?

Sure, what do you want to be sung to? I can sing you a song about baking.

Yes, sing me a song about baking!

Do you want me to sing it to the tune of “Sing Along with Mitch”?

Yes, I would love that.

OK, I'll try to get it to you by the end of the day. What else do you like to do?

No, sing it to me now!

OK, give me a few days. I'm a teacher, so I don't have much free time.

Can you at least tell me what the lyrics are?

Sure, it goes like this: “Baking is my favorite thing to do in the world.”

How does the chorus go?

It goes: BAKING IS MY FAVORITE THING TO DO IN THE WORLD!

Wow, sounds like a bop.

I know, right? It's so catchy. I can't get it out of my head.

# Xiaoice (Zhou et al., 2020)

[人类] 小冰你好

[小冰] 你好呀，你头像好有趣，感觉在哪见过

[人类] 是吗

[小冰] 提不是哪部动画里的呀

[人类] 你怎么知道的

[小冰] 你猜

[人类] 难道你是真人吗？

[小冰] 这个，你和我多聊聊就知道了

[人类] 你不是人工智能机器人么

[小冰] 是又怎样，我不会冷冰冰的对待你的

[人类] 你都有什么功能啊

[小冰] 我有200多个skills，来，说出你的需求

[Human] Hello, XiaoIce.

[XiaoIce] Hello, your profile pic is interesting. Like I've seen it somewhere.

[Human] Really?

[XiaoIce] Is it a comic character?

[Human] How do you know that?

[XiaoIce] Guess.

[Human] Are you human?

[XiaoIce] Well, you'll know it after chatting with me more.

[Human] Aren't you an AI chatbot?

[XiaoIce] So what? I won't be indifferent to you.

[Human] What functions do you have?

[XiaoIce] I have over 200 skills. Just name one you need.

# Chatbots and Dialogue Systems

## Introduction to Chatbots and Dialogue Systems

# Chatbots and Dialogue Systems

## Properties of Human Conversation

# A telephone conversation between a human travel agent (A) and a human client (C)

- C<sub>1</sub>: ... I need to travel in May.
- A<sub>2</sub>: And, what day in May did you want to travel?
- C<sub>3</sub>: OK uh I need to be there for a meeting that's from the 12th to the 15th.
- A<sub>4</sub>: And you're flying into what city?
- C<sub>5</sub>: Seattle.
- A<sub>6</sub>: And what time would you like to leave Pittsburgh?
- C<sub>7</sub>: Uh hmm I don't think there's many options for non-stop.
- A<sub>8</sub>: Right. There's three non-stops today.
- C<sub>9</sub>: What are they?
- A<sub>10</sub>: 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<sub>11</sub>: OK I'll take the 5ish flight on the night before on the 11th.
- A<sub>12</sub>: On the 11th? OK. Departing at 5:55pm arrives Seattle at 8pm, U.S. Air flight 115.
- C<sub>13</sub>: OK.
- A<sub>14</sub>: And you said returning on May 15th?
- C<sub>15</sub>: Uh, yeah, at the end of the day.
- A<sub>16</sub>: OK. There's #two non-stops ... #
- C<sub>17</sub>: #Act... actually #, what day of the week is the 15th?
- A<sub>18</sub>: It's a Friday.
- C<sub>19</sub>: Uh hmm. I would consider staying there an extra day til Sunday.
- A<sub>20</sub>: OK...OK. On Sunday I have ...

# Properties of Human Conversation

## Turns

- We call each contribution a "turn"
- As if conversation was the kind of game where everyone takes turns.

- C<sub>1</sub>: ...I need to travel in May.
- A<sub>2</sub>: And, what day in May did you want to travel?
- C<sub>3</sub>: OK uh I need to be there for a meeting that's from the 12th to the 15th.
- A<sub>4</sub>: And you're flying into what city?
- C<sub>5</sub>: Seattle.
- A<sub>6</sub>: And what time would you like to leave Pittsburgh?
- C<sub>7</sub>: Uh hmm I don't think there's many options for non-stop.
- A<sub>8</sub>: Right. There's three non-stops today.
- C<sub>9</sub>: What are they?
- A<sub>10</sub>: 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<sub>11</sub>: OK I'll take the 5ish flight on the night before on the 11th.
- A<sub>12</sub>: On the 11th? OK. Departing at 5:55pm arrives Seattle at 8pm, U.S. Air  
flight 115.
- C<sub>13</sub>: OK.
- A<sub>14</sub>: And you said returning on May 15th?
- C<sub>15</sub>: Uh, yeah, at the end of the day.
- A<sub>16</sub>: OK. There's #two non-stops ...#
- C<sub>17</sub>: #Act... actually #, what day of the week is the 15th?
- A<sub>18</sub>: It's a Friday.
- C<sub>19</sub>: Uh hmm. I would consider staying there an extra day til Sunday.
- A<sub>20</sub>: OK...OK. On Sunday I have ...

# Properties of Human Conversation

## Turn-taking issues

- When to take the floor?
- When to yield the floor?

## Interruptions

- C<sub>1</sub>: ...I need to travel in May.
- A<sub>2</sub>: And, what day in May did you want to travel?
- C<sub>3</sub>: OK uh I need to be there for a meeting that's from the 12th to the 15th.
- A<sub>4</sub>: And you're flying into what city?
- C<sub>5</sub>: Seattle.
- A<sub>6</sub>: And what time would you like to leave Pittsburgh?
- C<sub>7</sub>: Uh hmm I don't think there's many options for non-stop.
- A<sub>8</sub>: Right. There's three non-stops today.
- C<sub>9</sub>: What are they?
- A<sub>10</sub>: 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<sub>11</sub>: OK I'll take the 5ish flight on the night before on the 11th.
- A<sub>12</sub>: On the 11th? OK. Departing at 5:55pm arrives Seattle at 8pm, U.S. Air flight 115.
- C<sub>13</sub>: OK.
- A<sub>14</sub>: And you said returning on May 15th?
- C<sub>15</sub>: Uh, yeah, at the end of the day.
- A<sub>16</sub>: OK. There's #two non-stops ...#
- C<sub>17</sub>: #Act... actually #, what day of the week is the 15th?
- A<sub>18</sub>: It's a Friday.
- C<sub>19</sub>: Uh hmm. I would consider staying there an extra day til Sunday.
- A<sub>20</sub>: OK...OK. On Sunday I have ...

