Dialogue Systems And Evaluation

Thomas Gerald

February 16, 2024

Laboratoire Interdisciplinaire des Sciences du Numérique – LISN, CNRS
firstname.lastname@lisn.upsaclay.fr

Objectives of the course

Objectives : The Lesson

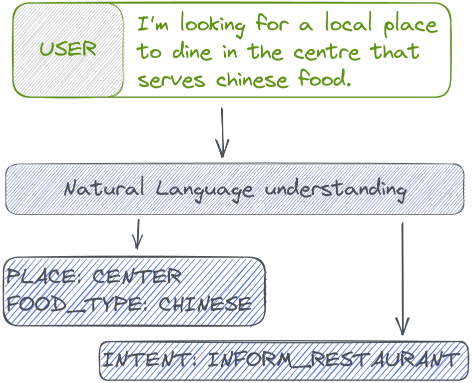
Overview of Dialogue systems:

- **What is a dialogue system?** Provide definitions to understand the differents aspect of dialogue systems.
- **How does it works?** Define the main objectives and the structure of the different approaches
- **What is the current state of the art and what next?** Provide a short overview of futures directions to better adress conversation tasks

Corpus & Evaluation:

- **What kind of data?** Overview of dialogue corpus
- **How to evaluate?** Presentation of the different evaluation approaches

What it is not covered here?
Details on all the approaches.



A task oriented system for restaurant booking:

- Build a simple NLU model (we will explain it later)
- Evaluate the different approaches
- Use a generative system to answer user utterance

introduction

Objectives of a Dialogue Systems

- Provide a service to help users to fullfill a task

Objectives of a Dialogue Systems

- Provide a service to help users to fullfill a task
  - Specific knowledge (e.g. medical, gathering informations, FAQ)

Objectives of a Dialogue Systems

- Provide a service to help users to fullfill a task
  - Specific knowledge (e.g. medical, gathering informations, FAQ)
  - Interact to complete a task with real world implication (e.g. booking, agenda etc..)

Objectives of a Dialogue Systems

- Provide a service to help users to fullfill a task
  - Specific knowledge (e.g. medical, gathering informations, FAQ)
  - Interact to complete a task with real world implication (e.g. booking, agenda etc..)
- Interract with the user in a sequential ways (dialogue)

Objectives of a Dialogue Systems

- Provide a service to help users to fullfill a task
  - Specific knowledge (e.g. medical, gathering informations, FAQ)
  - Interact to complete a task with real world implication (e.g. booking, agenda etc..)
- Interract with the user in a sequential ways (dialogue)
- Interract in natural language

Objectives of a Dialogue Systems

- Provide a service to help users to fullfill a task
  - Specific knowledge (e.g. medical, gathering informations, FAQ)
  - Interact to complete a task with real world implication (e.g. booking, agenda etc..)
- Interract with the user in a sequential ways (dialogue)
- Interract in natural language

Different terms for dialogue systems

- Conversationnal agent : implies an avatar (animated agent)
- Dialogue System : interract with a system (e.g. knowledge base, relationnal database...)
- Chatbots and intelligent assistant : more often used from application point of view

Define two actors

- **USER:** An user (person) which interact with the system. The user query the system

Define two actors

- **USER:** An user (person) which interact with the system. The user query the system
- **SYSTEM:** The chatbot system detect the intent of the user and answer accordingly

Define two actors

- **USER:** An user (person) which interact with the system. The user query the system
- **SYSTEM:** The chatbot system detect the intent of the user and answer accordingly

Concept definition

- **utterance:** text of a turn weither by user or the system
- **user intent:** intention given by the user (must be interpreted by the model)

Define two actors

- **USER:** An user (person) which interact with the system. The user query the system
- **SYSTEM:** The chatbot system detect the intent of the user and answer accordingly

Concept definition

- **utterance:** text of a turn weither by user or the system
- **user intent:** intention given by the user (must be interpreted by the model)

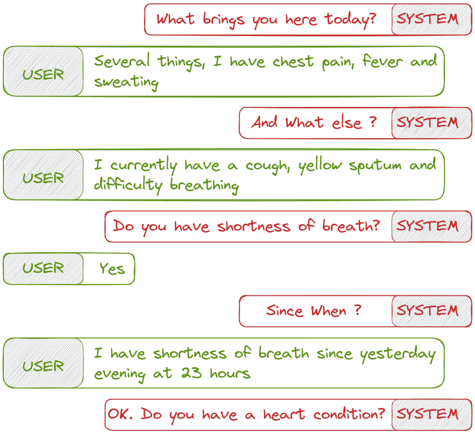


Figure 1: An example of dialogue in PVDial corpus

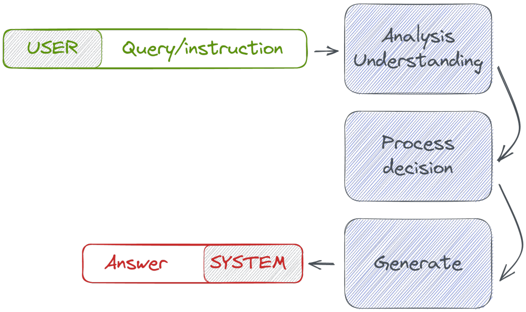


Figure 2: Base concepts of a Dialogue system

A simple dialogue system:

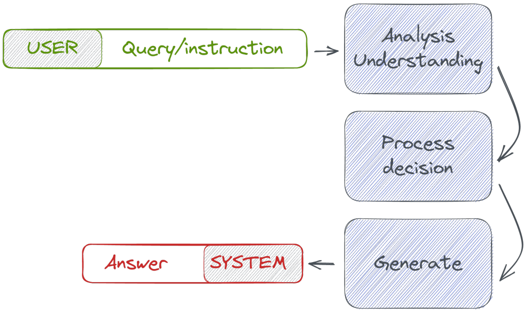


Figure 2: Base concepts of a Dialogue system

A simple dialogue system:

- **Analyse:** Analysing the user utterence to understant the user intent

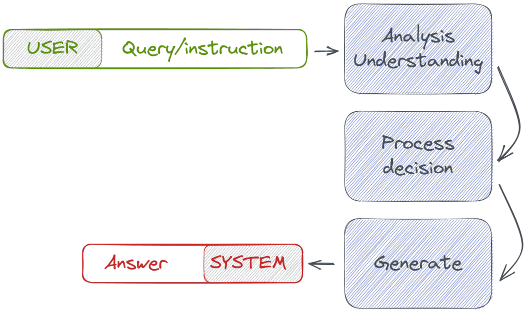


Figure 2: Base concepts of a Dialogue system

A simple dialogue system:

- **Analyse:** Analysing the user utterence to understant the user intent
- **Act/decide:** Select an action (e.g. query a database), next information to return

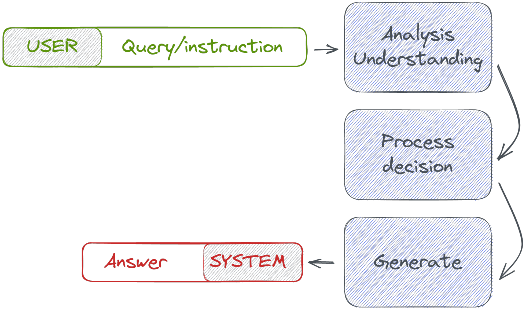


Figure 2: Base concepts of a Dialogue system

A simple dialogue system:

