

Natural Language Processing

IN2361

Prof. Dr. Georg Groh

Chapter 24

Dialog Systems and Chatbots

- content is based on [1]
- certain elements (e.g. equations or tables) were taken over or taken over in a modified form from [1]
- citations of [1] or from [1] are omitted for legibility
- errors are fully in the responsibility of Georg Groh
- BIG thanks to Dan and James for a great book!

Dialog Systems

- **Task-oriented dialog agents:**
 - designed for a particular task + short related conversations to get information from the user
 - examples: Siri, Cortana, Alexa, Google Now/Home, etc.
 - give travel directions, control home appliances, find restaurants, or help make phone calls or send texts, etc.
- **Chatbots:** designed for extended, casual conversations
 - examples: Microsoft's Xiaolce (小冰 *Xiǎobīng*) (2015), Eliza (1969)
- characteristic: **how many turns** can the system handle (one → mostly question answering up to many (e.g. in tutoring agents):

Chatbots: Example: Eliza

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

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

Chatbots: Example: Eliza

Men are all alike.

IN WHAT WAY

They're alw

CAN YOU THII

Well my bo

YOUR BOYFRI

He says I'm

I AM SORRY :

..

..

WHO ELSE IN

My father

YOUR FATHER

You are like

WHAT RESEMBI

You are not

WHAT MAKES ?

You don't a

WHY DO YOU ?

You are afri

DOES IT PLE

My father i:

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

- Rogerian psycho-therapy style
- rule-based: each rule linked to keyword in user response
- main contribution: transformations

function ELIZA GENERATOR(*user sentence*) **returns** *response*

Find the word *w* in *sentence* that has the highest keyword rank

if *w* exists

 Choose the highest ranked rule *r* for *w* that matches *sentence*

response \leftarrow Apply the transform in *r* to *sentence*

if *w* = 'my'

future \leftarrow Apply a transformation from the 'memory' rule list to *sentence*

 Push *future* onto memory stack

else (no keyword applies)

either

response \leftarrow Apply the transform for the NONE keyword to *sentence*

or

response \leftarrow Pop the top response from the memory stack

return(*response*)

Chatbots: Example: Eliza

- examples:
 - (* YOU * ME) [pattern] → (WHAT MAKES YOU THINK I 3 YOU) [transform]
 - *You hate me* → WHAT MAKES YOU THINK I HATE YOU
 - rule 1: (I *) [pattern] → (YOU SAY YOU 2) [transform]
 - I know everybody laughed at me → YOU SAY YOU KNOW EVERYBODY LAUGHED AT YOU
 - keyword *I* : very general → rank of rule low
 - rule 2: (I * everybody *) [pattern] → WHO IN PARTICULAR ARE YOU THINKING OF?
 - I know everybody laughed at me → WHO IN PARTICULAR ARE YOU THINKING OF?
 - although rule 1 may also be applied, keyword *everybody* : more specific → rank of this second rule higher
 - (MEMORY MY
 - (* MY * = LETS DISCUSS FURTHER WHY YOUR 3)
 - (* MY * = EARLIER YOU SAID YOUR 3)
 - (* MY * = DOES THAT HAVE ANYTHING TO DO WITH THE FACT THAT YOUR 3)

Properties of Human Conversation

- Turns & turn-taking patterns
(↔ endpoint detection)

C₁: ... I need to travel in May.
A₂: And, what day in May did you want to travel?
C₃: OK uh I need to be there for a meeting that's from the 12th to the 15th.
A₄: And you're flying into what city?
C₅: Seattle.
A₆: And what time would you like to leave Pittsburgh?
C₇: Uh hmm I don't think there's many options for non-stop.
A₈: Right. There's three non-stops today.
C₉: What are they?
A₁₀: 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.
C₁₁: OK I'll take the 5ish flight on the night before on the 11th.
A₁₂: On the 11th? OK. Departing at 5:55pm arrives Seattle at 8pm, U.S. Air
flight 115.
C₁₃: OK.
A₁₄: And you said returning on May 15th?
C₁₅: Uh, yeah, at the end of the day.
A₁₆: OK. There's #two non-stops ...#
C₁₇: #Act...actually #, what day of the week is the 15th?
A₁₈: It's a Friday.
C₁₉: Uh hmm. I would consider staying there an extra day til Sunday.
A₂₀: OK...OK. On Sunday I have ...

Figure 24.1 Part of a phone conversation between a human travel agent (A) and human client (C). The passages framed by # in A₁₆ and C₁₇ indicate overlaps in speech.

Properties of Human Conversation

- Turns & turn-taking patterns
(↔ endpoint detection)
- Speech acts:

C₁: ... I need to travel in May.
A₂: And, what day in May did you want to travel?
C₃: OK uh I need to be there for a meeting that's from the 12th to the 15th.
A₄: And you're flying into what city?
C₅: Seattle.
A₆: And what time would you like to leave Pittsburgh?
C₇: Uh hmm I don't think there's many options for non-stop.
A₈: Right. There's three non-stops today.
C₉: What are they?
A₁₀: The first one departs PGH at 10:00am arrives Seattle at 12:05 their time.

Constatives:	committing the speaker to something's being the case (<i>answering, claiming, confirming, denying, disagreeing, stating</i>)
Directives:	attempts by the speaker to get the addressee to do something (<i>advising, asking, forbidding, inviting, ordering, requesting</i>)
Commissives:	committing the speaker to some future course of action (<i>promising, planning, vowing, betting, opposing</i>)
Acknowledgments:	express the speaker's attitude regarding the hearer with respect to some social action (<i>apologizing, greeting, thanking, accepting an acknowledgment</i>)

A₂₀: OK...OK. On Sunday I have ...

Figure 24.1 Part of a phone conversation between a human travel agent (A) and human client (C). The passages framed by # in A₁₆ and C₁₇ indicate overlaps in speech.

