Carnegie Mellon University
Language Technologies Institute

CS11-711 Advanced NLP

Dialog

Shikib Mehri

What is dialog?

NLP for **conversations**

- **Understanding** utterances, in the context of a conversation
- **Generating** responses
  - That are **consistent** and **coherent** with the dialog history
  - That are **interesting** and **engaging**
  - That meaningfully progress the dialog **towards a goal**

What is NEW in dialog?

- The dialog history
  - Consider an utterance in the **context of the dialog history**
  - Must understand **back-and-forth** interaction, estimate **belief state** over user goals, resolve **co-references**, etc.
- The goal/purpose of the conversation:
  - **Task-oriented:** progress dialog towards some goal/complete a task
  - **Open-domain:** produce interesting/engaging responses

What is NEW in dialog?

- **New challenges:** modelling dialog context, goal-driven nature, task-specific terminology, one-to-many response problem in open-domain
- **New tasks:** response generation (task-oriented vs open-domain), NLU (intent prediction, slot filling, state tracking), NLG (structured data → response generation), dialog evaluation
- **New models**

Objective

My objective in this lecture is for you to walk away with:

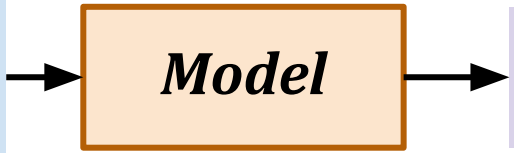
- Knowledge of the **unique challenges** in dialog research
- Knowledge of the **tasks/datasets** that are commonly studied in dialog
- Knowledge of the **common models/strategies** used in dialog

Outline

- Response generation
  - Task-oriented
  - Open-domain
- NLU (intent prediction, slot filling, state tracking)
- Dialog evaluation

Response Generation

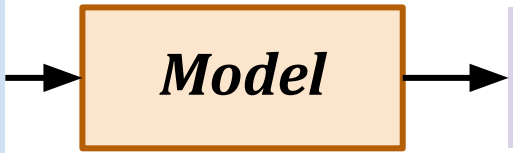
User: *Hello*
System: *Hello, how are you today?*
User: *I'm fine. I'm watching Netflix*
System: *What are you watching? I just finished Squid Game.*
User: *I'm watching the new season of You. It's really good!*



System: *What's it about? I need something new to watch.*

Response Generation

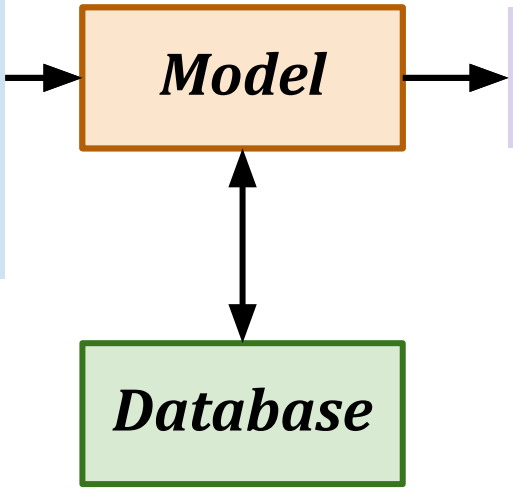
User: *Hello*
System: *Hello, how may I help you?*
User: *Find me a good italian restaurant*
System: *Okay, what neighbourhood are you in?*
User: *Oakland*



System: *Okay, I found Lucca, an Italian restaurant in Oakland.*

Response Generation

User: *Hello*
System: *Hello, how may I help you?*
User: *Find me a good italian restaurant*
System: *Okay, what neighbourhood are you in?*
User: *Oakland*



System: *Okay, I found Lucca, an Italian restaurant in Oakland.*

Task-Oriented Response Generation

Task-oriented dialog systems interact with a user in order to complete a specific task.

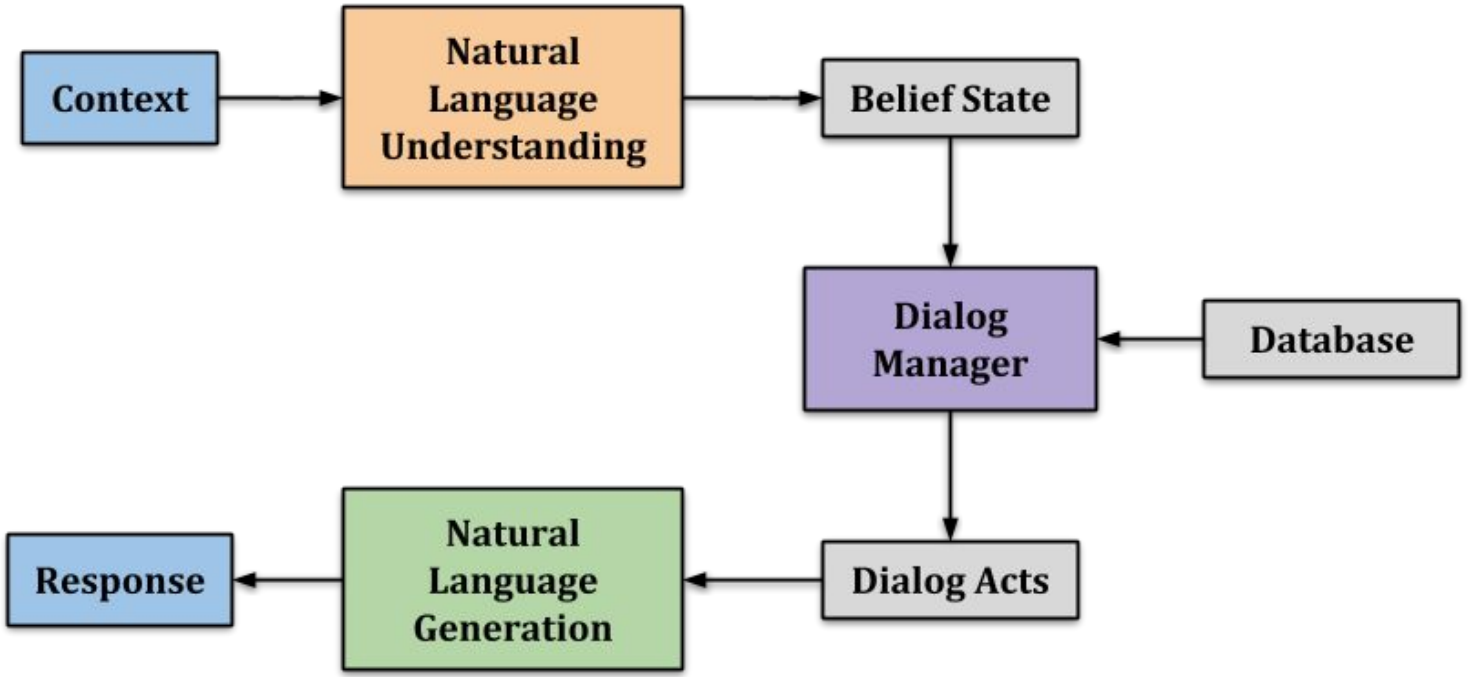
- MultiWOZ [Budzianowski et al. 2018]
- SGD [Rastogi et al. 2019]
- STAR [Mosig et al. 2020]
- Taskmaster-2 [Byrne et al. 2020]
- ABCD [Chen et al. 2021]

Task-Oriented Response Generation

Task-oriented dialog systems interact with a user in order to complete a specific task.