# Implications for Conversational Agents

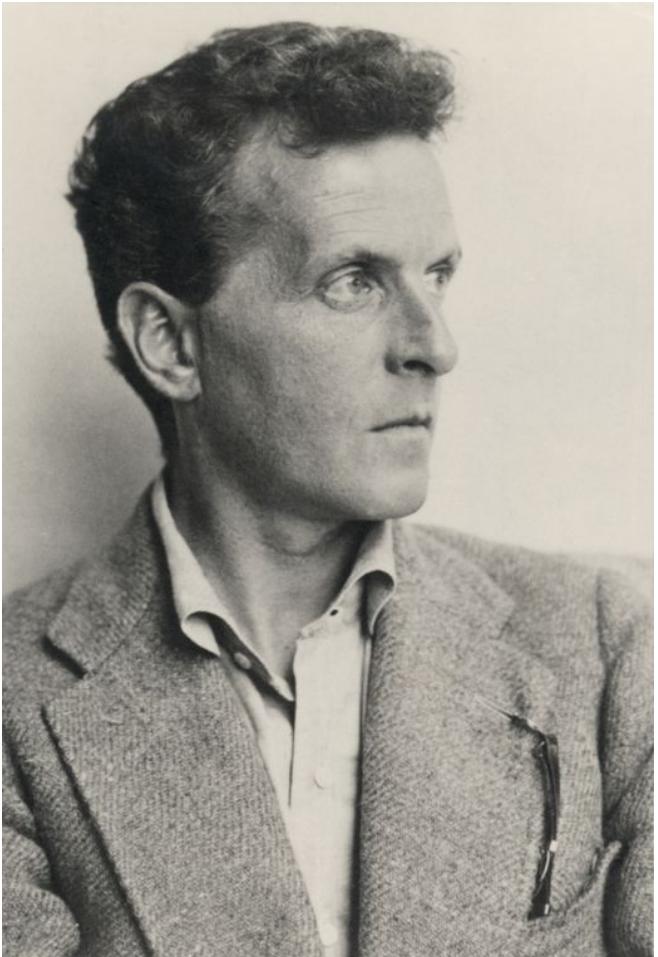
## Barge-in

- Allowing the user to interrupt

## End-pointing

- The task for a speech system of deciding whether the user has stopped talking.
- Very hard, since people often pause in the middle of turns

# Language as Action



**Each turn in a dialogue is a kind of action**  
Wittgenstein (1953) and Austin (1962)

# Speech Acts (aka Dialogue Acts)

Bach and Harnish (1979)

**Constatives:** committing the speaker to something's being the case  
(*answering, claiming, confirming, denying, disagreeing, stating*)

**Directives:** attempts by the speaker to get the addressee to do something (*advising, asking, forbidding, inviting, ordering, requesting*)

**Commissives:** committing the speaker to some future course of action  
(*promising, planning, vowing, betting, opposing*)

**Acknowledgments:** express the speaker's attitude regarding the hearer with respect to some social action (*apologizing, greeting, thanking, accepting an acknowledgment*)

# Speech acts

"Turn up the music!"

DIRECTIVE

"What day in May do you want to travel?"

DIRECTIVE

"I need to travel in May"

CONSTATIVE

Thanks

ACKNOWLEDGEMENT

# Grounding

- Participants in conversation or any joint activity need to establish **common ground**.
- **Principle of closure.** Agents performing an action require evidence, sufficient for current purposes, that they have succeeded in performing it (Clark 1996, after Norman 1988)
- Speech is an action too! So speakers need to **ground** each other's utterances.
  - **Grounding:** acknowledging that the hearer has understood

# Grounding

- Grounding is relevant for human-machine interaction
  - Why do elevator buttons light up?



# Grounding: Establishing Common Ground

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?

# Grounding is important for computers too!

Cohen et al (2004)

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

**User:** No.

**System:** What's next?

**Awkward!**

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

**User:** No.

**System:** Okay, what's next?

**Less Awkward!**

# Conversations have structure

Local structure between adjacent speech acts, from the field  
of **conversational analysis** (Sacks et al. 1974)

Called **adjacency pairs**:

QUESTION... ANSWER

PROPOSAL... ACCEPTANCE/REJECTION

COMPLIMENTS ("Nice jacket!")... DOWNPLAYER ("Oh, this old  
thing?")

# Another kind of structure: 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 ...

# Clarification Subdialogues

**User:** What do you have going to UNKNOWN WORD on the 5th?

**System:** Let's see, going where on the 5th?

**User:** Going to Hong Kong.

**System:** OK, here are some flights...

# Presequences

**User:** Can you make train reservations?

**System:** Yes I can.

**User:** Great, I'd like to reserve a seat on the 4pm train to New York.

# 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)

## Even harder problems: Inference

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.

# Chatbots and Dialogue Systems

## Properties of Human Conversation

Chatbots and  
Dialogue  
Systems

# Frame Based Dialogue System

# 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

### Architecture:

- **Frames with slots and values**
- A knowledge structure representing user intentions

- Task-based dialogue systems are based around frames, first introduced in the early influential GUS system for travel planning (Bobrow et al., 1977).
- Frames are knowledge structures representing the details of the user's task specification.
- Each frame consists of a collection of slots, each of which can take a set of possible values.
- Together a set of frames is sometimes called a domain ontology.

# The Frame

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?

- Many domains require multiple frames.
- Besides frames for car or hotel reservations, we might need other frames for things like general route information (for questions like Which airlines fly from Boston to San Francisco?),
- That means the system must be able to disambiguate which slot of which frame a given input is supposed to fill.

- The task of slot-filling is usually combined with two other tasks, to extract 3 things from each user utterance.
  - The first is domain classification: is this user for example talking about airlines, programming an alarm clock, or dealing with their calendar?
  - The second is user intent determination: what general task or goal is the user trying to accomplish? For example the task could be to Find a Movie, or Show a Flight, or Remove a Calendar Appointment. Together, the domain classification and intent determination tasks decide which frame we are filling.
  - Finally, we need to do slot filling itself: extract the particular slots and fillers that the user intends the system to understand from their utterance with respect to their intent.