Properties of Human Conversation

- Turns & turn-taking patterns (\leftrightarrow endpoint detection)
- Speech acts
- Grounding: acknowledge understanding

C₁: ... I need to travel in May.
A₂: And, what day in May did you want to travel?
C₃: OK uh I need to be there for a meeting that's from the 12th to the 15th.
A₄: And you're flying into what city?
C₅: Seattle.
A₆: And what time would you like to leave Pittsburgh?
C₇: Uh hmm I don't think there's many options for non-stop.
A₈: Right. There's three non-stops today.
C₉: What are they?
A₁₀: 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.
C₁₁: OK I'll take the 5ish flight on the night before on the 11th.
A₁₂: On the 11th? OK. Departing at 5:55pm arrives Seattle at 8pm, U.S. Air
flight 115.
C₁₃: OK.
A₁₄: And you said returning on May 15th?
C₁₅: Uh, yeah, at the end of the day.
A₁₆: OK. There's #two non-stops ...#
C₁₇: #Act...actually #, what day of the week is the 15th?
A₁₈: It's a Friday.
C₁₉: Uh hmm. I would consider staying there an extra day til Sunday.
A₂₀: OK...OK. On Sunday I have ...

Figure 24.1 Part of a phone conversation between a human travel agent (A) and human client (C). The passages framed by # in A₁₆ and C₁₇ indicate overlaps in speech.

Properties of Human Conversation

- Turns & turn-taking patterns
(↔ endpoint detection)
- Speech acts
- Grounding: acknowledge understanding
- Subdialogues and Dialogue Structure:

Conversational analysis:
adjacency pairs:

QUESTION - ANSWER

PROPOSAL – ACCEPTANCE / REJECTION

COMPLIMENT - DOWNPLAYER

C₁: ... I need to travel in May.
A₂: And, what day in May did you want to travel?
C₃: OK uh I need to be there for a meeting that's from the 12th to the 15th.
A₄: And you're flying into what city?
C₅: Seattle.
A₆: And what time would you like to leave Pittsburgh?
C₇: Uh hmm I don't think there's many options for non-stop.
A₈: Right. There's three non-stops today.
C₉: What are they?
A₁₀: 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.
C₁₁: OK I'll take the 5ish flight on the night before on the 11th.
A₁₂: On the 11th? OK. Departing at 5:55pm arrives Seattle at 8pm, U.S. Air
flight 115.
C₁₃: OK.
A₁₄: And you said returning on May 15th?
C₁₅: Uh, yeah, at the end of the day.
A₁₆: OK. There's #two non-stops ... #
C₁₇: #Act... actually #, what day of the week is the 15th?
A₁₈: It's a Friday.
C₁₉: Uh hmm. I would consider staying there an extra day til Sunday.
A₂₀: OK...OK. On Sunday I have ...

Figure 24.1 Part of a phone conversation between a human travel agent (A) and human client (C). The passages framed by # in A₁₆ and C₁₇ indicate overlaps in speech.

Properties of Human Conversation

- Turns & turn-taking patterns (\leftrightarrow endpoint detection)
- Speech acts
- Grounding: acknowledge understanding
- Subdialogues and Dialogue Structure:
subdialogue

C₁: ... I need to travel in May.
A₂: And, what day in May did you want to travel?
C₃: OK uh I need to be there for a meeting that's from the 12th to the 15th.
A₄: And you're flying into what city?
C₅: Seattle.
A₆: And what time would you like to leave Pittsburgh?
C₇: Uh hmm I don't think there's many options for non-stop.
A₈: Right. There's three non-stops today.
C₉: What are they?
A₁₀: 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.
C₁₁: OK I'll take the 5ish flight on the night before on the 11th.
A₁₂: On the 11th? OK. Departing at 5:55pm arrives Seattle at 8pm, U.S. Air
flight 115.
C₁₃: OK.
A₁₄: And you said returning on May 15th?
C₁₅: Uh, yeah, at the end of the day.
A₁₆: OK. There's #two non-stops ... #
C₁₇: #Act... actually #, what day of the week is the 15th?
A₁₈: It's a Friday.
C₁₉: Uh hmm. I would consider staying there an extra day til Sunday.
A₂₀: OK...OK. On Sunday I have ...

Figure 24.1 Part of a phone conversation between a human travel agent (A) and human client (C). The passages framed by # in A₁₆ and C₁₇ indicate overlaps in speech.

Properties of Human Conversation

- Turns & turn-taking patterns (\leftrightarrow endpoint detection)
- Speech acts
- Grounding: acknowledge understanding
- Subdialogues and Dialogue Structure:
clarification question:

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...

C₁: ... I need to travel in May.
A₂: And, what day in May did you want to travel?
C₃: OK uh I need to be there for a meeting that's from the 12th to the 15th.
A₄: And you're flying into what city?
C₅: Seattle.
A₆: And what time would you like to leave Pittsburgh?
C₇: Uh hmm I don't think there's many options for non-stop.
A₈: Right. There's three non-stops today.
C₉: What are they?
A₁₀: 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.
C₁₁: OK I'll take the 5ish flight on the night before on the 11th.
A₁₂: On the 11th? OK. Departing at 5:55pm arrives Seattle at 8pm, U.S. Air
flight 115.
C₁₃: OK.
A₁₄: And you said returning on May 15th?
C₁₅: Uh, yeah, at the end of the day.
A₁₆: OK. There's #two non-stops ... #
C₁₇: #Act... actually #, what day of the week is the 15th?
A₁₈: It's a Friday.
C₁₉: Uh hmm. I would consider staying there an extra day til Sunday.
A₂₀: OK...OK. On Sunday I have ...

Figure 24.1 Part of a phone conversation between a human travel agent (A) and human client (C). The passages framed by # in A₁₆ and C₁₇ indicate overlaps in speech.

Properties of Human Conversation

- Turns & turn-taking patterns (\leftrightarrow endpoint detection)
- Speech acts
- Grounding: acknowledge understanding
- Subdialogues and Dialogue Structure:
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.