- Must understand the **dialog context**
- Must track **belief state** over dialog context
- Often need to interpret **structured database output**
- Must follow task-specific **dialog policy**
- Must generate **natural language** responses

Pipeline Dialog System

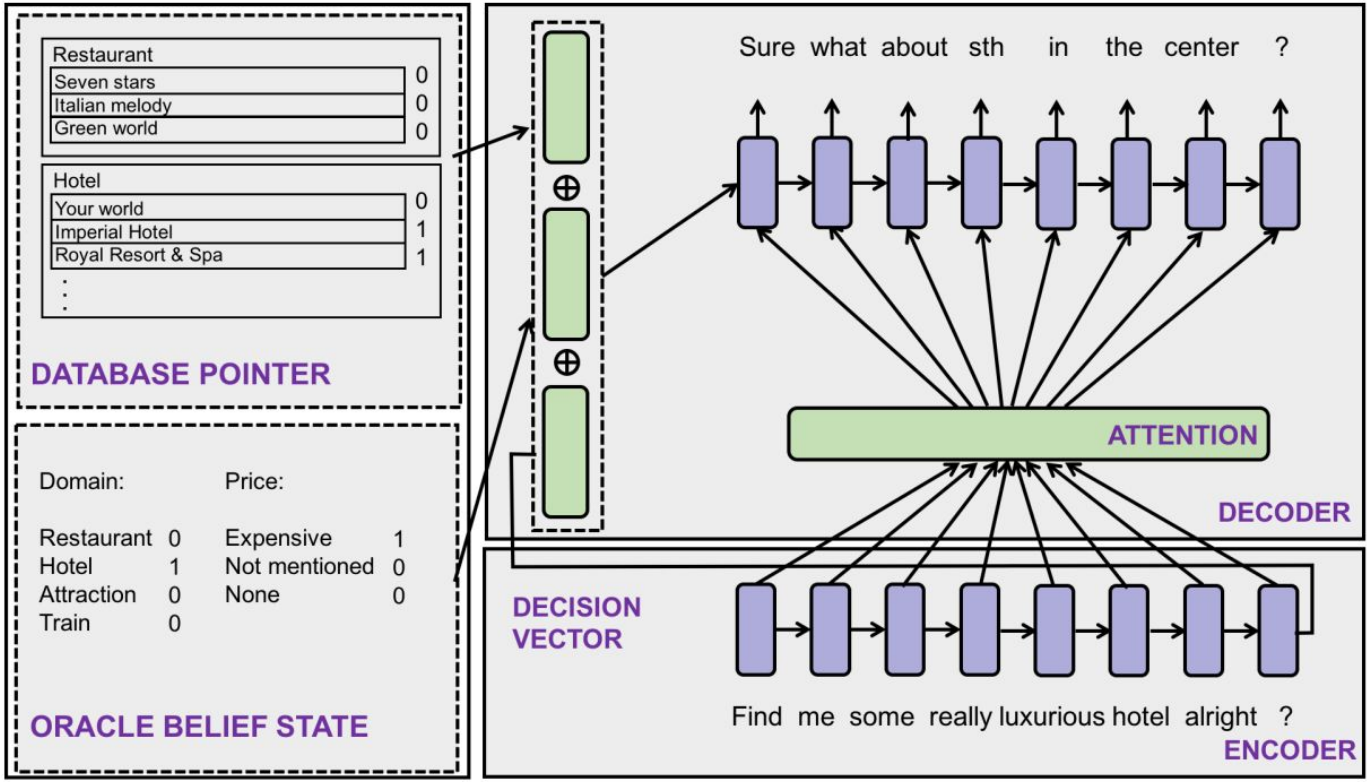


Task-Oriented Response Generation

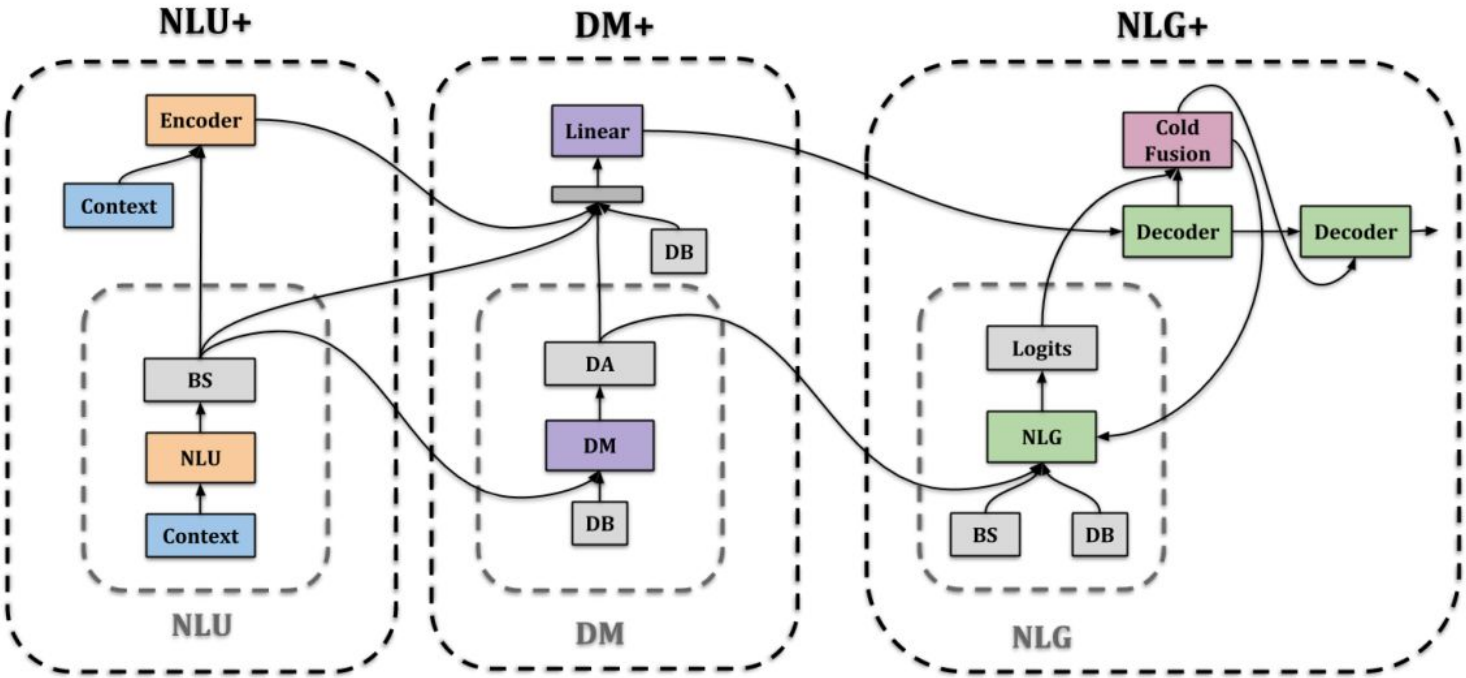
Task-oriented dialog systems interact with a user in order to complete a specific task.

- Must understand the **dialog context**
- Must track **belief state** over dialog context
- Often need to interpret **structured database output**
- Must follow task-specific **dialog policy**
- Must generate **natural language** responses

Seq2Seq with Attention [Budzianowski et al. 2018]

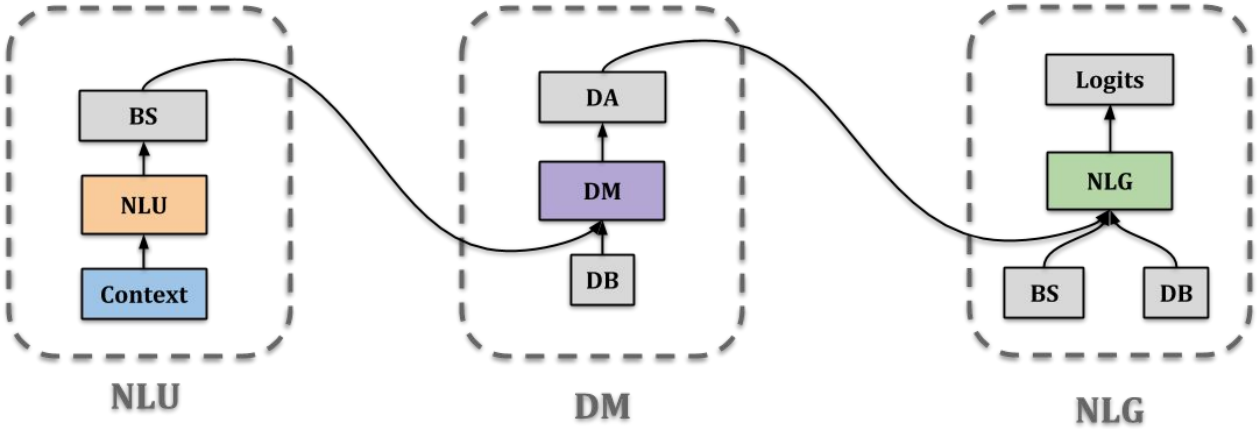


Structured Fusion Networks [Mehri et al. 2019]