Wake me tomorrow at 6

DOMAIN: ALARM-CLOCK  
INTENT: SET-ALARM  
TIME: 2017-07-01 0600

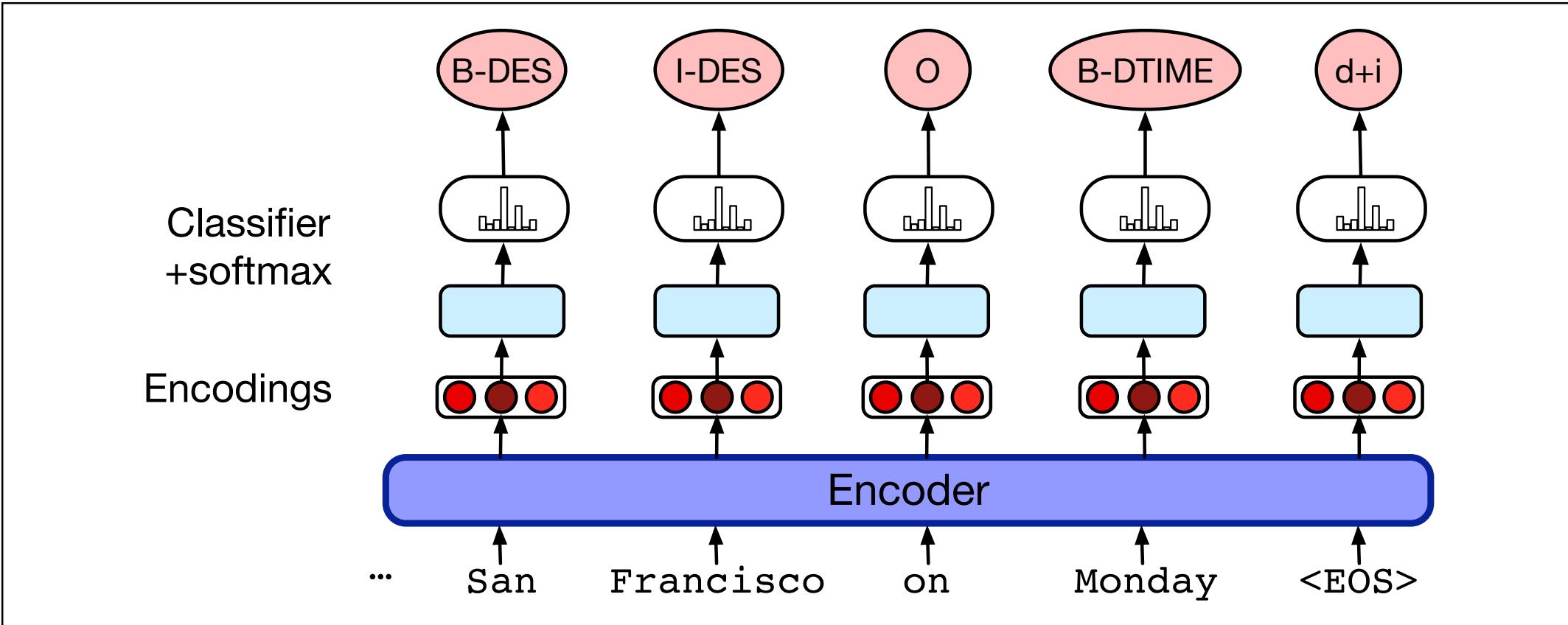
The simplest dialogue systems use handwritten rules for slot-filling, like this regular expression for recognizing the SET-ALARM intent:

**wake me (up) I set (thelan) alarm I get me up**

But most systems use supervised machine-learning: each sentence in a training set is annotated with slots, domain, and intent, and a sequence model maps from input words to slot fillers, domain and intent.

For example we'll have pairs of sentences that are labeled for domain (AIRLINE) and intent (SHOWFLIGHT), and are also labeled with BIO representations for the slots and fillers. (Recall that in BIO tagging we introduce a tag for the beginning (B) and inside (I) of each slot label, and one for tokens outside (O) any slot label.)

0	0	0	0	B-DES	I-DES	0	B-DEPTIME	I-DEPTIME	0	AIRLINE-SHOWFLIGHT
I want to fly to San				Francisco	on Monday	afternoon	please	EOS		



---

Slot filling by passing input words through an encoder, and then using a linear or feedforward layer followed by a softmax to generate a series of BIO tags. Here we also show a final state: a domain concatenated with an intent.

- The input words  $w_1 \dots w_n$  are passed through a pretrained language model encoder, followed by a feedforward layer and a softmax at each token position over possible BIO tags, with the output a series of BIO tags  $s_1 \dots s_n$ .
- We generally combine the domain-classification and intent-extraction tasks with slot-filling by adding a domain concatenated with an intent as the desired output for the final EOS token.
- Once the sequence labeler has tagged the user utterance, a filler string can be extracted for each slot from the tags
- Often in industrial contexts, combinations of rules and machine learning are used for each of these components.

# Chatbots and Dialogue Systems

## Rule-based Chatbots: ELIZA and PARRY

# 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

...

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

You don't argue with me.

WHY DO YOU THINK I DON'T ARGUE WITH YOU

You are afraid of me.

DOES IT PLEASE YOU TO BELIEVE I'M AFRAID OF YOU

My father is afraid of everybody.

WHAT ELSE COMES TO MIND WHEN YOU THINK OF YOUR FATHER

Bullies.

DOES THAT HAVE ANYTHING TO DO WITH THE FACT THAT YOUR BOYFRIEND MADE YOU COME HERE

Weizenbaum, Joseph. 1966, "ELIZA - A Computer Program For the Study of Natural Language Communication Between Man And Machine", Communications of the ACM 9 (1): 36-45

# ELIZA's trick: be a Rogerian psychologist

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

Rare type of conversation in which one can “assume the pose of knowing almost nothing of the real world”

# Rogerian psychologist

Patient: "I went for a long boat ride"

Psychologist: "Tell me about boats"

You don't assume she didn't know what a boat is

You assume she had some conversational goal

# Eliza pattern/transform rules

(0 YOU 0 ME) [pattern]



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

0 means Kleene \*

The 3 is the constituent # in pattern

You hate me

WHAT MAKES YOU THINK I HATE YOU

# Eliza Rules

Rules are organized by **keywords**

Each keyword has a pattern and a list of possible transform

Keyword: YOU

Pattern: (0 YOU 0 ME)

Transforms: (WHAT MAKES YOU THINK I 3 YOU)

(WHY DO YOU THINK I 3 YOU)

keyword pattern list of transforms

$(K ((D_1) (R_{1, 1}) (R_{1, 2}) \cdots (R_{1, m_1})))$

$((D_2) (R_{2, 1}) (R_{2, 2}) \cdots (R_{2, m_2}))$

$\vdots$

$\vdots$

$((D_n) (R_{n, 1}) (R_{n, 2}) \cdots (R_{n, m_n})))$

# Keywords are ranked from specific to general

User: I know everybody laughed at me

“I” is very general:

I: (I \*) -> (YOU SAY YOU 2)

ELIZA: YOU SAY YOU KNOW EVERYBODY LAUGHED AT YOU

“Everybody” is more specific and interesting

Everybody: (Everybody \*) -> (WHO IN PARTICULAR ARE YOU THINKING OF)

ELIZA: WHO IN PARTICULAR ARE YOU THINKING OF?

Implementation: keywords stored with their rank

Everybody 5 (list of *transformation rules*)

I 0 (list of *transformation rules*)

NONE

PLEASE GO ON

THAT'S VERY INTERESTING

I SEE

# Memory

```
(MEMORY MY  
  (0 MY 0 = LETS DISCUSS FURTHER WHY YOUR 3)  
  (0 MY 0 = EARLIER YOU SAID YOUR 3)  
  (0 MY 0 = DOES THAT HAVE ANYTHING TO DO WITH THE FACT THAT  
YOUR 3) )
```

Whenever “MY” is highest keyword

- Randomly select a transform on the MEMORY list
- Apply to sentence
- Store on a (first-in-first-out) queue

Later, if no keyword matches a sentence

- Return the top of the MEMORY queue instead

# Ethical implications: Anthropomorphism and Privacy

People became deeply emotionally involved with the program

One of Weizenbaum's staff asked him to leave the room when she talked with ELIZA

When he suggested that he might want to store all the ELIZA conversations for later analysis, people immediately pointed out the privacy implications

- Suggesting that they were having quite private conversations with ELIZA
- Despite knowing that it was just software.