- **Analyse:** Analysing the user utterence to understant the user intent
- **Act/decide:** Select an action (e.g. query a database), next information to return(inform/clarify)
- **Generate:** Return textual information about action/requeste of the users

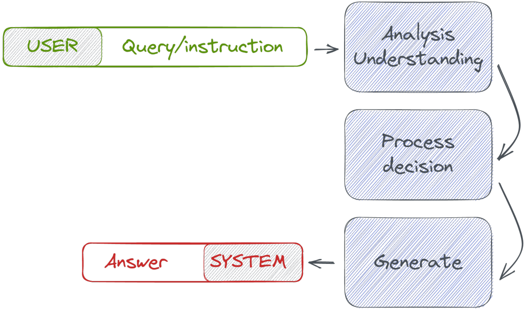


Figure 2: Base concepts of a Dialogue system

A simple dialogue system:

- **Analyse:** Analysing the user utterence to understant the user intent
- **Act/decide:** Select an action (e.g. query a database), next information to return(inform/clarify)
- **Generate:** Return textual information about action/requeste of the users

Needs:

- A combination of different systems (not always true, e.g. end-to-end system)
- Cover at least two differents objectives, accomplish a task and chat with user

Different kind of systems:

- **Task-oriented systems:** Help user to fullfill a task, obtain information on a specific topic such as booking restaurant, specific medical advice...

Chatbot different approaches

Different kind of systems:

- **Task-oriented systems:** Help user to fullfill a task, obtain information on a specific topic such as booking restaurant, specific medical advice...
- **ChitChat/purely conversational systems:** Finding the most natural reaction to an user utterance and maintain engagement

Chatbot different approaches

Different kind of systems:

- **Task-oriented systems:** Help user to fullfill a task, obtain information on a specific topic such as booking restaurant, specific medical advice...
- **ChitChat/purely conversational systems:** Finding the most natural reaction to an user utterance and maintain engagement
- **Question-answering systems:** Help the user to find informations or answers to a question (it can be document/facts or generated answers)

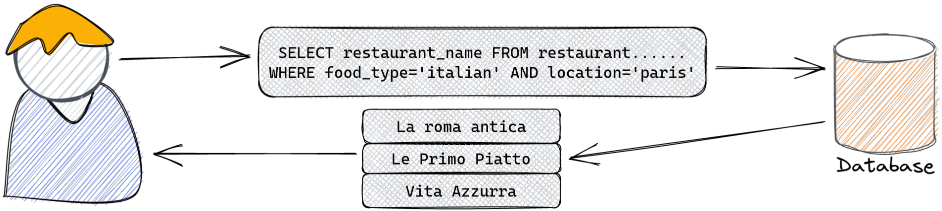
Chatbot different approaches

Different kind of systems:

- **Task-oriented systems:** Help user to fullfill a task, obtain information on a specific topic such as booking restaurant, specific medical advice...
- **ChitChat/purely conversational systems:** Finding the most natural reaction to an user utterance and maintain engagement
- **Question-answering systems:** Help the user to find informations or answers to a question (it can be document/facts or generated answers)

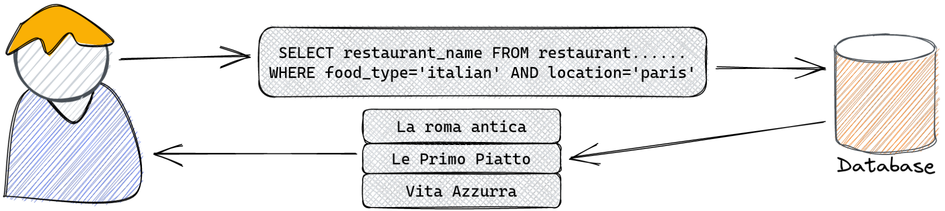
All those objectives can be combined !!!

Task Oriented Dialogue systems



Query a database:

- Query to get information from a database
- Update the database (e.g. to book a reservation, register task in an agenda...)



Query a database:

- Query to get information from a database
- Update the database (e.g. to book a reservation, register task in an agenda...)

Drawback
Not easy to interract for most people...

Query the system:

- Query the system in natural language (NLU system)
- The system will "transform"the query into database instruction

Query the system:

- Query the system in natural language (NLU system)
- The system will "transform"the query into database instruction

Drawback
Answer can be difficult to interpret

System Answer

- Translate the responses in natural language

System Answer

- Translate the responses in natural language

Drawback
Not able to understand context

System Answer

- Consider dilaogue history

Tasks Based dialogue systems

System:

- Provide a service, obtain specific informations or help user for a task
- Query/update a structured database or a knowledge base

Tasks exemples :

Tasks Based dialogue systems

System:

- Provide a service, obtain specific informations or help user for a task
- Query/update a structured database or a knowledge base

Tasks exemples :

- Dialogue to maintain an Agenda

Tasks Based dialogue systems

System:

- Provide a service, obtain specific informations or help user for a task
- Query/update a structured database or a knowledge base

Tasks exemples :

- Dialogue to maintain an Agenda
- Booking accomodations/restaurants

System:

- Provide a service, obtain specific informations or help user for a task
- Query/update a structured database or a knowledge base

Tasks exemples :

- Dialogue to maintain an Agenda
- Booking accomodations/restaurants
- Customer services

Tasks Based dialogue systems

System:

- Provide a service, obtain specific informations or help user for a task
- Query/update a structured database or a knowledge base

Tasks exemples :

- Dialogue to maintain an Agenda
- Booking accomodations/restaurants
- Customer services
- Medical assistant
- ...

What do we need ?

Tasks Based dialogue systems

System:

- Provide a service, obtain specific informations or help user for a task
- Query/update a structured database or a knowledge base

Tasks exemples :

- Dialogue to maintain an Agenda
- Booking accomodations/restaurants
- Customer services
- Medical assistant
- ...

What do we need ?

- Understand what is the user goal (intent)

Tasks Based dialogue systems

System:

- Provide a service, obtain specific informations or help user for a task
- Query/update a structured database or a knowledge base

Tasks exemples :

- Dialogue to maintain an Agenda
- Booking accomodations/restaurants
- Customer services
- Medical assistant
- ...

What do we need ?

- Understand what is the user goal (intent)
- Gather multiple turns informations related to user task

Tasks Based dialogue systems

System:

- Provide a service, obtain specific informations or help user for a task
- Query/update a structured database or a knowledge base

Tasks exemples :

- Dialogue to maintain an Agenda
- Booking accomodations/restaurants
- Customer services
- Medical assistant
- ...

What do we need ?

- Understand what is the user goal (intent)
- Gather multiple turns informations related to user task
- Act (filling a database), inform the user

Tasks Based dialogue systems

System:

- Provide a service, obtain specific informations or help user for a task
- Query/update a structured database or a knowledge base

Tasks exemples :

- Dialogue to maintain an Agenda
- Booking accomodations/restaurants
- Customer services
- Medical assistant
- ...

What do we need ?