C₁: ... I need to travel in May.
A₂: And, what day in May did you want to travel?
C₃: OK uh I need to be there for a meeting that's from the 12th to the 15th.
A₄: And you're flying into what city?
C₅: Seattle.
A₆: And what time would you like to leave Pittsburgh?
C₇: Uh hmm I don't think there's many options for non-stop.
A₈: Right. There's three non-stops today.
C₉: What are they?
A₁₀: 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.
C₁₁: OK I'll take the 5ish flight on the night before on the 11th.
A₁₂: On the 11th? OK. Departing at 5:55pm arrives Seattle at 8pm, U.S. Air
flight 115.
C₁₃: OK.
A₁₄: And you said returning on May 15th?
C₁₅: Uh, yeah, at the end of the day.
A₁₆: OK. There's #two non-stops ... #
C₁₇: #Act... actually #, what day of the week is the 15th?
A₁₈: It's a Friday.
C₁₉: Uh hmm. I would consider staying there an extra day til Sunday.
A₂₀: OK...OK. On Sunday I have ...

Figure 24.1 Part of a phone conversation between a human travel agent (A) and human client (C). The passages framed by # in A₁₆ and C₁₇ indicate overlaps in speech.

Properties of Human Conversation

- Turns & turn-taking patterns (\leftrightarrow endpoint detection)
- Speech acts
- Grounding: acknowledge understanding
- Subdialogues and Dialogue Structure
- Initiative

C₁: ... I need to travel in May.
A₂: And, what day in May did you want to travel?
C₃: OK uh I need to be there for a meeting that's from the 12th to the 15th.
A₄: And you're flying into what city?
C₅: Seattle.
A₆: And what time would you like to leave Pittsburgh?
C₇: Uh hmm I don't think there's many options for non-stop.
A₈: Right. There's three non-stops today.
C₉: What are they?
A₁₀: 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.
C₁₁: OK I'll take the 5ish flight on the night before on the 11th.
A₁₂: On the 11th? OK. Departing at 5:55pm arrives Seattle at 8pm, U.S. Air flight 115.
C₁₃: OK.
A₁₄: And you said returning on May 15th?
C₁₅: Uh, yeah, at the end of the day.
A₁₆: OK. There's #two non-stops ... #
C₁₇: #Act... actually #, what day of the week is the 15th?
A₁₈: It's a Friday.
C₁₉: Uh hmm. I would consider staying there an extra day til Sunday.
A₂₀: OK...OK. On Sunday I have ...

Figure 24.1 Part of a phone conversation between a human travel agent (A) and human client (C). The passages framed by # in A₁₆ and C₁₇ indicate overlaps in speech.

Properties of Human Conversation

- Turns & turn-taking patterns
(↔ endpoint detection)

- Speech acts

- Grounding: acknowledge understanding

- Subdialogues and Dialogue Structure

- Initiative

- Inference and Implications

C₁: ... I need to travel in May.
A₂: And, what day in May did you want to travel?
C₃: OK uh I need to be there for a meeting that's from the 12th to the 15th
A₄: And you're flying into what city?
C₅: Seattle.
A₆: And what time would you like to leave Pittsburgh?
C₇: Uh hmm I don't think there's many options for non-stop.
A₈: Right. There's three non-stops today.
C₉: What are they?
A₁₀: 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.
C₁₁: OK I'll take the 5ish flight on the night before on the 11th.
A₁₂: On the 11th? OK. Departing at 5:55pm arrives Seattle at 8pm, U.S. Air
flight 115.
C₁₃: OK.
A₁₄: And you said returning on May 15th?
C₁₅: Uh, yeah, at the end of the day.
A₁₆: OK. There's #two non-stops ...#
C₁₇: #Act...actually #, what day of the week is the 15th?
A₁₈: It's a Friday.
C₁₉: Uh hmm. I would consider staying there an extra day til Sunday.
A₂₀: OK...OK. On Sunday I have ...

Figure 24.1 Part of a phone conversation between a human travel agent (A) and human client (C). The passages framed by # in A₁₆ and C₁₇ indicate overlaps in speech.