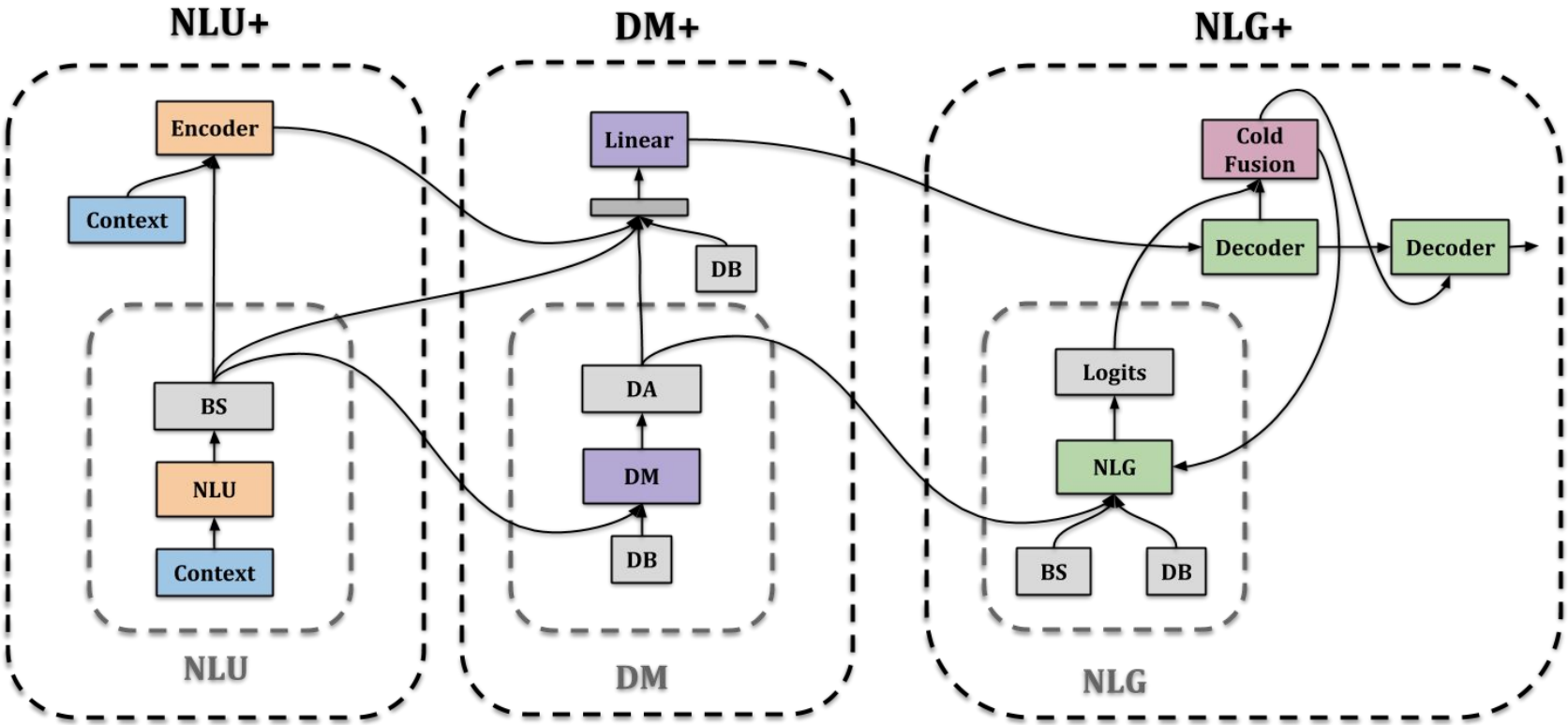


Dialog Modules

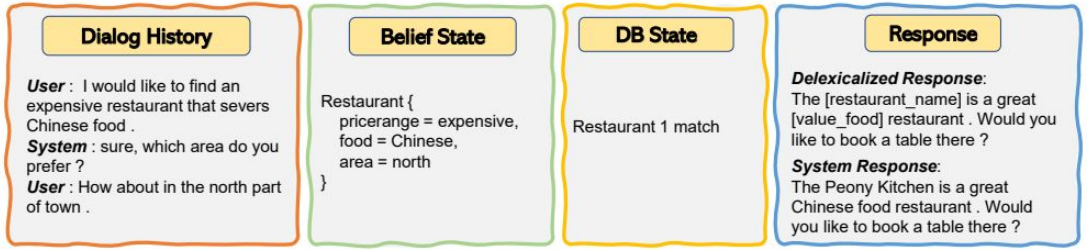
Start with **pre-trained** neural dialog modules



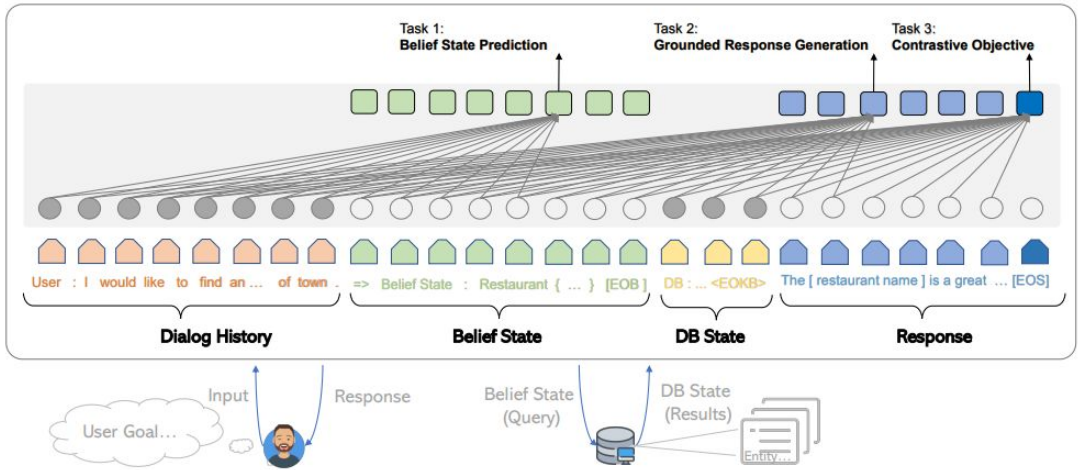
Structured Fusion Networks



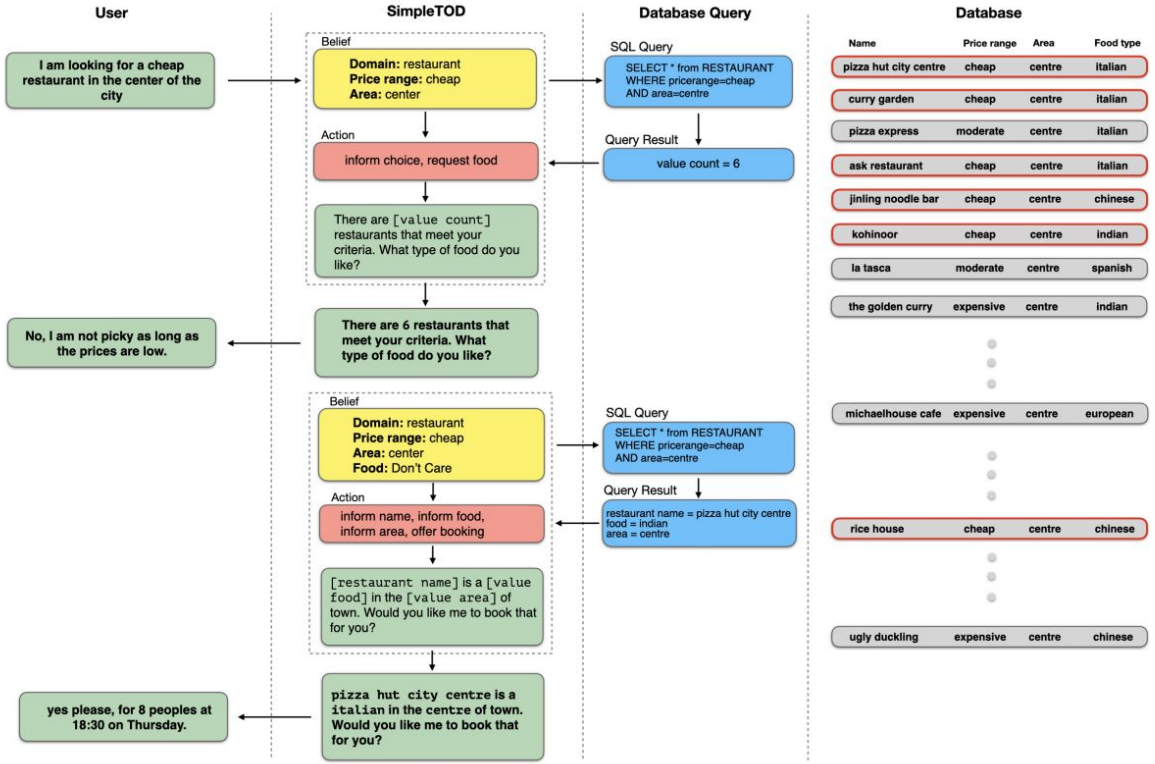
SOLOIST [Peng et al. 2020]