- Understand what is the user goal (intent)
- Gather multiple turns informations related to user task
- Act (filling a database), inform the user
- Answering in Natural language

A classical Pipeline

Different Modules :

A classical Pipeline

Different Modules :

- **NLU**(understanding): **N**atural **L**anguage **U**nderstanding system. It's objective is to interpret user utterances and retrieve, classify or select the different information needed by the system.

Different Modules :

- **NLU**(understanding): **N**atural **L**anguage **U**nderstanding system. It's objective is to interpret user utterances and retrieve, classify or select the different information needed by the system.
- **DST**(Dialogue Manager): **D**ialogue **S**tate **T**racking. The objective is to keep different information up-to-date (on multiple dialogue turns).

A classical Pipeline

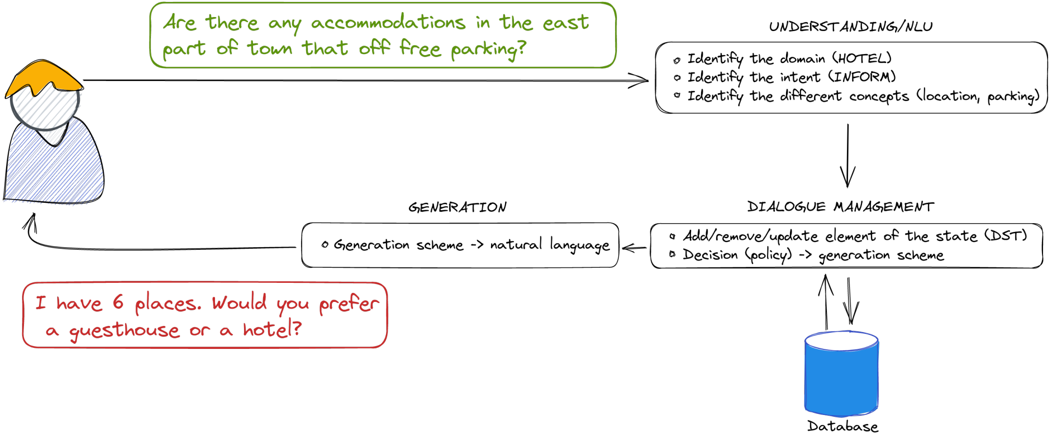
Different Modules :

- **NLU**(understanding): **N**atural **L**anguage **U**nderstanding system. It's objective is to interpret user utterances and retrieve, classify or select the different information needed by the system.
- **DST**(Dialogue Manager): **D**ialogue **S**tate **T**racking. The objective is to keep different information up-to-date (on multiple dialogue turns).
- **Policy**(Dialogue Manager): Selecting what to do, objective is to determine what actions or information the user request. It is often related to the state where the decision is often based upon it.

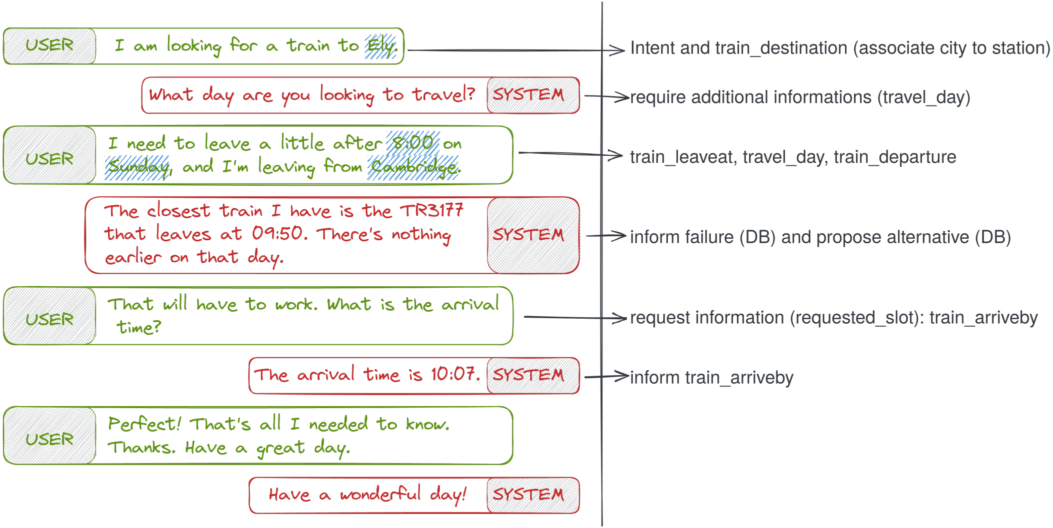
A classical Pipeline

Different Modules :

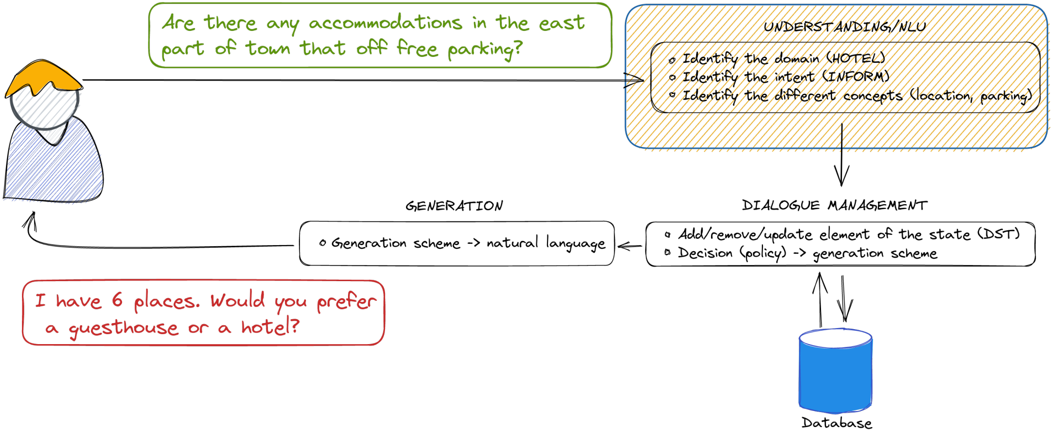
- **NLU**(understanding): **N**atural **L**anguage **U**nderstanding system. It's objective is to interpret user utterances and retrieve, classify or select the different information needed by the system.
- **DST**(Dialogue Manager): **D**ialogue **S**tate **T**racking. The objective is to keep different information up-to-date (on multiple dialogue turns).
- **Policy**(Dialogue Manager): Selecting what to do, objective is to determine what actions or information the user request. It is often related to the state where the decision is often based upon it.
- **NLG**: **N**atural **L**anguage **G**eneration aims to generate an answer based on dialogue manager informations, it process the "system intent" to make it understandable by user.


Are there any accommodations in the east part of town that off free parking?
UNDERSTANDING/NLU
Identify the domain (HOTEL)
Identify the intent (INFORM)
Identify the different concepts (location, parking)
DIALOGUE MANAGEMENT
Add/remove/update element of the state (DST)
Decision (policy) -> generation scheme
Database
GENERATION
Generation scheme -> natural language
I have 6 places. Would you prefer a guesthouse or a hotel?