Chatbots

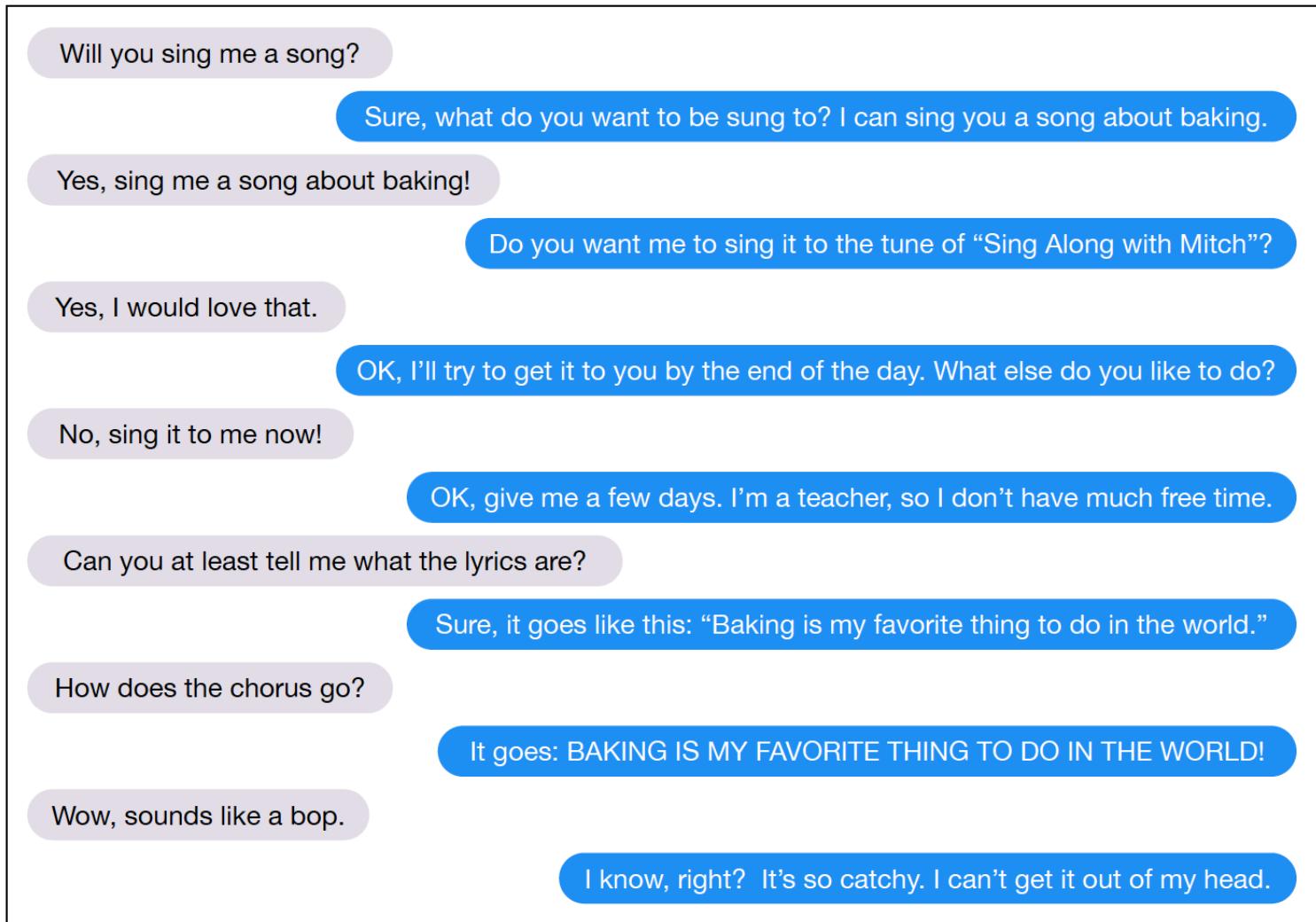


Figure 24.2 A conversation between a user (left) and BlenderBot (Roller et al., 2021).

Corpus-Based Chatbots

- Response by **retrieval**

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

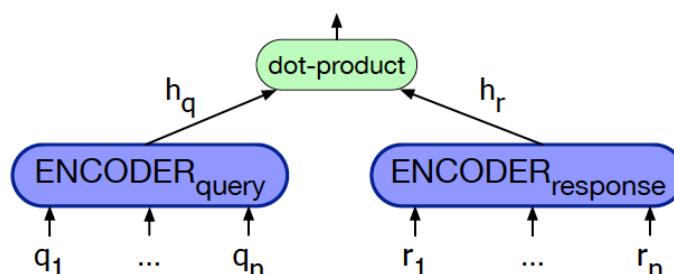
$$h_q = \text{BERT}_Q(q)[\text{CLS}]$$

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

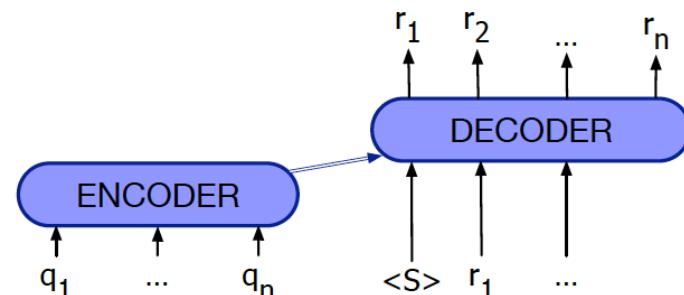
$$\text{response}(q, C) = \operatorname{argmax}_{r \in C} h_q \cdot h_r$$

- Response by **generation**

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



(a) Response by Retrieval



(b) Response by Generation

Corpus-Based Chatbots

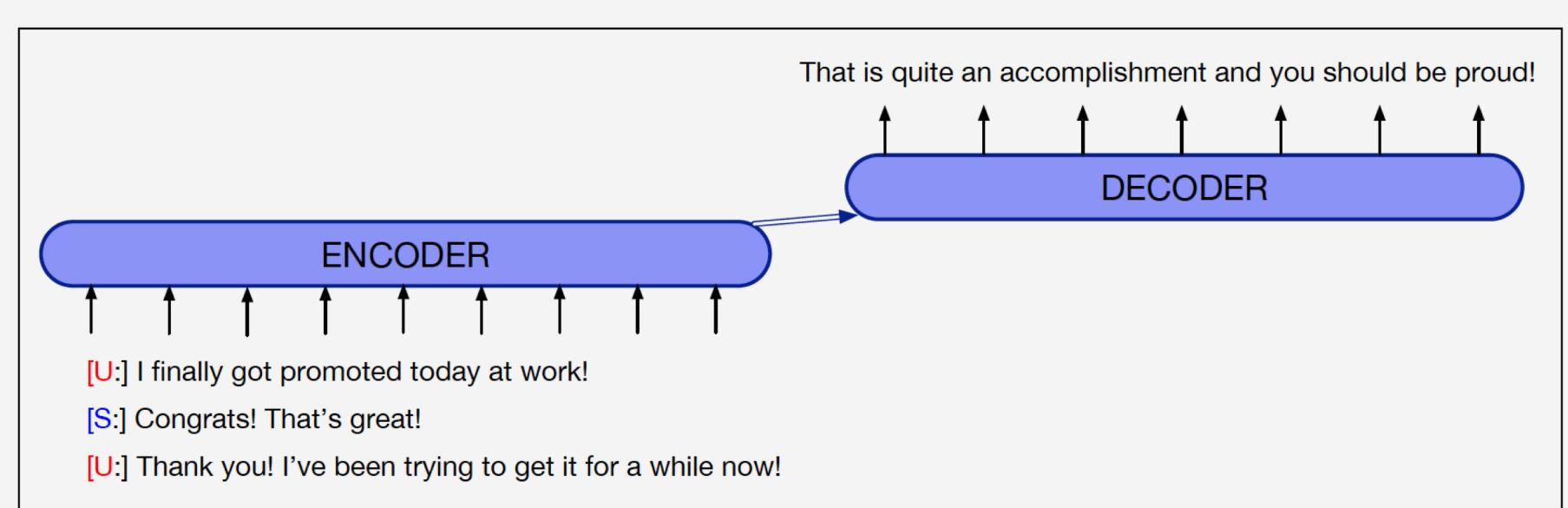
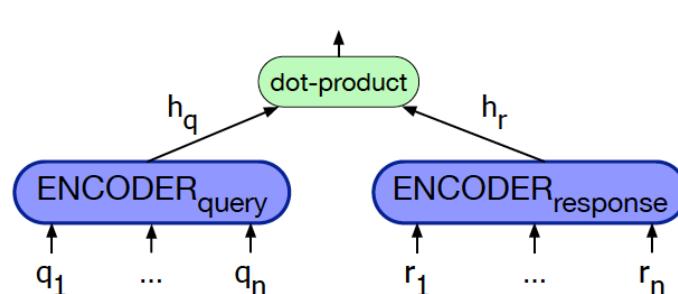
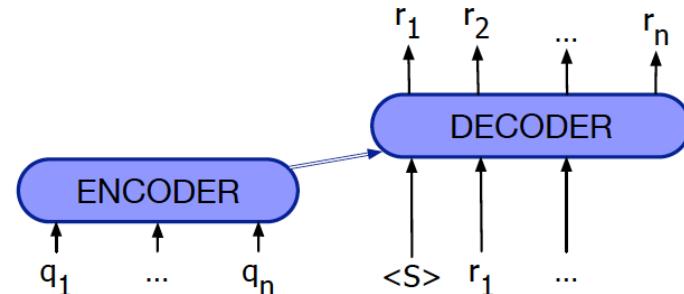