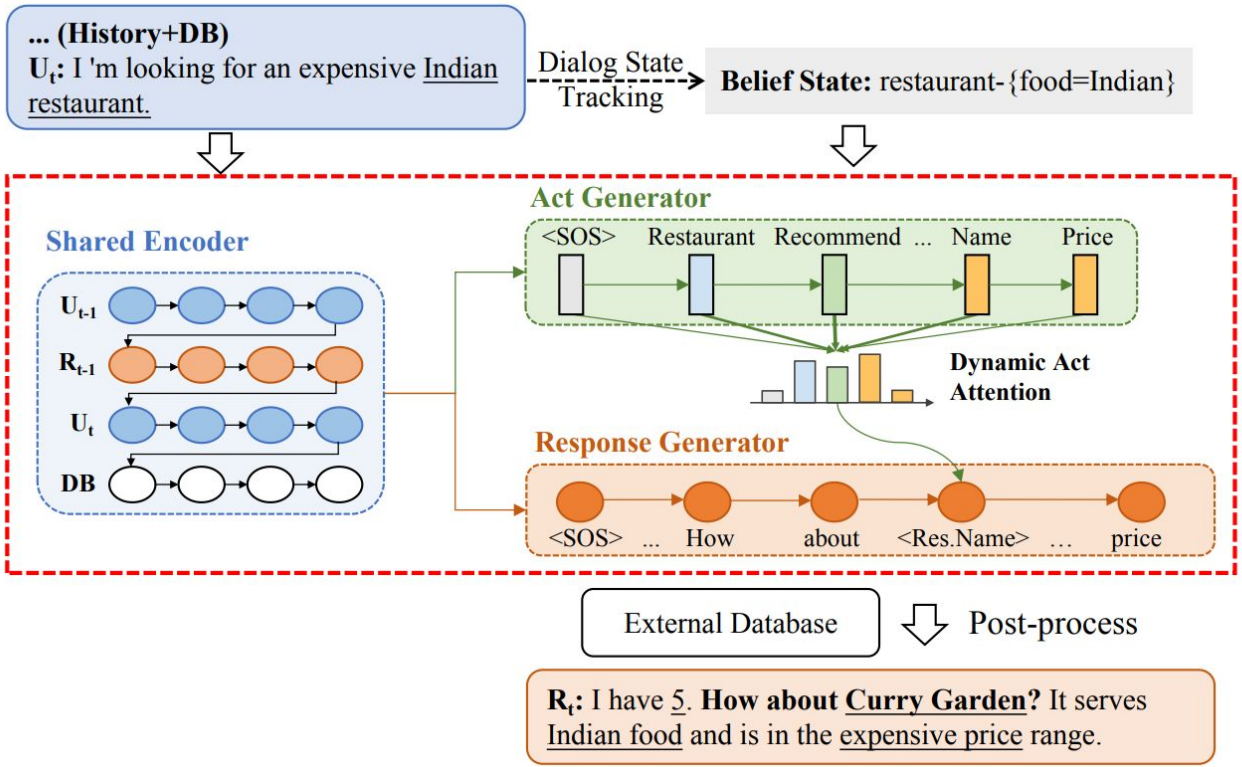
(b) Example snippets for the items compounding the input of SOLOIST model.



SimpleTOD [Hosseini-Asl et al. 2020]



MarCo [Wang et al. 2020]



Open-Domain Response Generation

Open-domain dialog systems must engage in **chit-chat** with a user

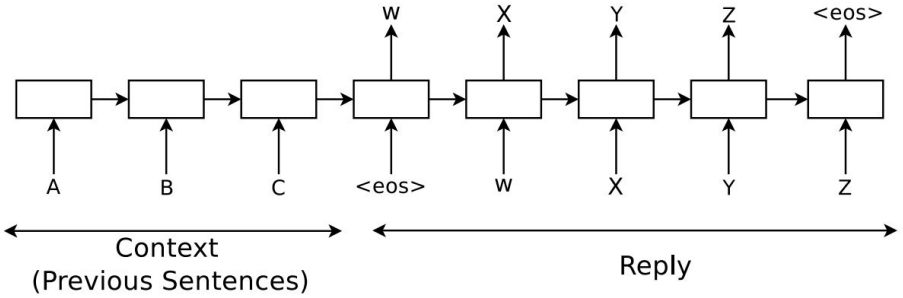
- DailyDialog [Li et al. 2017]
- PersonaChat [Zhang et al. 2018]
- Topical-Chat [Gopalakrishnan et al. 2019]
- Wizards of Wikipedia [Dinan et al. 2018]
- Empathetic Dialogs [Rashkin et al. 2019]

Open-Domain Response Generation

Open-domain dialog systems must engage in **chit-chat** with a user

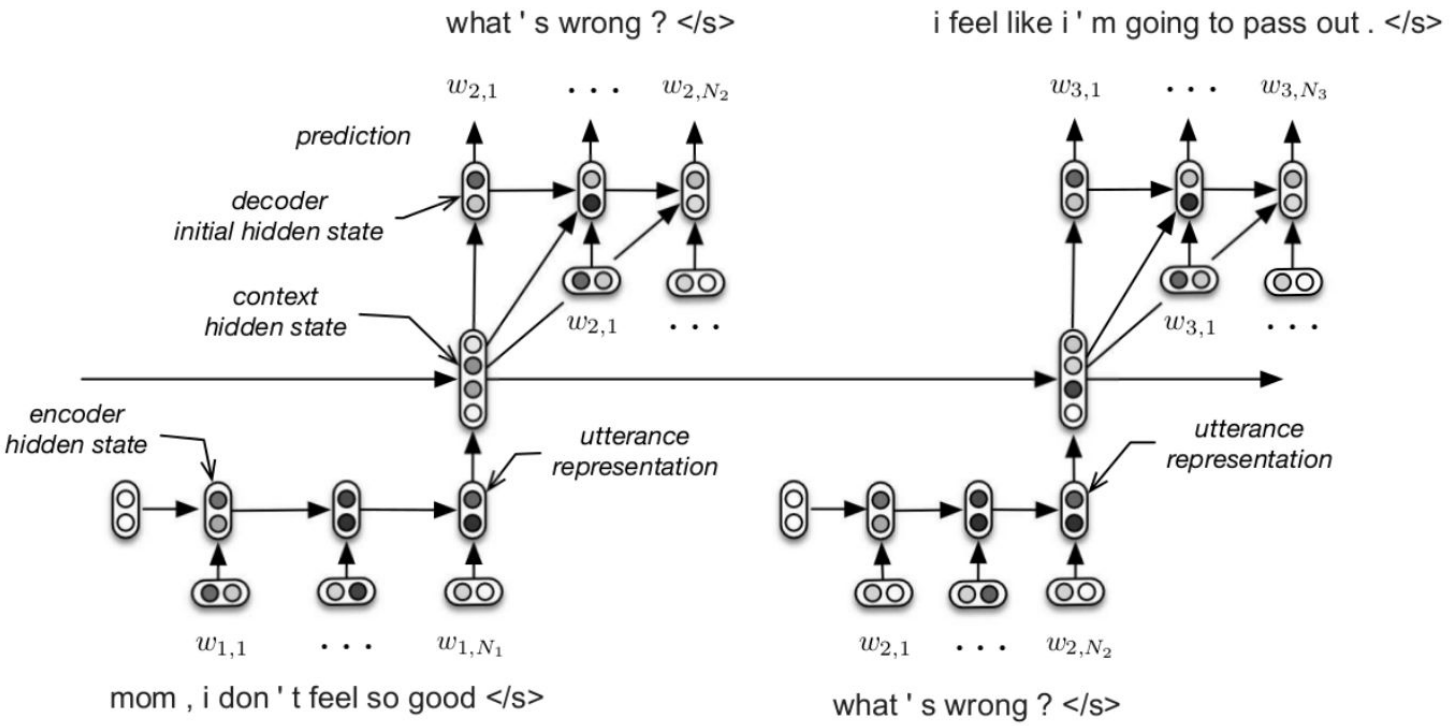
- Must understand the **dialog context**
- Must be able to discuss a **variety of topics**
- Must generate **natural language** responses
- Must generate **engaging/interesting** responses
- Must demonstrate **common sense reasoning**

Seq2Seq [Vinyals and Le. 2015]



Human: *what is the purpose of life ?*
Machine: *to serve the greater good .*
Human: *what is the purpose of living ?*
Machine: *to live forever .*
Human: *what is the purpose of existence ?*
Machine: *to find out what happens when we get to the planet earth .*
Human: *where are you now ?*
Machine: *i 'm in the middle of nowhere .*
Human: *what is the purpose of dying ?*
Machine: *to have a life .*
Human: *what is the purpose of being intelligent ?*
Machine: *to find out what it is .*
Human: *what is the purpose of emotions ?*
Machine: *i don 't know .*

HRED [Serban et al. 2016]



Diversity Promoting Objective [Li et al. 2016]

- To mitigate the **dull response problem** (“*I don’t know”*), Li et al. proposes a diversity-promoting objective function
  - Use MMI rather than cross entropy as a loss function
  - Penalize high-likelihood responses (anti LM objective)

Diversity Promoting Objective [Li et al. 2016]

message	SEQ2SEQ	MMI $P(T\|S) - \lambda P(T)$
I did not get the report from an mi-6 agent.	I don't think that's a good idea.	You did the right thing, did you?
You haven't been given an assignment in this case.	I don't know what you are talking about.	I've been looking all over for you.
I'm losing my grip.	I don't know what you are talking about.	I'm the only one in the world.
I am ready to help.	Come on, come on	I have something we need to talk about.
You programmed me to gather intelligence. That's all I've ever done.	You know that, don't you?	You do have fun, don't you?
I mean, we'd have to talk to him.	I mean, I don't know.	I mean, he's a good guy.
This has nothing to do with him.	You don't know anything about him, do you?	I know him as much.
How come you never say it?	I don't know	Because I don't want to hurt you
He is the one behind this scheme to crash x-net.	I don't know what you are talking about	Yes, he is, isn't he?
How much time do you have here?	I don't know.	Not long enough. Sorry, sir.