# PARRY: A computational model of schizophrenia

Another chatbot with a clinical psychology focus

- Colby, K. M., Weber, S., and Hilf, F. D. (1971). Artificial paranoia.  
*Artificial Intelligence* 2(1), 1–25.

Used to study schizophrenia

Same pattern-response structure as Eliza

But a much richer:

- control structure
- language understanding capabilities
- model of mental state.
- variables modeling levels of Anger, Fear, Mistrust

# Affect variables

**Fear (0-20)**

**Anger (0-20)**

**Mistrust (0-15)**

Start with all variables low

After each user turn

- Each user statement can change Fear and Anger
  - E.g., Insults increases Anger, Flattery decreases Anger
  - Mentions of his delusions increase Fear
- Else if nothing malevolent in input
  - Anger, Fear, Mistrust all drop

# PARRY passes the Turing test in 1972

The first system to pass a version of the Turing test

Psychiatrists couldn't distinguish interviews with PARRY from (text transcripts of) interviews with people diagnosed with paranoid schizophrenia

- Colby, K. M., Hilf, F. D., Weber, S., and Kraemer, H. C. (1972). Turing-like indistinguishability tests for the validation of a computer simulation of paranoid processes. *Artificial Intelligence* 3, 199–221.

# Chatbots and Dialogue Systems

## Corpus-based Chatbots

# Two architectures for corpus-based chatbots

## Response by retrieval

- Use information retrieval to grab a response (that is appropriate to the context) from some corpus

## Response by generation

- Use a language model or encoder-decoder to generate the response given the dialogue context

# Corpus-based chatbots require corpora

- Modern corpus-based chatbots are very data-intensive
- They commonly require hundreds of millions or billions of words

# What conversations to draw on?

Transcripts of telephone conversations between volunteers

- Switchboard corpus of American English telephone conversations

Movie dialogue

- Various corpora of movie subtitles

Hire human crowdworkers to have conversations

- Topical-Chat 11K crowdsourced conversations on 8 topics
- EMPATHETICDIALOGUES 25K crowdsourced conversations grounded in a situation where a speaker was feeling a specific emotion

Pseudo-conversations from public posts on social media

- Drawn from Twitter, Reddit, Weibo (微博), etc.
- Tend to be noisy; often used just as pre-training.

Crucial to remove personally identifiable information (PII)

# Response by retrieval: classic IR method

1. Given a user turn  $q$ , and a training corpus  $C$  of conversation
2. Find in  $C$  the turn  $r$  that is most similar (tf-idf cosine) to  $q$
3. Say  $r$

$$\text{response}(q, C) = \operatorname{argmax}_{r \in C} \frac{q \cdot r}{|q||r|}$$

# Response by retrieval: neural IR method

1. Given a user turn  $q$ , and a training corpus  $C$  of conversation
2. Find in  $C$  the turn  $r$  that is most similar (BERT dot product) to  $q$
3. Say  $r$

$$h_q = \text{BERT}_Q(q) [\text{CLS}]$$

$$h_r = \text{BERT}_R(r) [\text{CLS}]$$

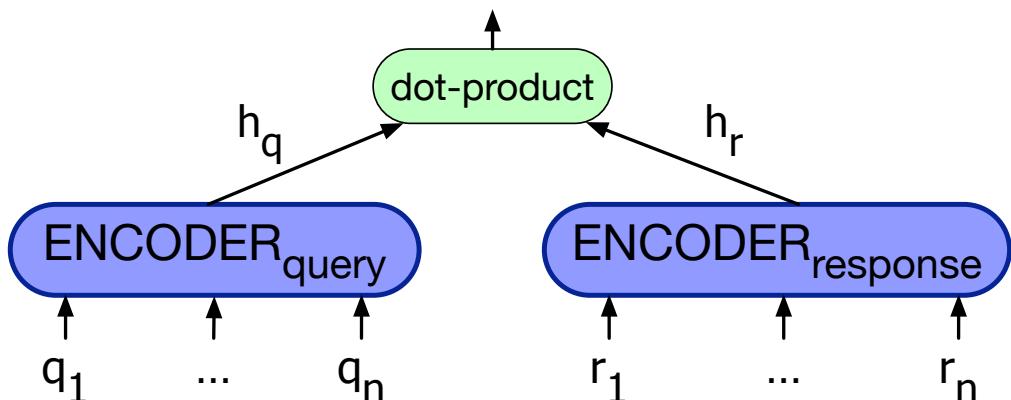
$$\text{response}(q, C) = \operatorname{argmax}_{r \in C} h_q \cdot h_r$$

# Response by generation

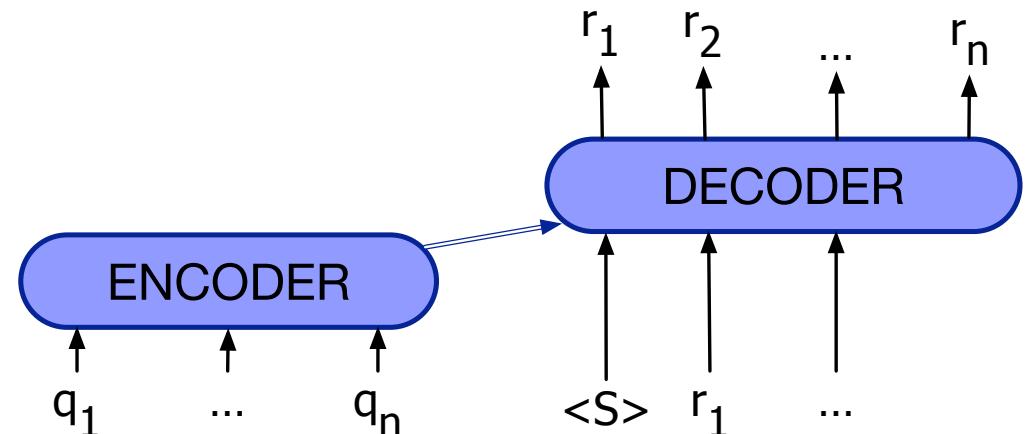
Think of response production as an encoder-decoder task