Figure 24.7 Example of encoder-decoder for dialogue response generation; the encoder sees the entire dialogue context.



(a) Response by Retrieval



(b) Response by Generation

Hybrid Architectures

- Examples: Alexa, Siri: **combine** neural approaches (e.g. involving GPT), entity linking, dialogue classification, slot filling etc.

#	Utterance	Entity	Commentary
1	User: let's chat Bot: 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 <i>FORCE_START</i> response to begin the conversation.
2	User: my name is chris Bot: 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 <i>STRONG_CONTINUE</i> response that needs a prompt, which is supplied by Neural Chat with <i>FORCE_START</i> priority (scripted transition).
3	User: hang out with my friends Bot: 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 <i>STRONG_CONTINUE</i> priority.
4	User: maybe watch a movie Bot: 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 <i>FORCE_START</i> response (overriding Neural Chat), setting current entity to <i>Film</i> .
5	User: i saw the matrix Bot: Nice! Did you like The Matrix? [Movies]	The Matrix	The Entity Linker and Entity Tracker identify <i>The Matrix</i> as the current entity.
6	User: i loved it neo is amazing Bot: 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 <i>Keanu Reeves</i> .
7	User: i want to talk about something else	None	Navigational Intent detects <i>negative intent</i> , so Entity Tracker rejects the current entity.

Figure 24.8 Part of a conversation with Chirpy Cardinal (Paranjape et al., 2020).

Simple Frame-Based Dialogue Systems

- **Task-based dialogue:** typical case. **GUS** architecture (1977): dialogue state (state machine), frames, slots + values, domain ontology

GUS: Hello. My name is GUS. I can help you plan a simple trip. 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 102 on Wednesday May 28 from San Jose to San Diego that leaves at 7.30 am and arrives at 9.15 am. I will call you when you get there. Goodbye

Types in GUS, as in modern frame-based dialogue agents, have hierarchical structure; for example the *date* type in GUS is itself a frame with slots with types like *integer* or members of sets of weekday names:

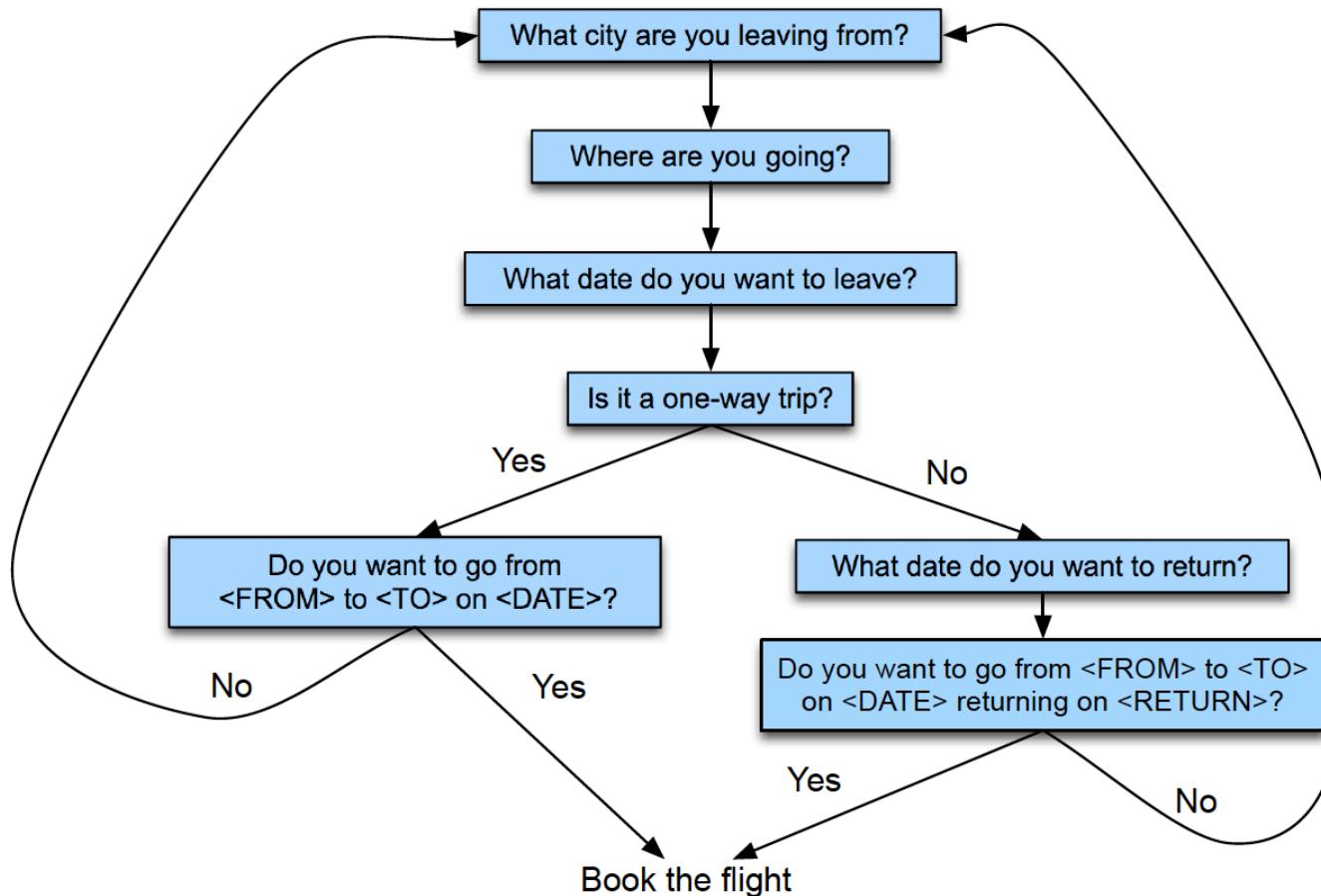
DATE
MONTH:NAME YEAR:INTEGER DAY:(BOUNDED-INTEGER 1 31)
WEEKDAY:(MEMBER (Sunday Monday Tuesday Wednesday))