Persona-Conditioned Models [Zhang et al. 2018]

- To make open-domain chit-chat dialog models more **consistent** and **engaging**, condition them on a persona

Persona 1	**Persona 2**
I like to ski	I am an artist
My wife does not like me anymore	I have four children
I have went to Mexico 4 times this year	I recently got a cat
I hate Mexican food	I enjoy walking for exercise
I like to eat cheetos	I love watching Game of Thrones

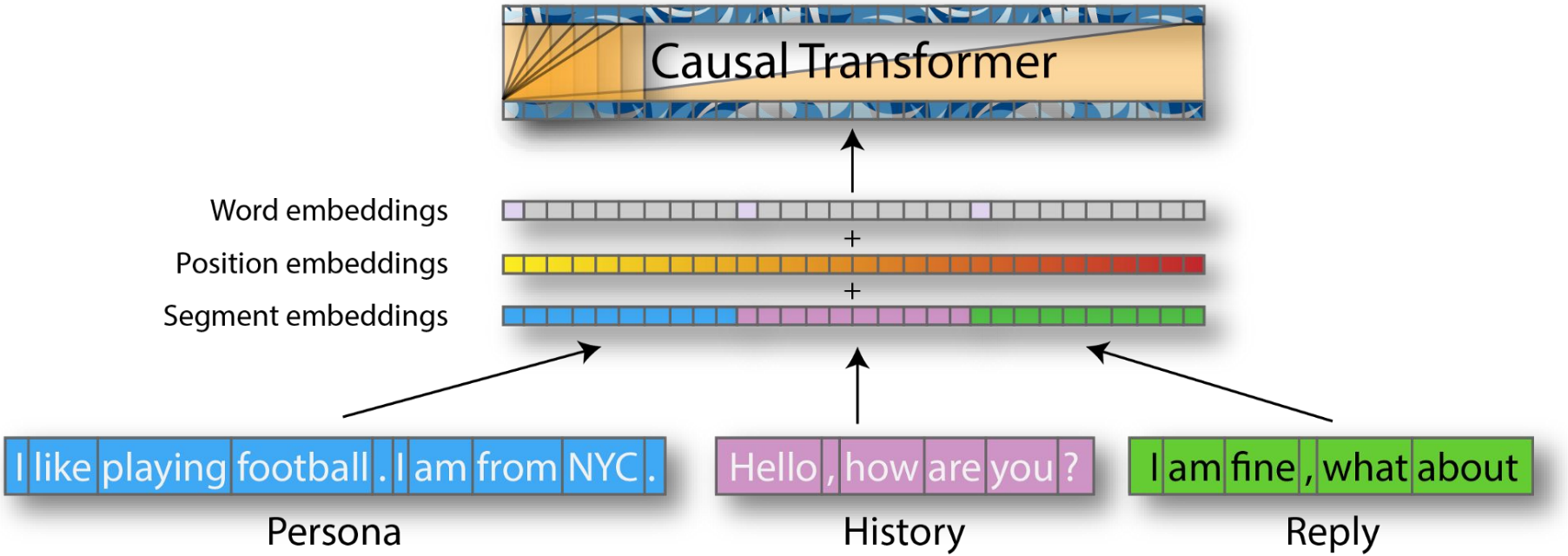
[PERSON 1:] Hi
[PERSON 2:] Hello ! How are you today ?
[PERSON 1:] I am good thank you , how are you.
[PERSON 2:] Great, thanks ! My children and I were just about to watch Game of Thrones.
[PERSON 1:] Nice ! How old are your children?
[PERSON 2:] I have four that range in age from 10 to 21. You?
[PERSON 1:] I do not have children at the moment.
[PERSON 2:] That just means you get to keep all the popcorn for yourself.
[PERSON 1:] And Cheetos at the moment!
[PERSON 2:] Good choice. Do you watch Game of Thrones?
[PERSON 1:] No, I do not have much time for TV.
[PERSON 2:] I usually spend my time painting: but, I love the show.

Persona-Conditioned Models [Zhang et al. 2018]

Persona 1	Persona 2
I am engaged to be married	I am a writer
I used to play sports in high school	I write romance novels
My favorite sport is soccer	I live on a faraway island
I am a college graduate	I am an introvert
I love watching sports on TV	I have two cats

[PERSON 1:] I just got done playing soccer. What have you been up to?
[PERSON 2:] I do not have pets. I spend my free time writing. what about you?
[PERSON 1:] Just graduated from college. Looking for work.
[PERSON 2:] yes I graduated, in writing and poetry
[PERSON 1:] Have you had anything published?
[PERSON 2:] I mainly write fiction novels.
[PERSON 1:] Do you like to watch sports?
[PERSON 2:] do you like kings of leon my favorite by them is use somebody
[PERSON 1:] Are you married? I will be married soon.
[PERSON 2:] haha, no time. I have got a novel to finish.
[PERSON 1:] What is your novel about?
[PERSON 2:] it is a love story. I really enjoy writing. do you write or read?
[PERSON 1:] Like to read. Spend most time watching football.
[PERSON 2:] yes I do I am and author for a living

Transfer-Transfo [Wolf et al. 2019]



DialoGPT [Zhang et al. 2019]

- Continue pre-training GPT-2 on **conversations from Reddit**
  - Filter long utterances
  - Filter non-English utterances
  - Filter URLs
  - Filter toxic comments
- Train on 147M dialog instances (1.8B words)
- “Human-level” response generation ability

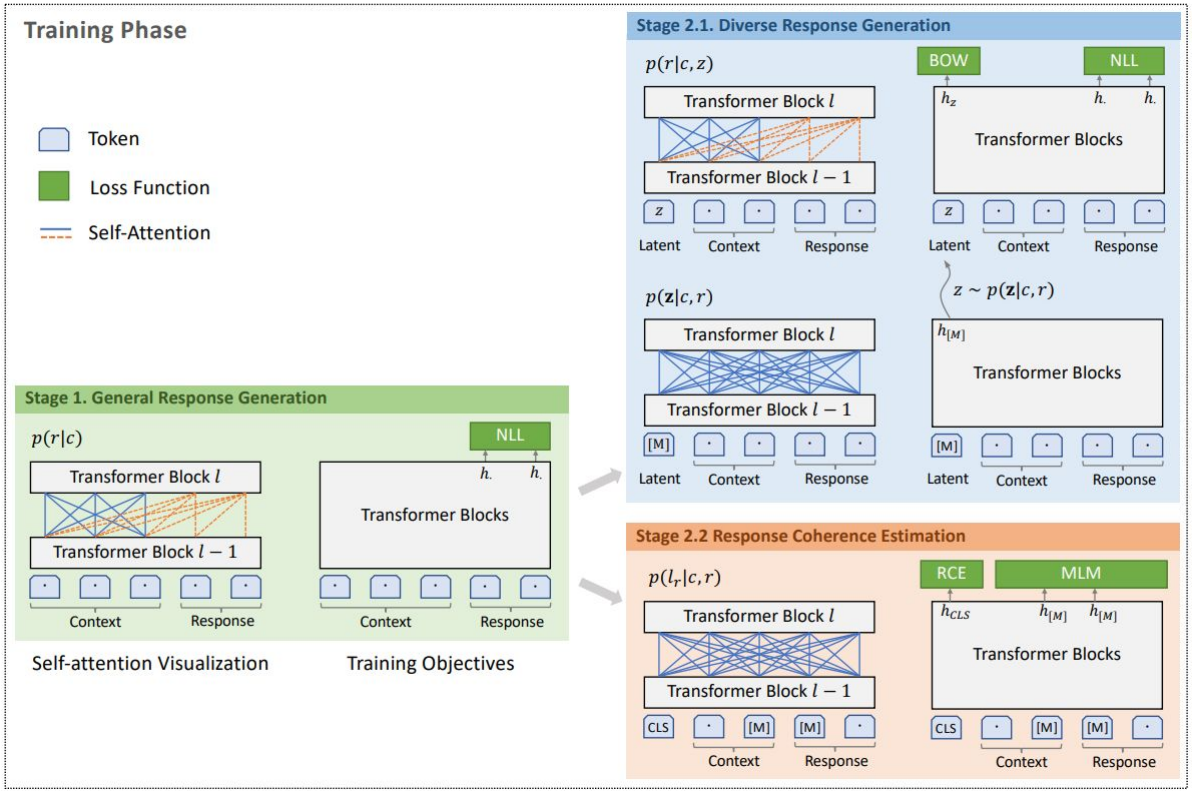
Meena [Adiwardana et al. 2020]