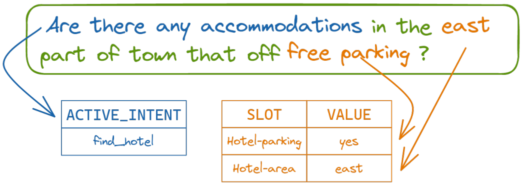
An exemple


USER
I am looking for a train to Ely.
Intent and train_destination (associate city to station)
What day are you looking to travel?
SYSTEM
require additional informations (travel_day)
USER
I need to leave a little after 8:00 on Sunday, and I'm leaving from Cambridge.
train_leaveat, travel_day, train_departure
The closest train I have is the TR3177 that leaves at 09:50. There's nothing earlier on that day.
SYSTEM
inform failure (DB) and propose alternative (DB)
USER
That will have to work. What is the arrival time?
request information (requested_slot): train_arriveby
The arrival time is 10:07.
SYSTEM
inform train_arriveby
USER
Perfect! That's all I needed to know. Thanks. Have a great day.
Have a wonderful day!
SYSTEM

The NLU task



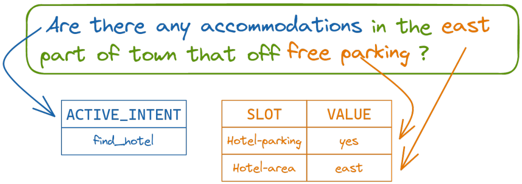
Are there any accommodations in the east part of town that off free parking?
UNDERSTANDING/NLU
Identify the domain (HOTEL)
Identify the intent (INFORM)
Identify the different concepts (location, parking)
DIALOGUE MANAGEMENT
Add/remove/update element of the state (DST)
Decision (policy) -> generation scheme
GENERATION
Generation scheme -> natural language
I have 6 places. Would you prefer a guesthouse or a hotel?
Database



Objective:
Retrieve in utterance the relevant information:

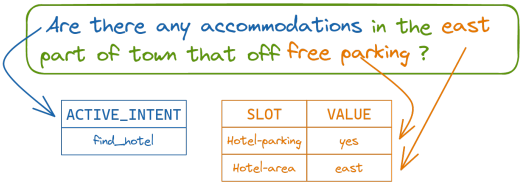
- **Domain**

The NLU task : different subtasks



Objective:
Retrieve in utterance the relevant information:

- **Domain** $\longrightarrow$ HOTEL
- **Intent**



Objective:
Retrieve in utterance the relevant information:

- **Domain** $\longrightarrow$ HOTEL
- **Intent** $\longrightarrow$ find_hotel
- **Concepts:**

The NLU task : different subtasks

Objective:
Retrieve in utterance the relevant information:

- **Domain** $\longrightarrow$ HOTEL
- **Intent** $\longrightarrow$ find_hotel
- **Concepts:**
  - hotel-parking $\longrightarrow$ yes
  - hotel-area $\longrightarrow$ east

Often a set of substasks:

- Detection/identification of :
  - Domain (Multi-domain applications)
  - Intention (Always)
- Concept detection:
  - slot-type (class of the concept, what kind of information)
  - slot-value (value of the concept, what value to store)

Early 2010s [DM11] it is highly related to the development of statistical approaches

Formalisation

Let be u_i and c_i respectivelly the i^{th} utterance and c_i the domain-intent associated to the utterance. We define D_k :

$$D_k = [(u_i, c_i) | \forall i \in \mathbb{N}, i \leq k]$$

The NLU task : Classify the intent

Formalisation
Let be u_i and c_i respectivelly the i^{th} utterance and c_i the domain-intent associated to the utterance. We define D_k :

$$D_k = [(u_i, c_i) | \forall i \in \mathbb{N}, i \leq k]$$

Domains Identification

- Supervised Learning Classification
- The objective is to train a classifier using past dialogue/intent (D_k).

The NLU task : Classify the intent

Formalisation
Let be u_i and c_i respectivelly the i^{th} utterance and c_i the domain-intent associated to the utterance. We define D_k :

$$D_k = [(u_i, c_i) | \forall i \in \mathbb{N}, i \leq k]$$

Domains Identification

- Supervised Learning Classification
- The objective is to train a classifier using past dialogue/intent (D_k).

What kind of classifiers?

- SVM (on BOW/TF-IDF/Word embedding), Entity based . . .

The NLU task : Classify the intent

Formalisation
Let be u_i and c_i respectivelly the i^{th} utterance and c_i the domain-intent associated to the utterance. We define D_k :

$$D_k = [(u_i, c_i) | \forall i \in \mathbb{N}, i \leq k]$$

Domains Identification

- Supervised Learning Classification
- The objective is to train a classifier using past dialogue/intent (D_k).

What kind of classifiers?

- SVM (on BOW/TF-IDF/Word embedding), Entity based ...
- CRF, RNN, LSTM

Formalisation

Let be u_i and c_i respectivelly the i^{th} utterance and c_i the domain-intent associated to the utterance. We define D_k :

$$D_k = [(u_i, c_i) | \forall i \in \mathbb{N}, i \leq k]$$

Domains Identification

- Supervised Learning Classification
- The objective is to train a classifier using past dialogue/intent (D_k).

What kind of classifiers?

- SVM (on BOW/TF-IDF/Word embedding), Entity based ...
- CRF, RNN, LSTM
- Transformer (BERT like) [WHSX20] ([CLS] classification)

The NLU task: Slots

What is a slot
A slot is an information that we would store to better answer, it will additionnaly allow to update the state.

exemple : "symptom" (slot name), "fever" (slot value)

What is a slot
A slot is an information that we would store to better answer, it will additionnaly allow to update the state.

exemple : "symptom" (slot name), "fever" (slot value)

Slots & slots value :
In most task based dialogue different slots can be defined. We can consider two types of slots:

- **Categorical slots:** The differents possible values for the slot is finite, but are not always explicit in the text

The NLU task: Slots

What is a slot
A slot is an information that we would store to better answer, it will additionnaly allow to update the state.

exemple : "symptom" (slot name), "fever" (slot value)

Slots & slots value :
In most task based dialogue different slots can be defined. We can consider two types of slots:

- **Categorical slots:** The differents possible values for the slot is finite, but are not always explicit in the text

  Example $\rightarrow$ a day in a week (7 values from Monday to Sunday)

- **Extractive slots:** It correspond to an entity or a part that cannot be categorized in the text such as most of named entities

The NLU task: Slots

What is a slot
A slot is an information that we would store to better answer, it will additionnaly allow to update the state.

exemple : "symptom" (slot name), "fever" (slot value)

Slots & slots value :
In most task based dialogue different slots can be defined. We can consider two types of slots:

- **Categorical slots:** The differents possible values for the slot is finite, but are not always explicit in the text

  Example $\rightarrow$ a day in a week (7 values from Monday to Sunday)

- **Extractive slots:** It correspond to an entity or a part that cannot be categorized in the text such as most of named entities

  Example 1 $\rightarrow$ a date (not recommended to categorize all posibilities)

  Example 2 $\rightarrow$ a name of a place (e.g. restaurant, hotel, airport, ...)