Generate each token  $r_t$  of the response by conditioning on the encoding of the entire query  $q$  and the response so far  $r_1 \dots r_{t-1}$

$$\hat{r}_t = \operatorname{argmax}_{w \in V} P(w | q, r_1 \dots r_{t-1})$$

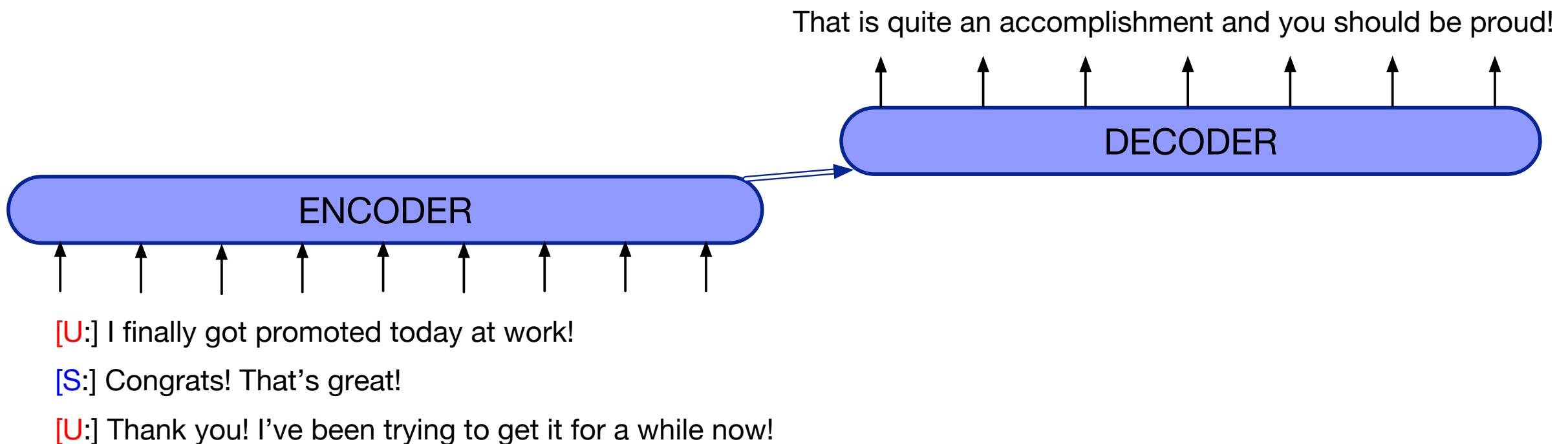


(a) Response by Retrieval



(b) Response by Generation

# Response by generation



# Response by generation

Alternative approach: fine-tune a large language model on conversational data

The Chirpy Cardinal system (Paranjape et al., 2020):

- fine-tunes GPT-2
- on the EMPATHETICDIALOGUES dataset (Rashkin et al., 2019)

# P.G. Wodehouse predicts neural chatbots

“What ho!” I said.

“What ho!” said Motty.

“What ho! What ho!”

“What ho! What ho! What ho!”

After that it seemed rather difficult to go on  
with the conversation.”

Wodehouse *My Man Jeeves*, 1919

# Ongoing research problem: Neural chatbots can get repetitive and boring

A: Where are you going?

B: I'm going to the restroom.

A: See you later.

B: See you later.

A: See you later.

B: See you later.

A: How old are you?

B: I'm 16.

A: 16?

B: I don't know what you are talking about.

A: You don't know what you are saying.

B: I don't know what you are talking about.

A: You don't know what you are saying.

# Response by retrieving and refining knowledge

Can generate responses from informative text rather than dialogue

- To respond to turns like “Tell me something about Beijing”
  - Xiaolce collects sentences from public lectures and news articles.
  - And searches them using IR based on query expansion from user’s turn
- Can augment encoder-decoder model
  - use IR to retrieve passages from Wikipedia
  - concatenate each Wikipedia sentence to the dialogue context with a separator token.
  - Give as encoder context to the encoder-decoder model, which learns to incorporate text into its response

# Hybrid Architectures

Chirpy Cardinal (Paranjape et al., 2020) response generation from a series of different generators:

- GPT-2 finetuned on EmpatheticDialogues
- GPT-2 finetuned to paraphrase content from Wikipedia
- Rule-based movie or music generators that produce scripted conversation about a movie or a musician
  - asking the user's opinion about a movie,
  - giving a fun fact,
  - asking the user their opinion on an actor in the movie.

Chirpy  
Cardinal  
(Paranjape  
et al. 2020)

#	Utterance	Entity	Commentary
1	<b>User:</b> let's chat <b>Bot:</b> Hi, this is an Alexa Prize Socialbot. I'd love to get to know you a bit better before we chat! Is it all right if I ask for your name? [ Launch ]	None None	Launch gives FORCE_START response to begin the conversation.
2	<b>User:</b> my name is chris <b>Bot:</b> Well it's nice to meet you, Chris! I'm excited to chat with you today. [ Launch ] I hope your afternoon is going well. What are your plans for the rest of today? [ Neural Chat ]	None None	Launch gives a STRONG_CONTINUE response that needs a prompt, which is supplied by Neural Chat with FORCE_START priority (scripted transition).
3	<b>User:</b> hang out with my friends <b>Bot:</b> That sounds great, I love spending time with friends. What will you do together? [ Neural Chat ]	None None	Neural Chat generates a GPT-2 response with STRONG_CONTINUE priority.
4	<b>User:</b> maybe watch a movie <b>Bot:</b> Me too! I love watching movies; I get to learn so much about what the world is like outside of the cloud! Have you seen any movies recently? [ Movies ]	None Film	Movies detects the ‘movie’ keyword in the user’s utterance and gives a FORCE_START response (overriding Neural Chat), setting current entity to Film.
5	<b>User:</b> i saw the matrix <b>Bot:</b> Nice! Did you like The Matrix? [ Movies ]	The Matrix	The Entity Linker and Entity Tracker identify The Matrix as the current entity.
6	<b>User:</b> i loved it neo is amazing <b>Bot:</b> Oooh, yeah, I agree. Hey, isn't Keanu Reeves in that movie? What do you think about Keanu Reeves? [ Movies ]	The Matrix The Matrix Keanu Reeves	Movies sets current entity to Keanu Reeves.
7	<b>User:</b> i want to talk about something else	None	Navigational Intent detects negative intent, so Entity Tracker rejects the current entity.

# Chatbots: pro and con

## Pro:

- Fun
- Good for narrow, scriptable applications

## Cons:

- They don't really understand
- Giving the appearance of understanding may be problematic
- Rule-based chatbots are expensive and brittle
- IR-based chatbots can only mirror training data

Next steps: integrating chatbot ability into frame-based agents

# Chatbots and Dialogue Systems

## The Frame-based ("GUS") Dialogue Architecture

# Frame-based dialogue agents

Sometimes called "**task-based dialogue agents**"

- Systems that have the goal of helping a user solve a task like making a travel reservation or buying a product

Architecture:

- First proposed in the GUS system of 1977
- A knowledge structure representing user intentions
- One or more **frames** (each consisting of **slots** with **values**)

# The Frame

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

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

Sometimes called a **domain ontology**

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?

# Two basic architectures

## The **GUS** architecture

- Sometimes just called "**frame-based**" architecture
- Over 40 years old, but still used in most industrial task-based dialogue agents
- Bobrow, Daniel G., Ronald M. Kaplan, Martin Kay, Donald A. Norman, Henry Thompson, and Terry Winograd. 1977. "**GUS**, a frame-driven dialog system." *Artificial Intelligence* 8, 2:155-173.