Slot	Type	Question Template
ORIGIN CITY	city	"From what city are you leaving?"
DESTINATION CITY	city	"Where are you going?"
DEPARTURE TIME	time	"When would you like to leave?"
DEPARTURE DATE	date	"What day would you like to leave?"
ARRIVAL TIME	time	"When do you want to arrive?"
ARRIVAL DATE	date	"What day would you like to arrive?"

Figure 24.10 A frame in a frame-based dialogue system, showing the type of each slot and a question used to fill the slot.

Simple Frame-Based Dialogue Systems

- GUS (also Siri, Alexa etc.): control structure \leftrightarrow slot filling in frames
- GUS: production rules determine state transitions



Simple Frame-Based Dialogue Systems

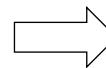
- Determining **Domain, Intent, Slot-Filling** from user utterance:
 - **domain:** what is the user talking about? (→ classifier)
 - **intent:** what is the task or goal of the user? (→ classifier)
 - **slot filling:** (→ hand-written rules)

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

Wake me tomorrow at 6



DOMAIN: ALARM-CLOCK
INTENT: SET-ALARM
TIME: 2017-07-01 0600-0800

Simple Frame-Based Dialogue Systems

- hand-written **rules** for slot filling

- Reg-ex for SET_ALARM intent:

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

- Semantic grammar (Phoenix system, 1994)

SHOW	→ show me i want can i see ...
DEPART_TIME_RANGE	→ (after around before) HOUR morning afternoon evening
HOUR	→ one two three four... twelve (AMPM)
FLIGHTS	→ (a) flight flights
AMPM	→ am pm
ORIGIN	→ from CITY
DESTINATION	→ to CITY
CITY	→ Boston San Francisco Denver Washington

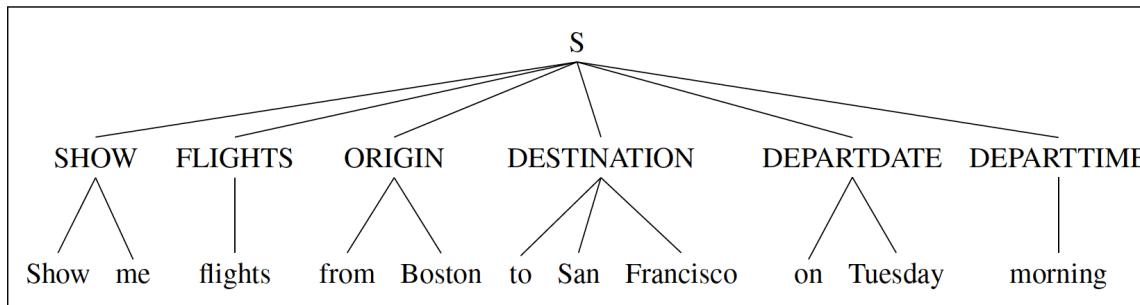


Figure 24.11 A semantic grammar parse for a user sentence, using slot names as the internal parse tree nodes.

Simple Frame-Based Dialogue Systems

- Other components

- make components (e.g. speech to text, output language model) dependent on dialogue state (e.g. use special restricted grammar)

- simple systems: system utterances: fixed hand-written templates

What time do you want to leave CITY-ORIG?

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

- use grounding:

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

Caller: No.

System: What's next?