The NLU task: slot, an example

Slot Name	Description	Categorical
train-arriveby	arrival time of the train	FALSE
train-leaveat	leaving time for the train	FALSE
train-duration	duration of the travel	FALSE
train-departure	departure location of the train	TRUE
train-destination	destination of the train	TRUE
train-day	day of the train	TRUE
train-bookpeople	how many train tickets you need	TRUE
train-price	price of the train	FALSE
train-ref	reference number of the train booking	FALSE
train-trainid	id of the train	FALSE

Table 1: The different slots for the "train" domain in MultiWoZ

In the NLU System :

- Extract part of the utterance reffering to categorical slots
- Classify utterances or text extracted to retrieve the different categorical pair slots values (e.g. mapping extracted value to closest value in the dataset)

The NLU task: slots, scheme example

CATEGORICAL SLOT SCHEME


```
"name": "restaurant-bookday",
"description": "day of the restaurant booking",
"possible_values": [
  "monday",
  "tuesday",
  "wednesday",
  "thursday",
  "friday",
  "saturday",
  "sunday"
```

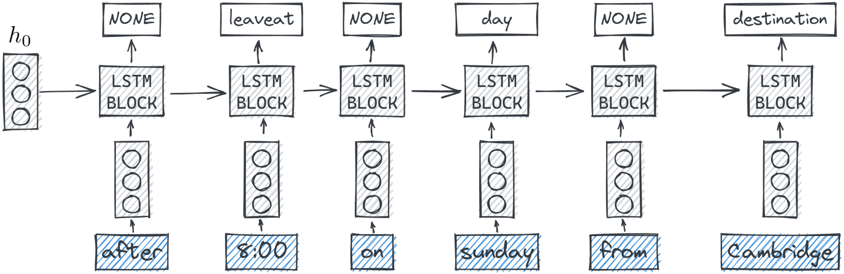

INTENT SCHEME


```
"name": "find_restaurant",
"description": "search for places to wine and dine",
"is_transactional": false,
"required_slots": [],
"optional_slots": {
  "restaurant-pricerange": "dontcare",
  "restaurant-area": "dontcare",
  "restaurant-food": "dontcare",
  "restaurant-name": "dontcare",
  "restaurant-bookday": "dontcare",
  "restaurant-bookpeople": "dontcare",
  "restaurant-booktime": "dontcare"
}
```


EXTRACTIVE SLOT SCHEME

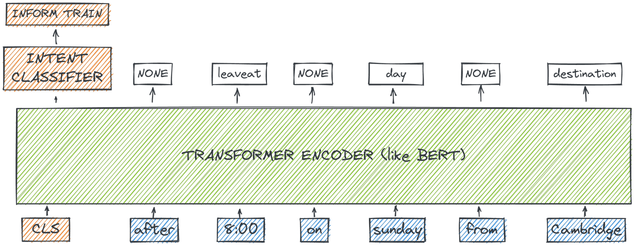

```
"name": "restaurant-booktime",
"description": "time of the restaurant booking",
"possible_values": [],
"is_categorical": false
```


Figure 4: An example of a scheme in MultiWoZ (restaurant domain)



RNN:

- LSTM with moving average (and deep LSTM with stacked LSTM) [YPZ$^+$14]
- Reccurent CRF [MDY$^+$15] (+learning features)
- Bi-LSTM [GSR20] (comparison of different word embeddings)



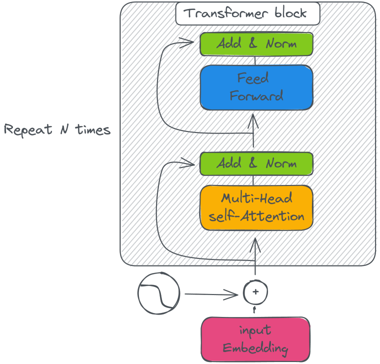


Figure 5: Transformer block architecture

Transformer approaches

- Joint Slot Filling and Intent Classification [WDLX20]
- Comparison on the French Corpus Fine-Tunning a CammenBERT model [GSR20]

Other approaches

Other kind of approaches exists but are more rare today (today mainly based on transformers models)

- Logic formula and λ calculus [vJY04]
- Regular expression [LBZR16]
- Context free grammar [GFG$^+$95]

Joint Slot filling and Intent classification

Learning both in the mean time tends to improve performances (notice that some models previously cited are joint models)

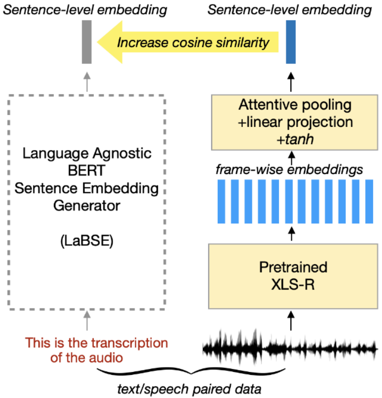


Figure 6: Architecture of SAMU-XLSR [KLG22]

ASR & SLU

An other approach consists in retrieving slot values directly from audio.

- **A**utomatic **S**peech **R**ecognition transform audio to texte
- **S**poken **L**anguage **U**nderstanding interpret the Language

Approaches:

- SAMU-XLSR [KLG22] (training ASR for semantic analysis)
- Few-shot/zero-shot SLU [LPR+22]

What are the approaches nowadays

What are the approaches nowadays

- Mostly transformer classification approaches

What are the approaches nowadays

- Mostly transformer classification approaches
- Some consider generative approaches (often DST models in the mean time)

Direction

What are the approaches nowadays

- Mostly transformer classification approaches
- Some consider generative approaches (often DST models in the mean time)

Direction

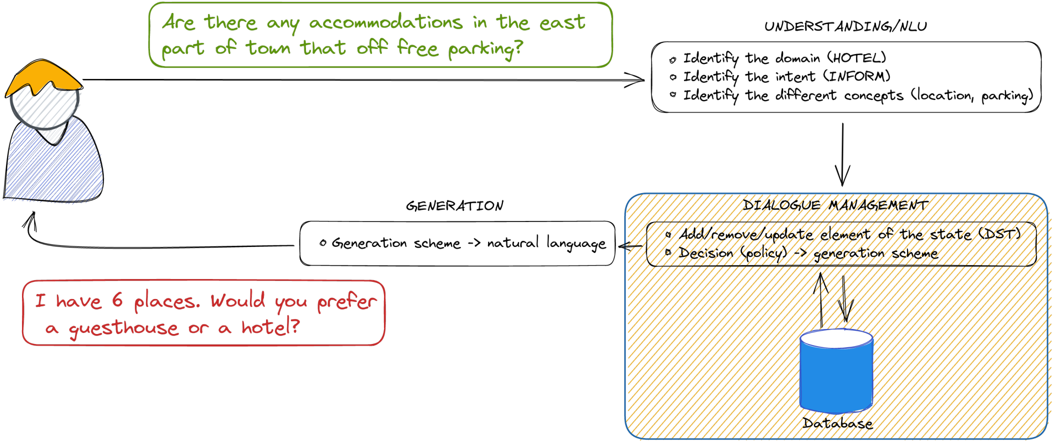
- Transfer for new languages and languages with low ressources [CRS21] (multilingual corpus)
- Lifelong learning, learning new terms, concepts during the dialogue [VRGB21]

NLU and DST:
Today most of the approaches does not only base on an NLU (Intent + Slot) approach, but also consider in the meantime the DST and answer generation !!!