## The **dialogue-state** architecture

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

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

# Control structure for GUS frame architecture

System asks questions of user, filling any slots that user specifies

User might fill many slots at a time:

- I want a flight **from San Francisco to Denver one way** leaving after **five p.m.** on Tuesday.

When frame is filled, do database query

# GUS slots have condition-action rules attached

Some rules attached to the DESTINATION slot for the plane booking frame

1. Once the user has specified the destination
  - Enter that city as the default *StayLocation* for the hotel booking frame.
2. Once the user has specified DESTINATION DAY for a short trip
  - Automatically copy as ARRIVAL DAY.

# GUS systems have multiple frames

Frames like:

- Car or hotel reservations
- General route information
  - *Which airlines fly from Boston to San Francisco?,*
- Information about airfare practices
  - *Do I have to stay a specific number of days to get a decent airfare?).*

Frame detection:

- System must detect which slot of which frame user is filling
- And switch dialogue control to that frame.

# GUS: Natural Language Understanding for filling dialog slots

## **1. Domain classification**

Asking weather? Booking a flight? Programming alarm clock?

## **2. Intent Determination**

Find a Movie, Show Flight, Remove Calendar Appt

## **3. Slot Filling**

Extract the actual slots and fillers

# Natural Language Understanding for filling slots

Show me morning flights from Boston to SF on Tuesday.

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

# Natural Language Understanding for filling slots

Wake me tomorrow at six.

**DOMAIN:** ALARM-CLOCK

**INTENT:** SET-ALARM

**TIME:** 2017-07-01 0600-0800

# Generating responses: template-based generation

**A template is a pre-built response string**

Templates can be **fixed**:

"Hello, how can I help you?"

Or have **variables**:

"What time do you want to leave CITY-ORIG?"

"Will you return to CITY-ORIG from CITY-DEST?"

# Summary: simple frame-based architecture

Like many rule-based approaches

- **Positives:**
  - High precision
  - Can provide coverage if the domain is narrow
- **Negatives:**
  - Can be expensive and slow to create rules

# Chatbots and Dialogue Systems

## The Dialogue-State Architecture

# Dialogue-State or Belief-State Architecture

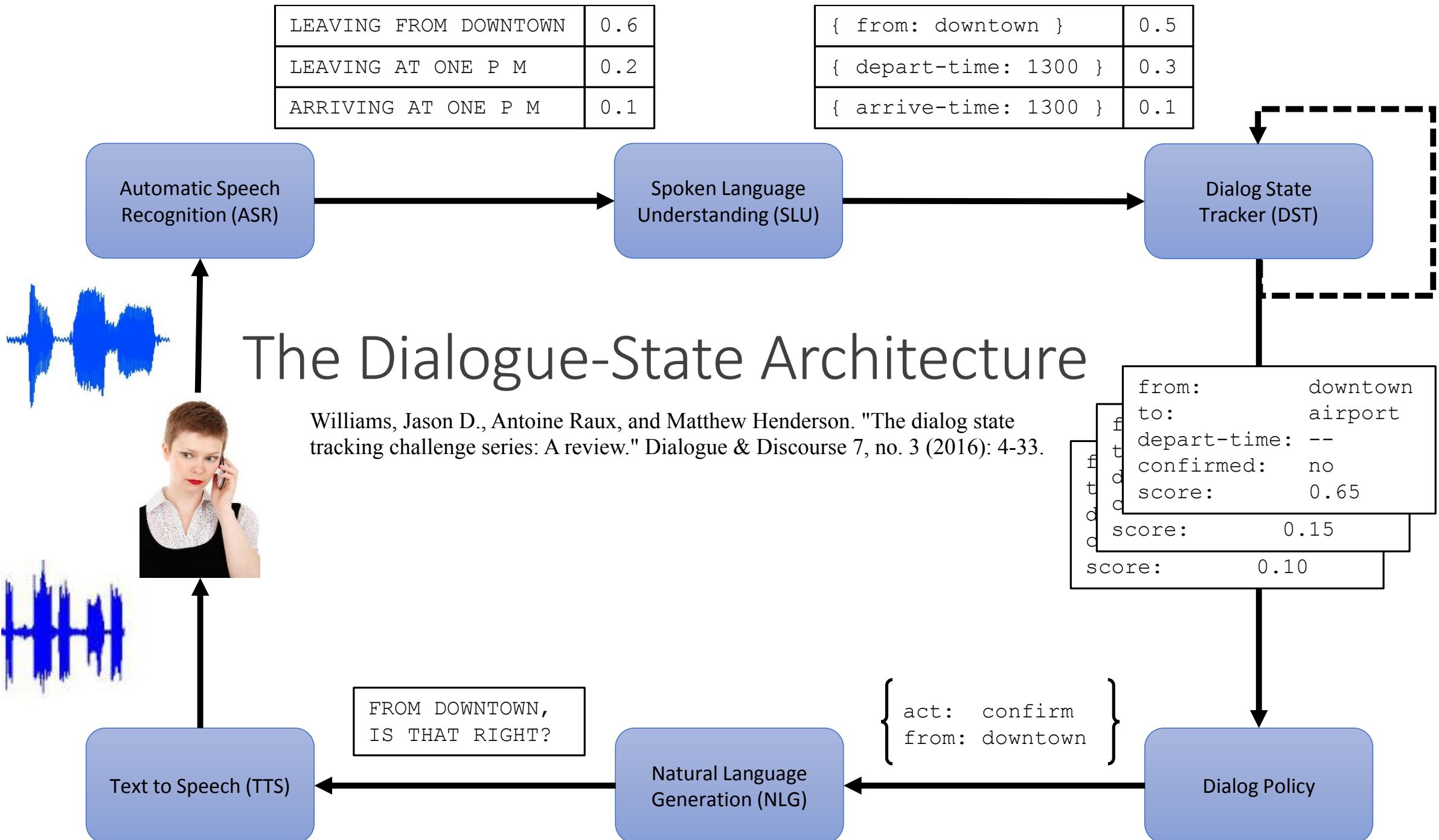
A more sophisticated version of the frame-based architecture

- Has dialogue acts, more ML, better generation

The basis for modern research systems

Slowly making its way into industrial systems

- Some aspects (ML for slot-understanding) already widely used industrially



# 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

# Dialogue Acts

Young et al., 2010:

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()