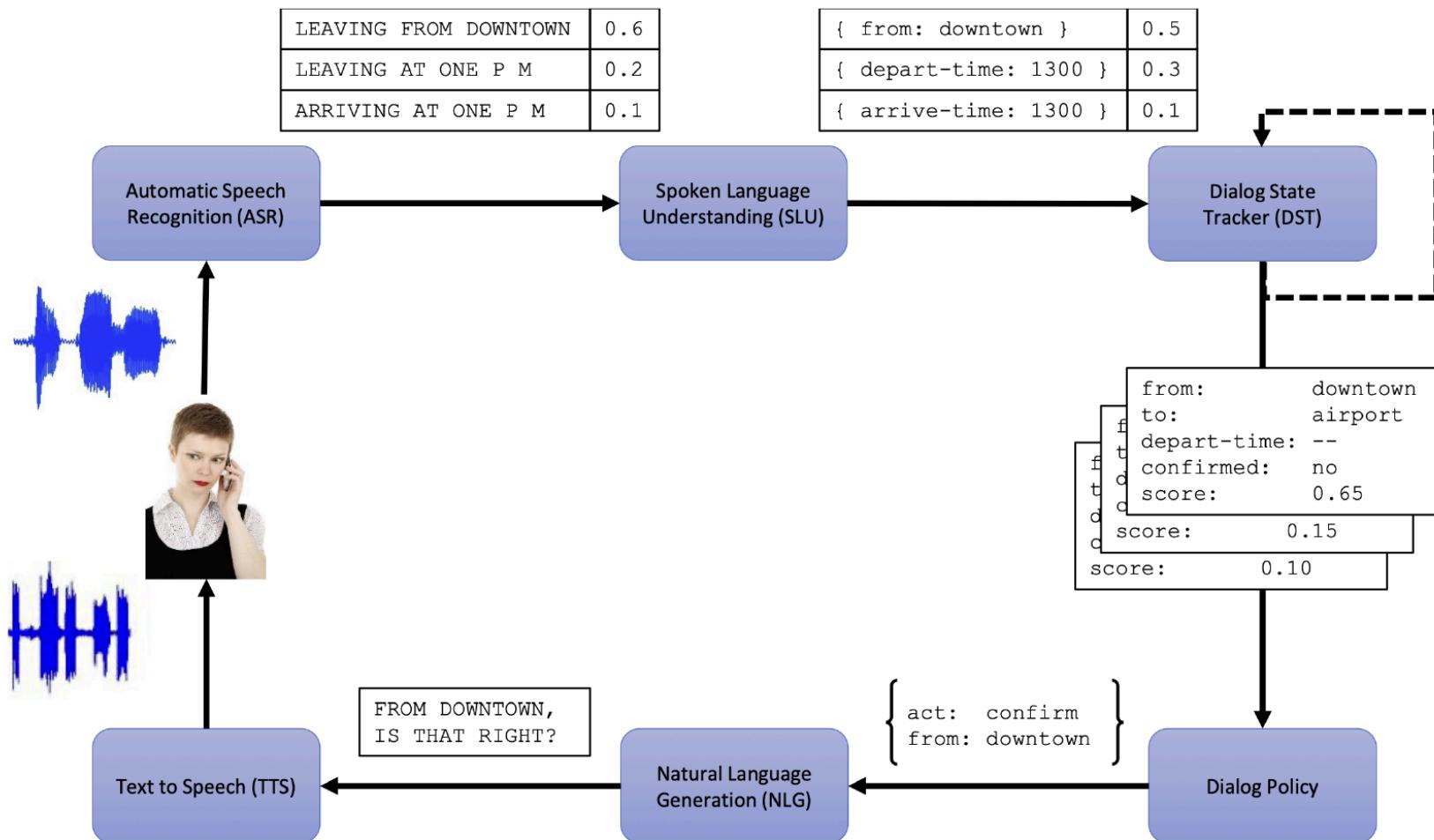


System: Did you want to review some more of your personal profile?

Caller: No.

System: *Okay*, what's next?

Task Oriented Dialogue Systems: Dialogue State Architecture



Dialogue State Architecture: Dialogue Acts

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

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()

NN Slot Filling – NN Sequence Labeling Task

I want to fly to San Francisco on Monday afternoon please

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

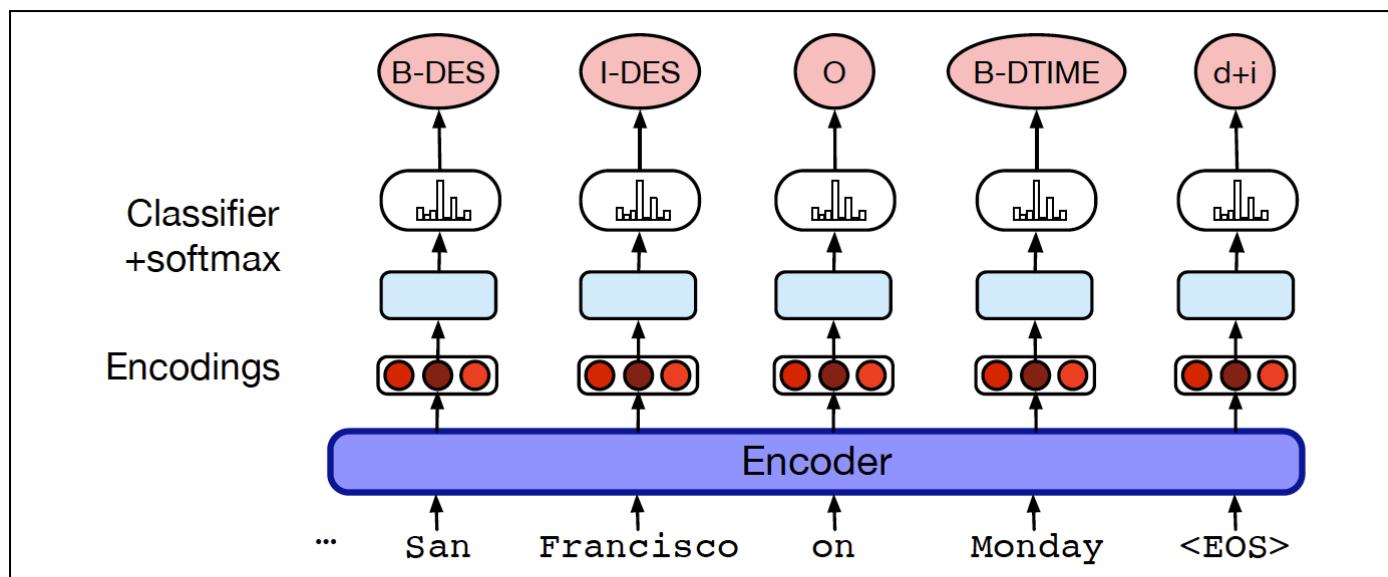


Figure 24.15 A simple architecture for slot filling, mapping the words in the input through contextual embeddings like BERT to an output classifier layer (which can be linear or something more complex), followed by softmax to generate a series of BIO tags (and including a final state consisting of a domain concatenated with an intent).