- A simple review on slot filling and intent classification [LL22]

Dialogue Management

Dialogue Management



Understanding in context : multi-turn information
Intent user information rarely depends only on one utterance

- Gathering Information all along the dialog
- Eventually change and update this informations (slot type/value)

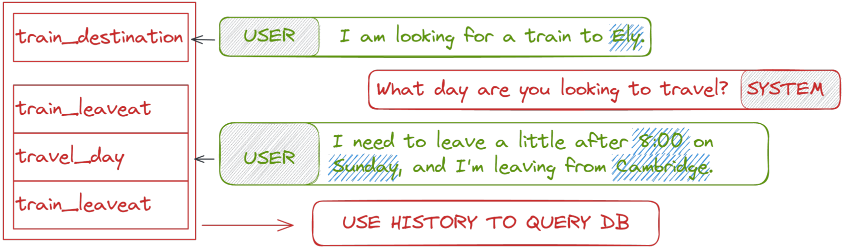
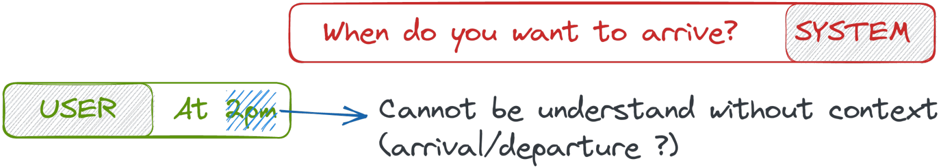
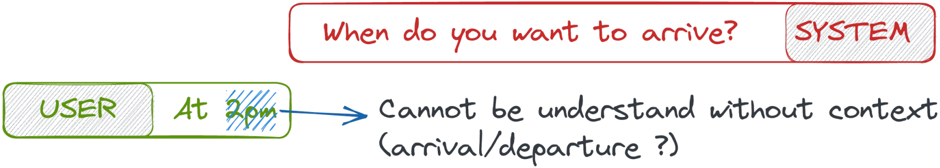


Figure 7: Gathering information for many utterances

Contextualise





Understanding in context
Utterance make reference to previous one

- Some request are reffering to a previous utterance
- Finding what it reffered to (here arrival_time)

Corresponding slot in the database

USER I have heart condition → correspond to hight blood pressure in the database

USER Yes, on monday → correspond to 2023/02/20 in the database

Corresponding slot in the database

Transform slot values (entity linking)

- Different values extracted reffering to same entities
- Information expressed contextually, need to transform to use it (date, time...)

Entity Linking

For entity linking mostly depends on the domain or the Database, different approaches can be envisioned

- Distance between mentions and entities (token/entity)
- specific ontologies (for instance in medical domain)

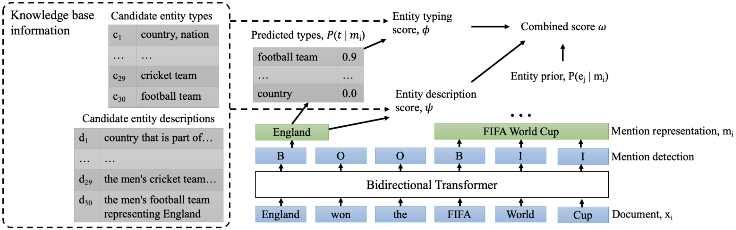


Figure 8: The ReFinED approach for entity linking basing on the distance of contextual embedding and both entity_type and description of the entity [ATF+]

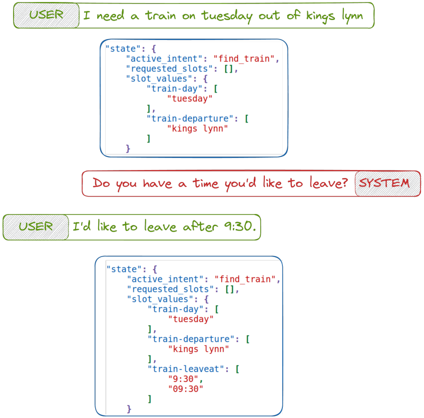
Contextual undertanding what approaches?
Often done by the **D**ialogue **S**tate **T**racking

Contextual undertanding what approaches?
Often done by the **D**ialogue **S**tate **T**racking

- Knowledge Based approaches [LBZR16]

Contextual undertanding what approaches?
Often done by the **D**ialogue **S**tate **T**racking

- Knowledge Based approaches [LBZR16]
- Neural Networks
  - LSTM encoding jointly previous utterances [HHW15]
  - Memory Network [CHTT+16]
  - Transformer with jointly encoded utterances (most of today approaches) [CL19]
  - Transformer based prompt (generative) [LCO21]



Dialiog State Tracking:
Provide a representation of what user want at each utterances [Hen15] :

- Using a set of slots type and value (see previous approaches)
- Using vector or frame (e.g one hot vector) [BA23]
- Update the information (the state vector or pairs of slots type/value)

Decision is then made on current state of the dialogues...

Objectives:
Dialogue control has to decide what is the next step using the user input and the current state.

For instance :

DM and Dialogue control

Objectives:
Dialogue control has to decide what is the next step using the user input and the current state.
For instance :

- Asking new information (e.g. for booking "number of peoples", "times" ...)

DM and Dialogue control

Objectives:
Dialogue control has to decide what is the next step using the user input and the current state.
For instance :

- Asking new information (e.g. for booking "number of peoples", "times" ...)
- Clarrify the the user's input/intent

DM and Dialogue control

Objectives:
Dialogue control has to decide what is the next step using the user input and the current state.
For instance :

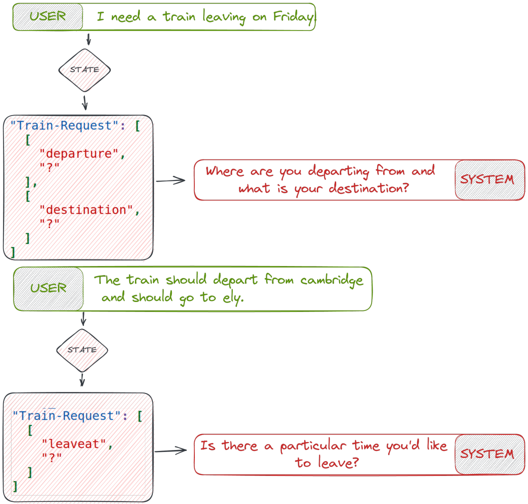
- Asking new information (e.g. for booking "number of peoples", "times" ...)
- Clarrify the the user's input/intent
- giving information to the user

DM and Dialogue control

Objectives:
Dialogue control has to decide what is the next step using the user input and the current state. For instance :

- Asking new information (e.g. for booking "number of peoples", "times" ...)
- Clarrify the the user's input/intent
- giving information to the user
- update the KB or database

We call this set of information the **D**alogue **A**ct (DA)



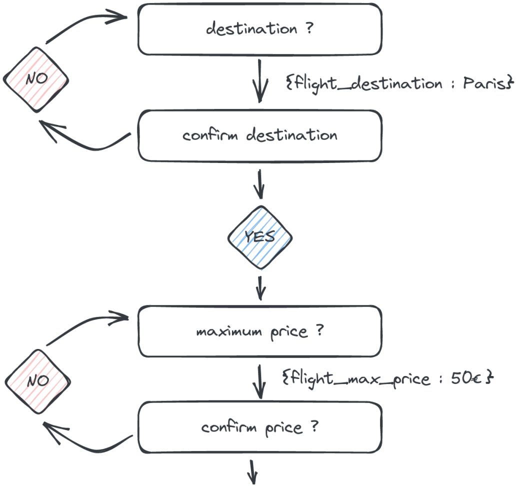
Finite state automata

- Nodes represent the question of the systems
- transition represent answers
- The graph specifies all legal dialogues

Pros and cons

- Limited number of interactions
- Used in simple tasks and structured domain (few slots and number of values)
- Cannot manage complex dialogues states

(In the application we will use this kind of approach!!!)