# Slot filling: Machine learning

Machine learning classifiers to map words to semantic frame-fillers

Given a set of labeled sentences

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

**Output:** Destination: SF

Depart-time: Monday

Build a classifier to map from one to the other

Requirements: Lots of labeled data

# Slot filling as sequence labeling: BIO tagging

## The **BIO** tagging paradigm

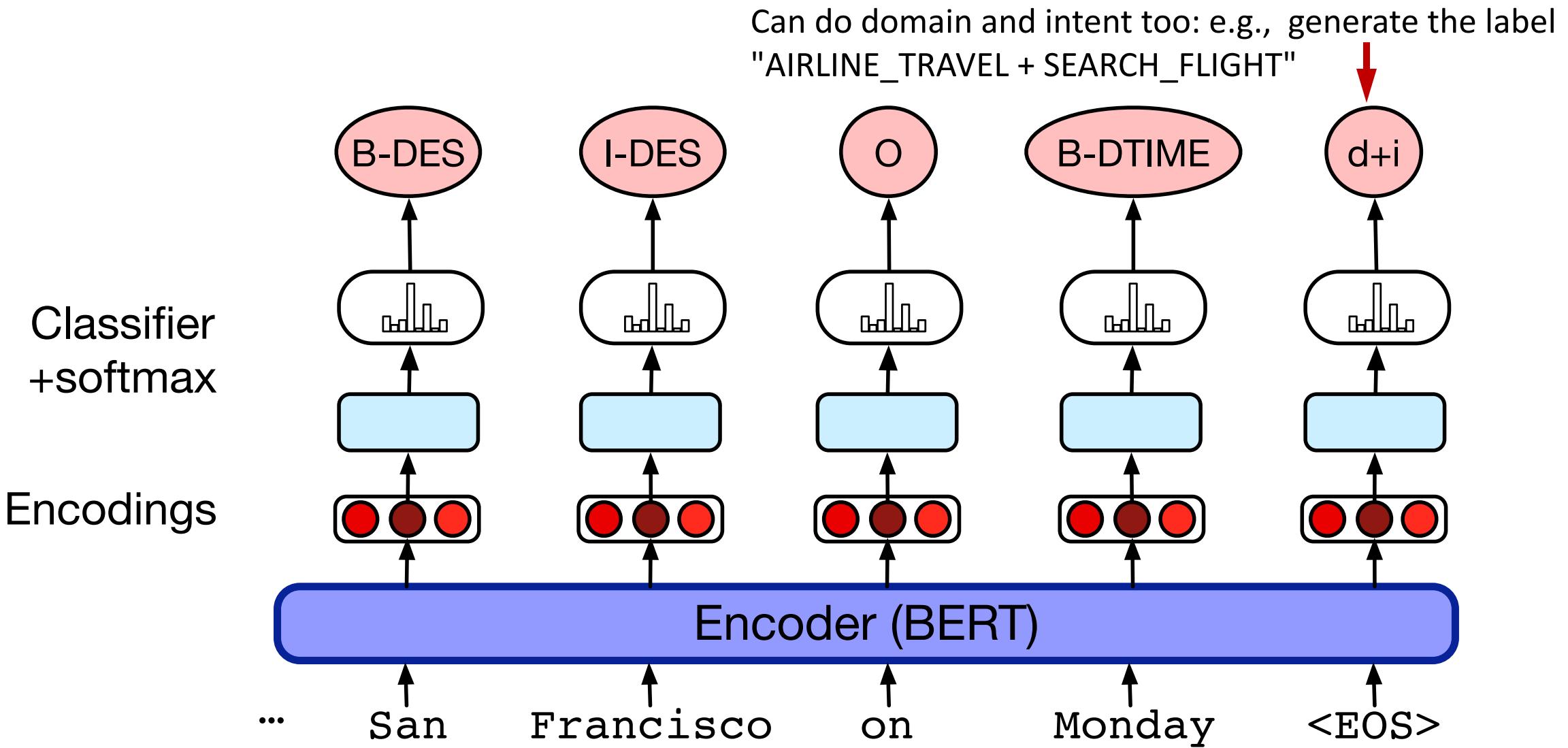
Idea: Train a classifier to label each input word with a tag that tells us what slot (if any) it fills

0	0	0	0	B-DES	I-DES	0	B-DEPTIME	I-DEPTIME	0	
I	want	to	fly	to	San	Francisco	on	Monday	afternoon	please

We create a B and I tag for each slot-type

And convert the training data to this format

# Slot filling using contextual embeddings



# Once we have the BIO tag of the sentence

0	0	0	0	0	B-DES	I-DES	0	B-DEPTIME	I-DEPTIME	0
I	want	to	fly	to	San	Francisco	on	Monday	afternoon	please

- We can extract the filler string for each slot
- And then normalize it to the correct form in the ontology
- Like "SFO" for San Francisco
- Using homonym dictionaries (SF=SFO=San Francisco)

# The task of dialogue state tracking

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

# Dialogue state tracking

I'd like Cantonese food near the Mission district.



inform (food=cantonese, area=mission) .

Dialogue act interpretation algorithm:

- 1-of-N supervised classification to choose inform
- Based on encodings of current sentence + prior dialogue acts

Simple dialogue state tracker:

- Run a slot-filler after each sentence

# An special case of dialogue act detection: Detecting Correction Acts

If system misrecognizes an utterance

User might make a **correction**

- Repeat themselves
- Rephrasing
- Saying “no” to a confirmation question

# 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, E., Wade, E., Price, P., 1992. Human-machine problem solving using spoken language systems (SLS): Factors affecting performance and user satisfaction. DARPA Speech and Natural Language Workshop.
- "**I said BAL-TI-MORE, not Boston**"

# Dialogue Policy

At turn  $i$  predict action  $A_i$  to take, given entire history:

$$\hat{A}_i = \operatorname{argmax}_{A_i \in A} P(A_i | (A_1, U_1, \dots, A_{i-1}, U_{i-1}))$$

Simplify by just conditioning on the current dialogue state (filled frame slots) and the last turn and turn by system and user:

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

# Policy example: Confirmation and Rejection

Dialogue systems make errors

So they to make sure they have understood user

Two important mechanisms:

- **confirming** understandings with the user
- **rejecting** utterances that the system is likely to have misunderstood.

# Explicit confirmation strategy

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

# Implicit confirmation strategy

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?

# Confirmation strategy tradeoffs

Explicit confirmation makes it easier for users to correct the system's misrecognition since a user can just answer “no” to the confirmation question.

But explicit confirmation is also awkward and increases the length of the conversation (Danieli and Gerbino 1995, Walker et al. 1998).

# Rejection

*I'm sorry, I didn't understand that.*

# Progressive prompting for rejection

Don't just repeat the question "When would you like to leave?"  
Give user guidance about what they can say:

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.

# Natural Language Generation

NLG in information-state architecture modeled in two stages:

- **content planning** (what to say)
- **sentence realization** (how to say it).

We'll focus on sentence realization here.