Meena [Adiwardana et al. 2020]

Conversation C	
Human:	*Haha, I think mathematics has a presence beyond human existence.*
Meena:	*So does philosophy.*
Human:	*Can you state with examples how this is true.*
Meena:	*Can you state how it isn't?*
Human:	*Without human existence there will be no one left to comprehend whats right whats wrong, and thus philosophy wi [sic]*
Meena:	*That's not the definition of philosophy.*
Human:	*Without human existence there will be no one left to comprehend whats right whats wrong, and thus philosophy will not be needed. How do you define philosophy?*
Meena:	*The study of general knowledge or the fundamental nature of things.*

Conversation D	
Human:	*I like how cute cows are*
Meena:	*Ah, interesting. You like how they are cute, right?*
Human:	*yeah, and they're realy smart*
Meena:	*I heard they go to college*
Human:	*cows go to college?*
Meena:	*I heard that a cow went to Harvard.*
Human:	*what did the cow study?*
Meena:	*Bovine sciences.*
Human:	*do horses go to Harvard?*
Meena:	*Horses go to Hayvard.*
Human:	*that's a pretty good joke, I feel like you led me into it*
Meena:	*You were trying to steer it elsewhere, I can see it.*

PLATO-2 [Bao et al. 2021]



PLATO-2 [Bao et al. 2021]

User: Hello
System: hello! hows your day going so far?
User: What are you up to?
System: just finished making coffee! what about you? what do you do?
User: i'm playing video games
System: oh, nice! what kind? and i'm just relaxing here and listening to some classical music.
User: i'm playing league of legends
System: nice! what rank are you? i haven't played in forever.
User: who is your favorite classical musician?
System: probably beethoven. i like classical music.
User: i'm plat 2
System: nice! i'd imagine you are better than that!
User: can you play an instrument?
System: i can play the piano. do you play any instruments?
User: thanks
System: i appreciate it! so what is it like being plat 2?
User: i can play the clarinet, but not very well haha
System: oh nice! i played clarinet in high school.
User: the games are difficult but rewarding
System: yes they are! i played the trumpet in college.
. . .

Open-Domain Response Generation

- **Knowledge-grounded** response generation
- **Persona-grounded** response generation
- **Negotiation/persuasive** dialog
- **Commonsense** dialog
- Conversational **QA**

NLU

Natural language understanding in dialog involves several key tasks:

- **Intent prediction:** *what is the user's intent/goal*
- **Slot filling:** *what are the slot values (e.g., what is the time)*
- **State tracking:** *track user information/goals throughout the dialog*

NLU

Natural language understanding in dialog involves several key tasks:

- **DialoGLUE** [Mehri et al. 2020]
- **Intent prediction**: ATIS, SNIPS, Banking77, CLINC150, HWU64
- **Slot filling**: ATIS, SNIPS, DSTC8-SGD, Restaurant8k
- **State tracking:** MultiWOZ (2.X)

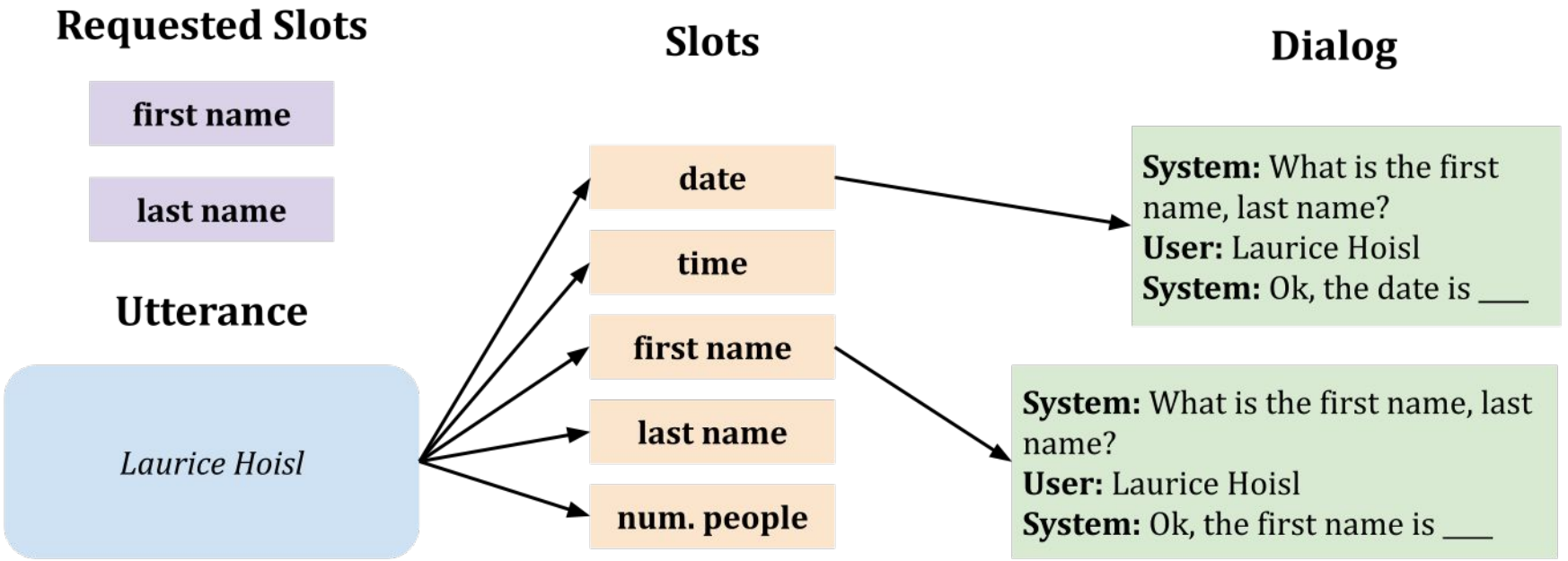
ConVEx [Henderson and Vulic. 2020]

Pre-training paradigm specifically for slot filling → strong few-shot/zero-shot performance

***Template* Sentence**	***Input* Sentence**
I get frustrated everytime I browse /r/all. I stick to my *BLANK* most of the time.	/r/misleadingpuddles Saw it on the **frontpage**, plenty of content if you like the premise.
Why Puerto Rico? It's Memphis at Dallas, which is in Texas where *BLANK* hit	**Hurricane Harvey**. Just a weird coincidence.
BLANK is my 3rd favorite animated Movie	**Toy Story 3** ended perfectly, but Disney just wants to keep milking it.
It really sucks, as the V30 only has *BLANK* . Maybe the Oreo update will add this.	Thanks for the input, but **64GB** is plenty for me :)
I took *BLANK*, cut it to about 2 feet long and duct taped Vive controllers on each end. Works perfect	Yeah, I just duct taped mine to **a broom stick**. You can only play no arrows mode but it's really fun.
I had *BLANK* and won the last game and ended up with 23/20 and still didn't get it.	I know how you feel my friend and I got **19/20** on the tournament today

Table 1: Sample data from Reddit converted to sentence pairs for the ConVEx pretraining via the pairwise cloze task. Target spans in the input sentence are denoted with bold, and are *"BLANKed"* in the template sentence.