Dialogue State Tracking

- Dialogue state tracker: determine current filling status and state of frame

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

- State tracking and slot filling: can be done jointly
- Special case: detecting (user) correction acts

features	examples
lexical	words like “no”, “correction”, “I don’t”, swear words, utterance length
semantic	similarity (word overlap or embedding dot product) between the candidate correction act and the user’s prior utterance
phonetic	phonetic overlap between the candidate correction act and the user’s prior utterance (i.e. “WhatsApp” may be incorrectly recognized as “What’s up”)
prosodic	hyperarticulation, increases in F0 range, pause duration, and word duration, generally normalized by the values for previous sentences
ASR	ASR confidence, language model probability

Dialogue Policy

- Decide what dialogue act to generate next

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

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

possible approach: reinforcement learning

- Example policy : confirmation and rejection

o explicit

confirmation:

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

Dialogue Policy

- implicit confirmation:

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?

- rejection:
progressive prompting:

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.

- rejection:
rapid re-prompting: system rejects an utterance just by saying “I’m sorry?” or “What was that?”

- confidence based:

$< \alpha$	low confidence	reject
$\geq \alpha$	above the threshold	confirm explicitly
$\geq \beta$	high confidence	confirm implicitly
$\geq \gamma$	very high confidence	don't confirm at all

Dialogue State Model: Language Generation

- Content planning, sentence realisation:

```
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.
```

Figure 24.16 Two examples of inputs to the sentence realization phase of NLG, showing the dialogue act and attributes prespecified by the content planner. Line 1-2 and 3-4 show different possible output sentences to be generated by the sentence realizer. From the restaurant recommendation system of [Nayak et al. \(2017\)](#).

```
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.
```

Figure 24.17 Delexicalized sentences that can be used for generating many different relexicalized sentences. From the restaurant recommendation system of [Nayak et al. \(2017\)](#).

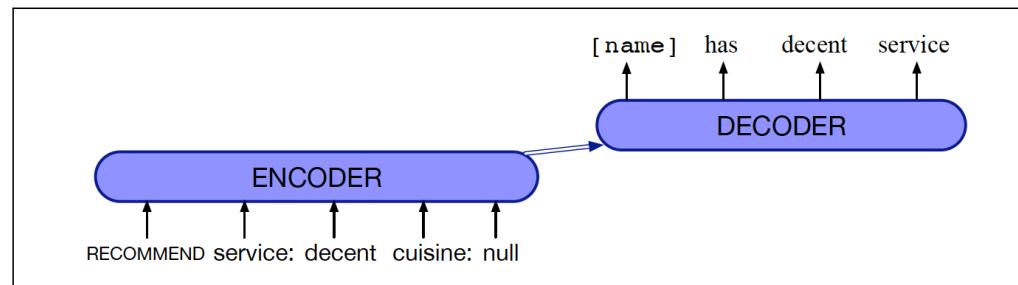


Figure 24.18 An encoder-decoder sentence realizer mapping slots/fillers to English.

Dialogue System Evaluation

<p>Hi how are you today</p> <p>Hi, I'm pretty good! Just listening to some aerosmith, they're my fave :) what're you up to?</p> <p>I am listening to some italian music</p> <p>Italian music, nice! What do you do for work?</p> <p>I'm a professional wrestler</p> <p>Oh wow! I'm a construction worker, I've been doing it since I was 16.</p> <p>What do you do for a living</p> <p>I'm a construction worker haha, just told you! What's your favorite type of food?</p> <p>My favorite is garlic bread</p> <p>Yum! I love mexican food, so good :) do you have any hobbies?</p> <p>I like to workout and go to the gym</p> <p>We're a bit different- I love watching nascar and ufc. They're so fun!</p> <p>Who would you prefer to talk to for a long conversation?</p> <p><input type="radio"/> I would prefer to talk to Speaker 1 <input type="radio"/> I would prefer to talk to Speaker 2</p> <p>Please provide a brief justification for your choice (a few words or a sentence)</p> <p>Please enter here...</p>	<p>Hello there, how are you?</p> <p>I am doing great. How are you?</p> <p>I am great, I did something crazy for me and colored my hair blue!</p> <p>I have a daughter and a son who also love blue colored balls. You should meet them</p> <p>Well that's neat, I got a new car my mother gave so maybe I could see them!</p> <p>It is a beautiful city. And, I try to be... Just cannot afford a bigger house atm.</p> <p>I am sorry to hear that, I feel bad going out of town for spring break now.</p> <p>Ok. I'm going to school in the spring for casino manager</p> <p>Well I turn 29 next week, I wonder if that is a good age to apply as one.</p> <p>My grandmother just died from lung cancer, sucks</p>
---	---

Figure 24.19 The ACUTE-EVAL method asks annotators to compare two dialogues and choose between Speaker 1 (light blue) and Speaker 2 (dark blue), independent of the gray speaker. Figure from Li et al. (2019a).

Dialogue System Evaluation: Task-Based Dialogue Systems

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

Figure 24.20 User satisfaction survey, adapted from [Walker et al. \(2001\)](#).

$$\text{Slot Error Rate for a Sentence} = \frac{\text{\# of inserted/deleted/substituted slots}}{\text{\# of total reference slots for sentence}}$$

Bibliography

- (1) Dan Jurafsky and James Martin: Speech and Language Processing (3rd ed. draft, version Jan 2022); Online: <https://web.stanford.edu/~jurafsky/slp3/> (URL, Oct 2022)

Recommendations for Studying

- **minimal approach:**
work with the slides and understand their contents! Think beyond instead of merely memorizing the contents
- **standard approach:**
minimal approach + read the corresponding pages in Jurafsky [1]
- **interested students**
== standard approach