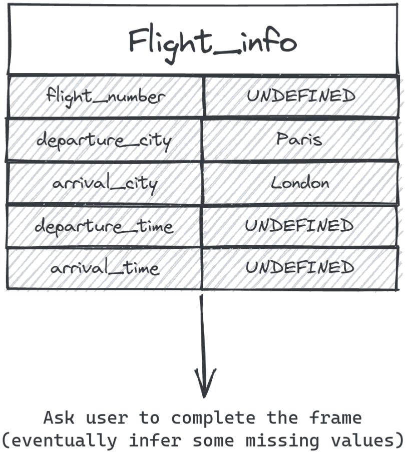


Frame-based approaches

- A frame is a set of slots (no constraint on the utterance order of the user)
- Possible to have many frames as substasks and more complex frames

How does it works

- A frame represents what the system has to solve
- The slots are filled with values from user utterance



Reinforcement Learning
Learning to take decision/action

- Take a decision given a (dialogue) state
- State are modelled with
  - **MDP:** Markov Decision Process [LPE00]
  - **POMDP:** Partially Observable Markov Decision Process [You06]
- use a reinforce strategy, dialogue policy optimization

Pros and Cons

- Not necessarly need to review all possibilities
- Data is needed

Neural Approaches
Relatively a new kind of approach [WVM$^+$16]

- Learn to map history/dialogue slots (on previous and current utterance) to an answer
- Hybrid approaches [WAZ17, HLLK19]
- Fully generative approach (feed with history) [YLQ21](transformer)

Pros and Cons
Mostly the same as RL approaches

- Not necessarly need to review all possibilities
- Huge amount of data is needed to train (particularly for transformer)

DM and Dialogue Control : Strengths and weakness

Rule based approaches
Often for Automata and frame based approaches

Strength

- Easy to interpret
- Few data needed

Those approaches are mostly used in industry

Weakness

- May be difficult to maintain
- Long and time consuming to maintaining rules (how to address most of possibilities) for complex systems

Statistical and Neural approaches
For RL and Neural Networks

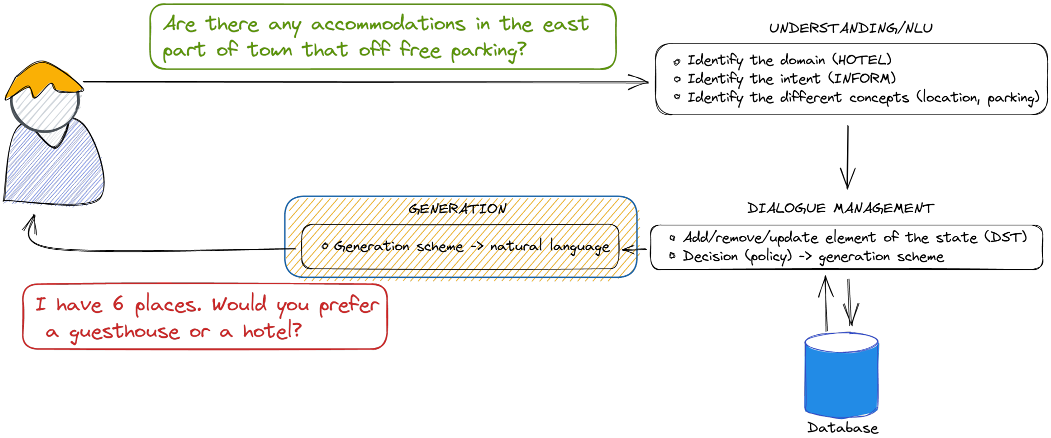
Strength

- Based on data (more natural response)
- Do not need to give explicitly the different cases

Weakness

- Data needed (huge amount particularly for end to end)
- Debugging and interpretation is often difficult

Generate


Are there any accommodations in the east part of town that off free parking?
UNDERSTANDING/NLU
Identify the domain (HOTEL)
Identify the intent (INFORM)
Identify the different concepts (location, parking)
DIALOGUE MANAGEMENT
Add/remove/update element of the state (DST)
Decision (policy) -> generation scheme
GENERATION
Generation scheme -> natural language
I have 6 places. Would you prefer a guesthouse or a hotel?
Database

Objectives:
Transform information returned by DM and generate an answer. Decide how to present the newt utterances words.

- Take DM informations such next requested informations or confirmation (asked previously by user)
- Decide how and what to say

Different approaches:

- Template based :
  Exemple : You have a train leaving at $confirm_departure_value (using information of the frame)
- Statistical Models such as Neural Networks such as Transformer models [PZL+20] (using DA)

Different ressources

Corpus and dataset (non-exhaustive)

- **ATIS:** Airline Travel Information Systems (exists in multi lingual settings)
- **MultiWOZ:** a multi domain dataset (accomodation, taxi, restaurant etc..) for booking or requesting informations
- **Microsoft Dialogue Challenge:** Multi-domain (movie-ticket, restaurant, taxi)
- **SNIPS:** Multi-domain (alarm, weather...)
- **PVDial** : Medical dialogue (in french) [a]

[a] https://pvdial.limsi.fr/

Libraries and software

- RASA (create a conversationnal systems) [a]
- ConvLab2, Convlab3 [b]

[a] https://rasa.com/blog/attention-dialogue-and-learning-reusable-patterns/
[b] https://github.com/ConvLab/ConvLab-3

A small reminder

Task-oriented systems:
A task oriented dialogues syste is a complex system that aims to provide specific informations an a subset of domains. In most cases it rely on a KB or a relationnal database where to find/update informations. The pipeline require different components:

- (ASR) **A**utomatic **S**peech **R**ecognition
- (NLU) **N**atural **L**anguage **U**nderstanding
- (DM) **D**ialogue **M**anagement (state tracking and dialogue control)
- (NLG) **N**atural **L**anguage **G**eneration
- (TS) **T**ext to **S**peech

Approaches:
For each componenents a variety of approaches exists such as rules with explicit domain knowledge or statistical and neural approaches(generative/extractive).