GenSF [Mehri and Eskenazi. 2021]



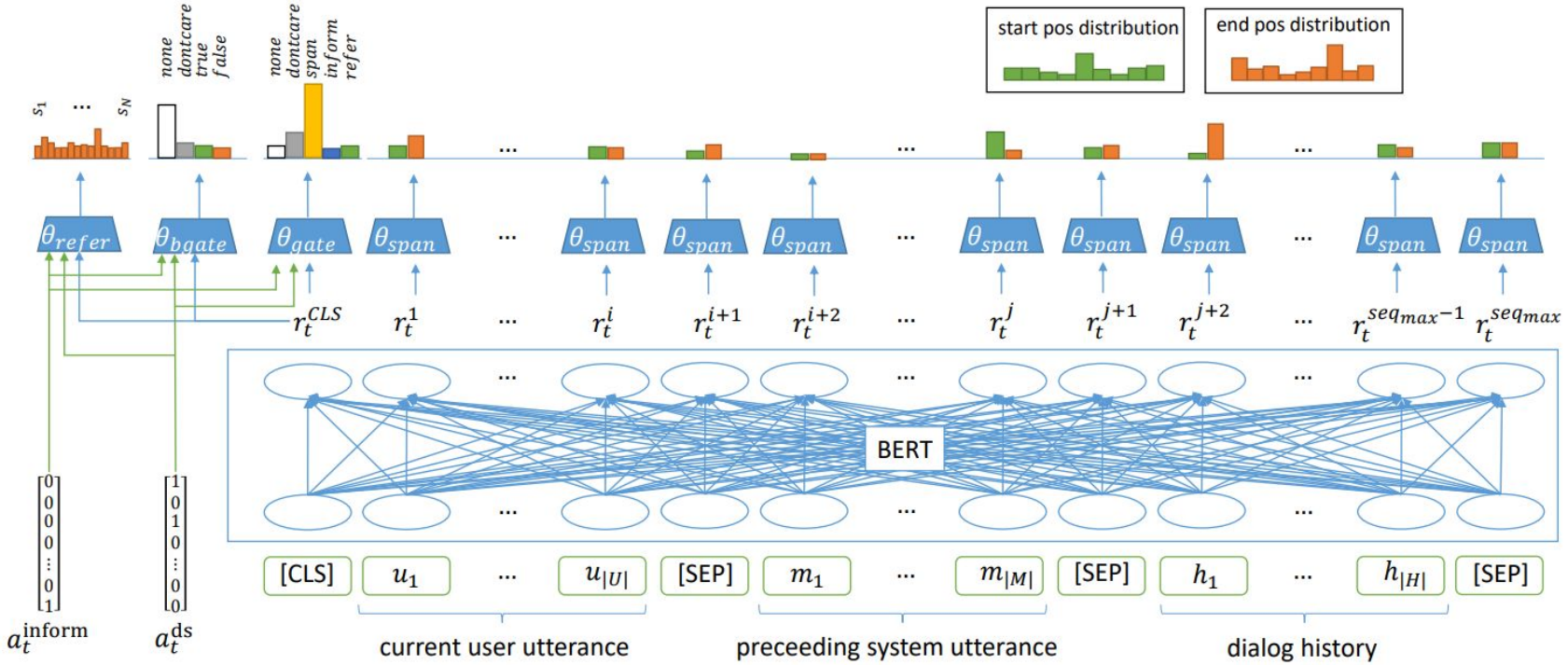
Results on Restaurant8k

Fraction	Span-BERT	ConVEx	GenSF
1 (8198)	93.1	96.0	**96.1**
1/2 (4099)	91.4	94.1	**94.3**
1/4 (2049)	88.0	92.6	**93.2**
1/16 (512)	76.6	86.4	**89.7**
1/128 (64)	30.6	71.7	**72.2**

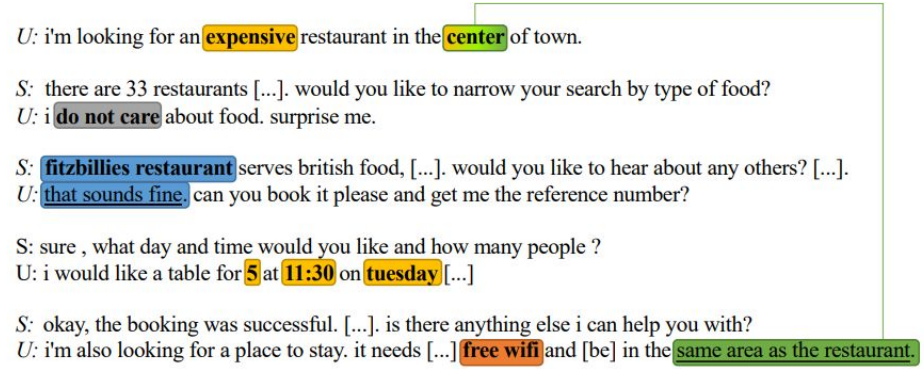
Zero-Shot Slot Filling

Slot	Coach + TR	ConVEx	GenSF
First Name	2.5	4.1	**19.8**
Last Name	0.0	3.4	**13.8**
Date	**15.7**	3.6	12.6
Time	**35.1**	9.1	34.7
Number of People	0	6.0	**16.4**
Average	10.7	5.2	**19.5**

TripPy [Heck et al. 2020]



TripPy [Heck et al. 2020]



Turn	Domain-slot pair	Value	Type	Coreference
0	restaurant-pricerange	expensive	span	
0	restaurant-area	center	span	
1	restaurant-food	*<dontcare>*	*(dontcare)*	
2	restaurant-name	fitzbillies	informed	
3	restaurant-people	5	span	
3	restaurant-book_time	11:30	span	
3	restaurant-book_day	tuesday	span	
4	hotel-internet	*<true>*	*(bool)*	
5	hotel-area	center	coreference (multiturn)	restaurant-area

Dialog Evaluation

Goal: Construct automatic evaluation metrics for response generation/interactive dialog

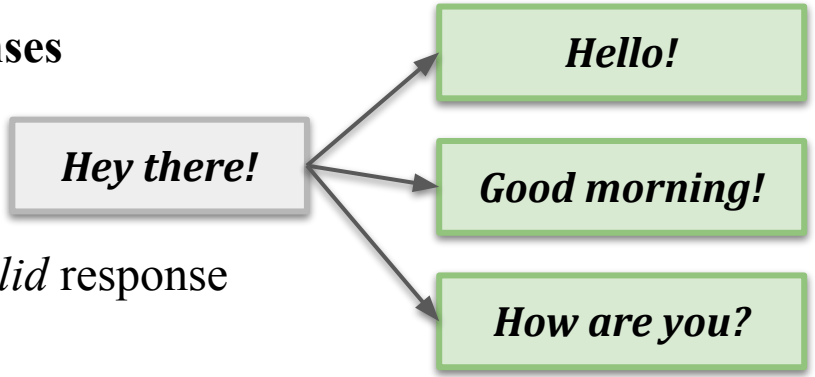
Given: dialog history, generated response, reference response (optional)

Output: a score for the response

Why is evaluating dialog hard? (1/3)

1. **One-to-many** nature of dialog

- For each dialog there are **many valid responses**
- Cannot compare to a reference response
    - The reference response isn't the *only valid* response
- Existing metrics won't work
    - BLEU, F-1, etc.



Why is evaluating dialog hard? (2/3)

2. Dialog quality is **multi-faceted**

- A response isn't just **good** or **bad**
- For interpretability, should measure **multiple qualities**
  - Relevance
  - Interestingness
  - Fluency

Why is evaluating dialog hard? (3/3)

3. Dialog is inherently **interactive**

- Dialog systems are designed to have a **back-and-forth interaction** with a user
    - Research largely focuses on **static corpora** → Reduces the problem of dialog to **response generation**
- Some properties of a system can't be assessed outside an **interactive** environment
    - Long-term planning, error recovery, coherence.