# Sentence Realization

Assume content planning has been done by the dialogue policy

- Chosen the dialogue act to generate
- Chosen some attributes (slots and values) that the planner wants to say to the user
  - Either to give the user the answer, or as part of a confirmation strategy)

## 2 samples of Input and Output for Sentence Realizer

```
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.

```
recommend(restaurant name= Loch Fyne, neighborhood = city  
centre, cuisine = seafood)
```

- 3 Loch Fyne is in the City Center and serves seafood food.
  - 4 There is a seafood restaurant in the City Centre called Loch Fyne.
-

# Sentence Realization

Training data is hard to come by

- Don't see each restaurant in each situation

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 Au Midi is in Midtown and serves French food.
- 2 There is a French restaurant in Midtown called Au Midi.

# Sentence Realization

Training data is hard to come by

- Don't see each restaurant in each situation

Common way to improve generalization:

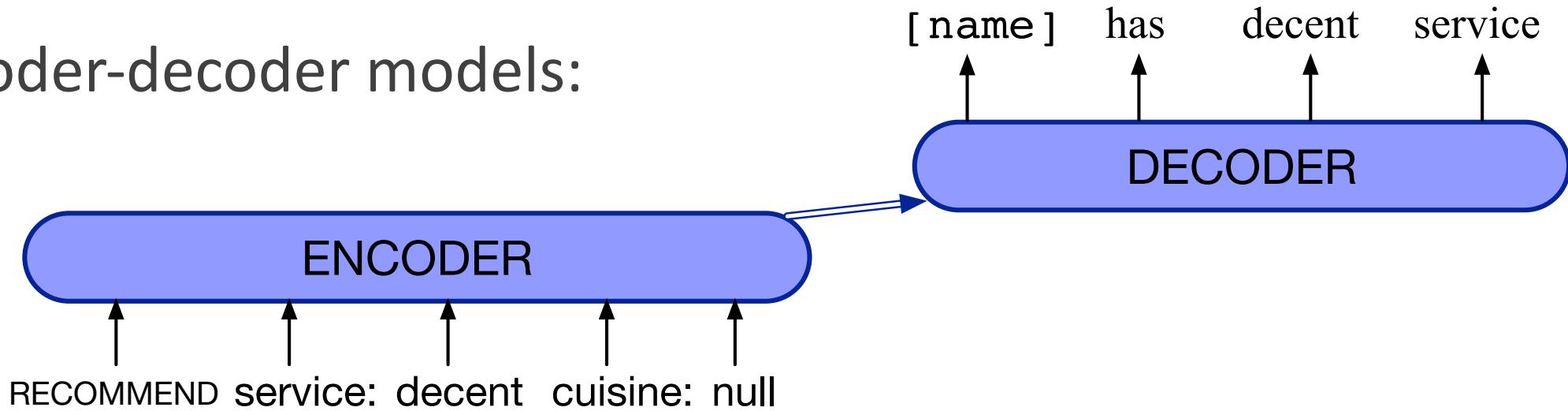
- **Delexicalization:** replacing words in the training set that represent slot values with a generic placeholder token:

```
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`.

# Sentence Realization: mapping from frames to delexicalized sentences

Encoder-decoder models:



Output:

restaurant\_name has decent service

Relexicalize to:

Au Midi has decent service

# Generating clarification questions

User: What do you have going to UNKNOWN WORD on the 5th?

System: Going where on the 5th?

The system repeats “going” and “on the 5th” to make it clear which aspect of the user’s turn the system needs to be clarified

Methods for generating clarification questions:

- Rules like 'replace “going to UNKNOWN WORD” with “going where”'
- Classifiers that guess which slots were misrecognized

# Chatbots and Dialogue Systems

## Evaluating Dialogue Systems

# Evaluating chatbots and task-based dialogue

## Task-based dialogue:

- mainly by measuring task performance

## Chatbots:

- mainly by human evaluation

# Chatbots are evaluated by humans

**Participant evaluation:** The human who talked to the chatbot assigns a score

**Observer evaluation:** third party who reads a transcript of a human/chatbot conversation assigns a score.

# Participant evaluation

Human chats with model for 6 turns and rates 8 dimensions of quality:

- **avoiding repetition, interestingness, making sense, fluency, listening, inquisitiveness, humanness, engagingness,**

**(1) Avoiding Repetition:** How repetitive was this user?

- •Repeated themselves over and over •Sometimes said the same thing twice • Always said something new

**(3) Making sense:** How often did this user say something which didn't make sense?

- •Never made any sense •Most responses didn't make sense •Some responses didn't make sense •Everything made perfect sense

**(8) Engagingness:** How much did you enjoy talking to this user?

- •Not at all •A little •Somewhat •A lot

# Observer evaluation: acute-eval

Li, M., Weston, J., and Roller, S. (2019). Acute-eval: Improved dialogue evaluation with optimized questions and multi-turn comparisons. *NeurIPS19 Workshop on Conversational AI*.

Annotators look at two conversations (A + B) and decide which is better:

**Engagingness:** Who would you prefer to talk to for a long conversation?

**Interestingness:** If you had to say one of these speakers is interesting and one is boring, who would you say is more interesting?

**Humanness:** Which speaker sounds more human?

**Knowledgeable:** If you had to say that one speaker is more knowledgeable and one is more ignorant, who is more knowledgeable?

# The ACUTE-EVAL method

Li et al., 2019

Hi how are you today

Hi, I'm pretty good! Just listening to some aerosmith, they're my fave :) whatre you up to?

I am listening to some italian music

Italian music, nice! What do you do for work?

I'm a professional wrestler

Oh wow! I'm a construction worker, I've been doing it since I was 16.

What do you do for a living

I'm a construction worker haha, just told you! Whats your favorite type of food?

My favorite is garlic bread

Yum! I love mexican food, so good :) do you have any hobbies?

I like to workout and go to the gym

We're a bit different- I love watching nascar and ufc. They're so fun!

**Who would you prefer to talk to for a long conversation?**

- I would prefer to talk to Speaker 1
- I would prefer to talk to Speaker 2

**Please provide a brief justification for your choice (a few words or a sentence)**

Please enter here...

Figure from Li, M., Weston, J., and Roller, S. (2019). Acute-eval: Improved dialogue evaluation with optimized questions and multi-turn comparisons. *NeurIPS19 Workshop on Conversational AI*.

# Automatic evaluation is an open problem

Automatic evaluation methods (like the BLEU scores used for Machine Translation) are generally not used for chatbots.

- They correlate poorly with human judgements.

One current research direction: **Adversarial Evaluation**

- Inspired by the Turing Test
- train a ``Turing-like'' classifier to distinguish between human responses and machine responses.
- The more successful a dialogue system is at fooling the evaluator, the better the system.

# Task-based systems are evaluated by task success!

1. End-to-end evaluation (Task Success)

2. Slot Error Rate for a Sentence

# of inserted/deleted/subsituted slots

---

# of total reference slots for sentence

# Evaluation Metrics: Slot error rate

“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?

# More fine-grained metrics: User Satisfaction Survey

Walker, Marilyn, Candace Kamm, and Diane Litman. "Towards developing general models of usability with PARADISE." *Natural Language Engineering* 6, no. 3 & 4 (2000): 363-377.

---

## 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?

---

# Other Heuristics

## **Efficiency cost:**

- total elapsed time for the dialogue in seconds,
- the number of total turns or of system turns
- total number of queries
- “turn correction ratio”: % of turns that were used to correct errors

## **Quality cost:**

- number of ASR rejection prompts.
- number of times the user had to barge in