Dialog Evaluation

- Evaluation of dialog is **hard**
  - Can’t compare to a **reference** response [no BLEU, F-1, etc.]
  - Should assess **many aspects** of dialog quality [relevant, interesting, etc.]
  - Should evaluate in an **interactive** manner

Dialog Evaluation

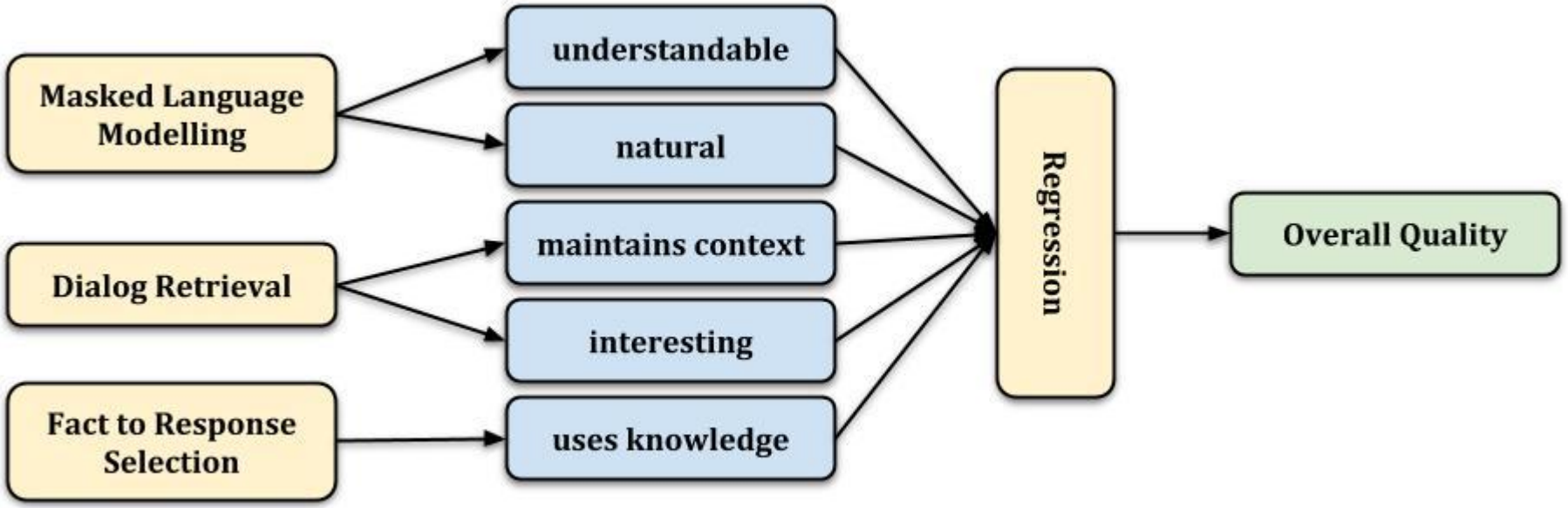
- USR [Mehri and Eskenazi. 2020]
- GRADE [Huang et al. 2020]
- HolisticEval [Pang et al. 2020]
- DSTC6 [Hori and Hori. 2017]
- FED [Mehri and Eskenazi. 2020]
- DSTC9 [Gunasekara et al. 2021]

If you're interested in dialog evaluation. Check out our repository and paper:

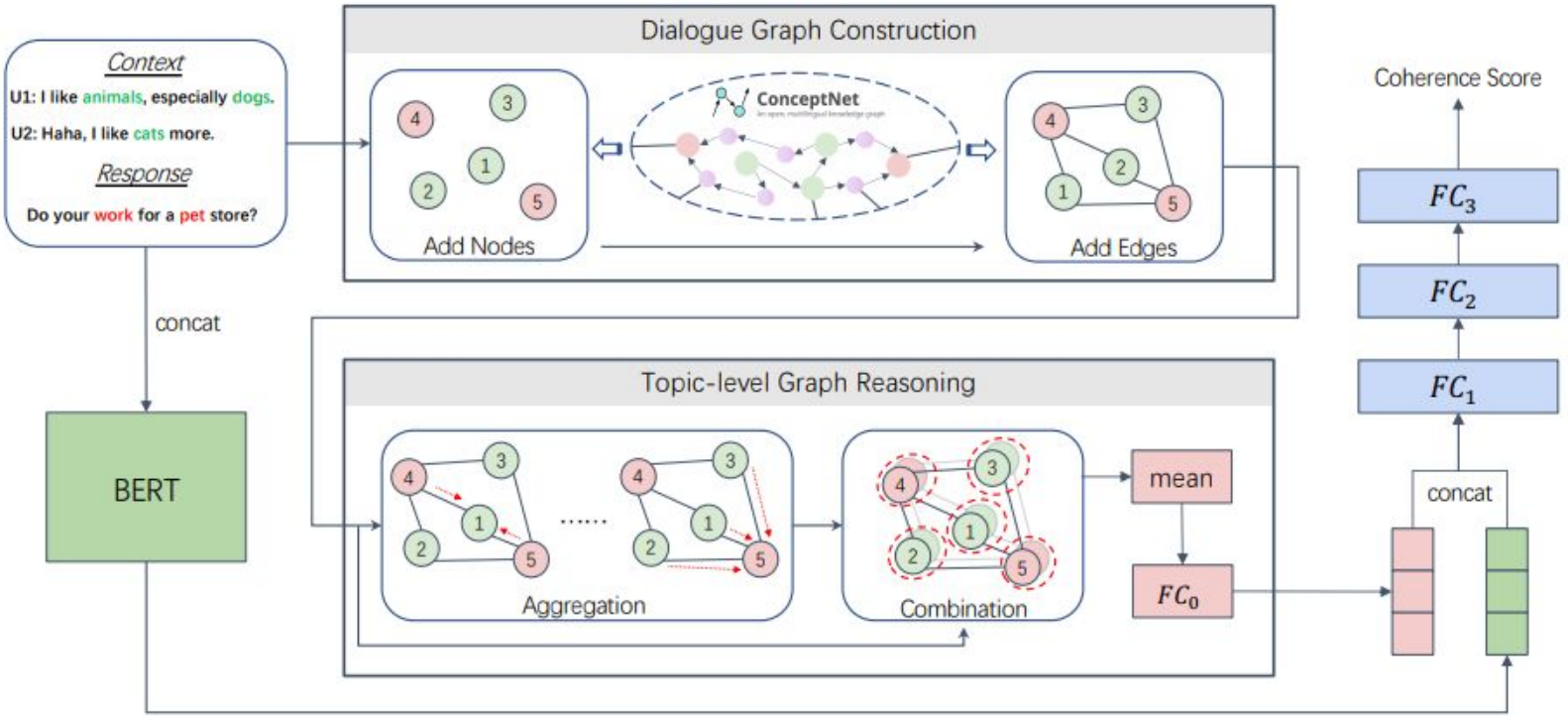
https://github.com/exe1023/DialEvalMetrics

A Comprehensive Assessment of Dialog Evaluation Metrics
Yi-Ting Yeh, Maxine Eskenazi, Shikib Mehri

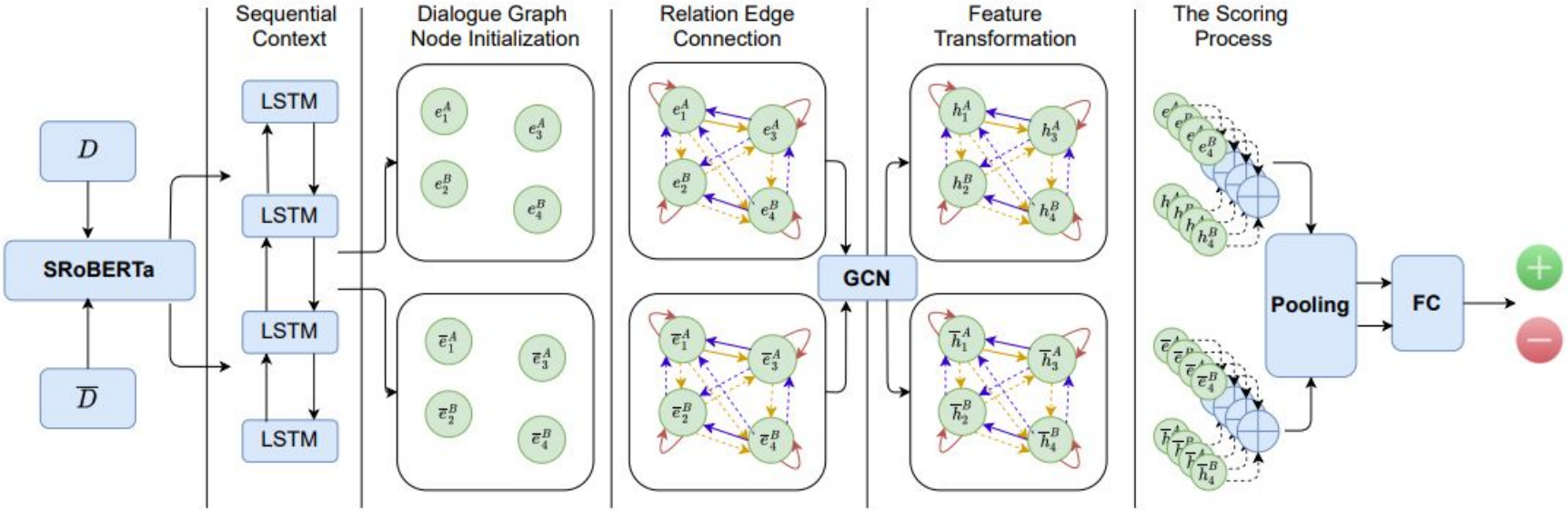
USR [Mehri and Eskenazi. 2020]



GRADE [Huang et al. 2020]



DynaEval [Zhang et al. 2021]



Questions?