Chit-Chat and Purely Conversational Systems

Chit-chat Bot

Objectives:

- General conversation abilities
- Generate most appropriate reaction given the context and the previous user utterance

What is needed:

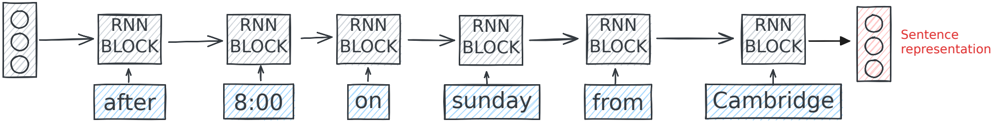
- Huge collection of dialogs

[HUMAN:]	Hey!
[BOT:]	hello there how are you? do you like dogs?
[HUMAN:]	Good ! I like dogs, I grew up in a farm
[BOT:]	that is cool. i've a german shepard.
[HUMAN:]	what is your dog's name ?
[BOT:]	his name is pepper. do you have any pets?
[HUMAN:]	I used to have a dog but he died.....
[BOT:]	that is so sad! do you watch nhl hockey?
[HUMAN:]	no... not a big fan of sport ?
[BOT:]	no, i love watching it. i love the nhl

Figure 9: Extract of a dialogue for the Second Conversational Intelligence Challenge [DLM+19]

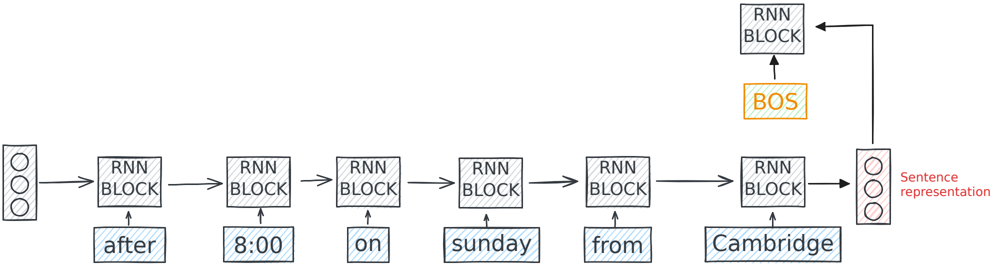
Sequence to Sequence

Model a sequence to sequence task with RNN :
→ Use *"many to one"* variant to encode the input sequence
→ Use *"one to many"* variant to generate a new sequence



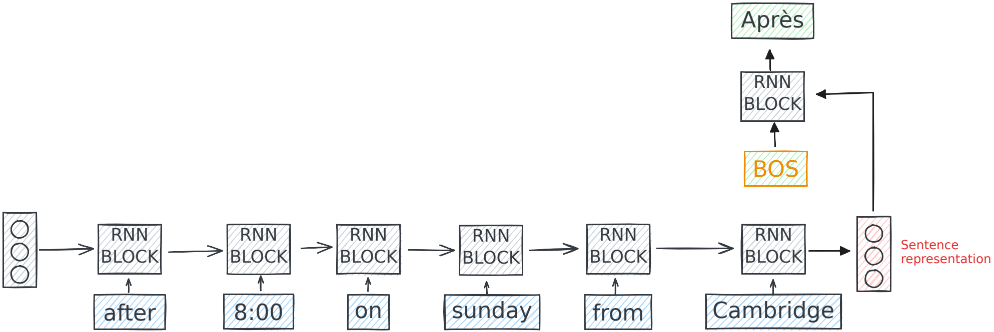
Sequence to Sequence

Model a sequence to sequence task with RNN :
$\rightarrow$ Use *"many to one"* variant to encode the input sequence
$\rightarrow$ Use *"one to many"* variant to generate a new sequence



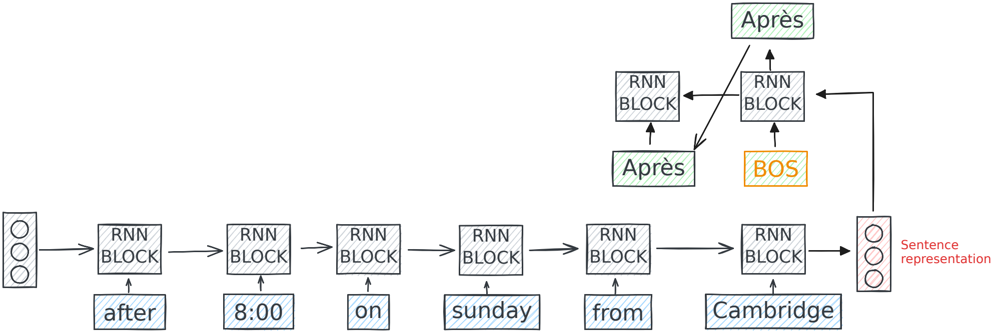
Sequence to Sequence

Model a sequence to sequence task with RNN :
$\rightarrow$ Use *"many to one"* variant to encode the input sequence
$\rightarrow$ Use *"one to many"* variant to generate a new sequence



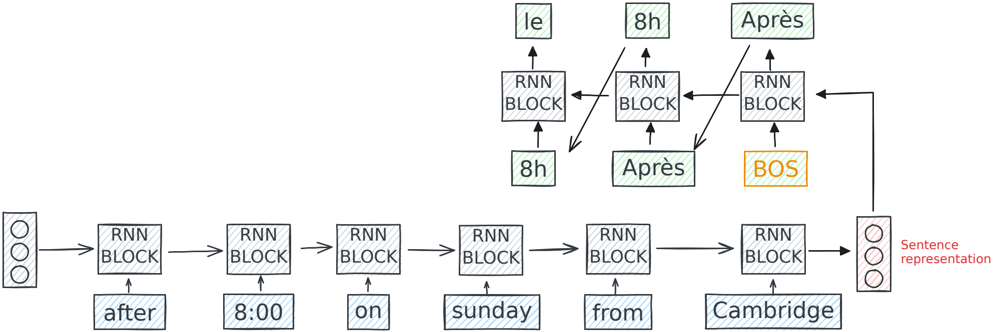
Sequence to Sequence

Model a sequence to sequence task with RNN :
→ Use *"many to one"* variant to encode the input sequence
→ Use *"one to many"* variant to generate a new sequence



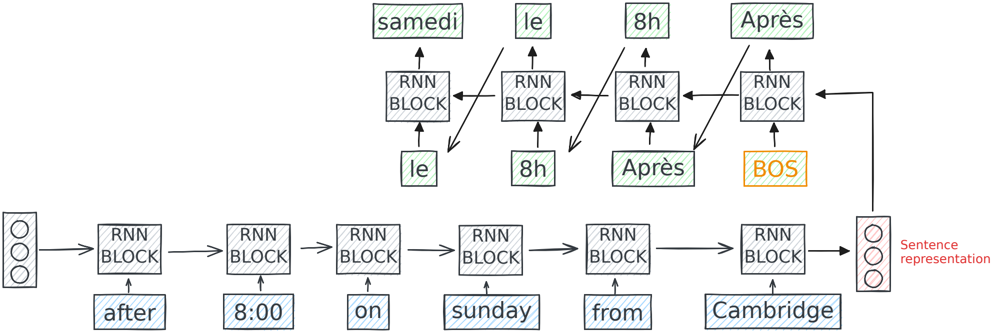
Sequence to Sequence

Model a sequence to sequence task with RNN :
→ Use *"many to one"* variant to encode the input sequence
→ Use *"one to many"* variant to generate a new sequence



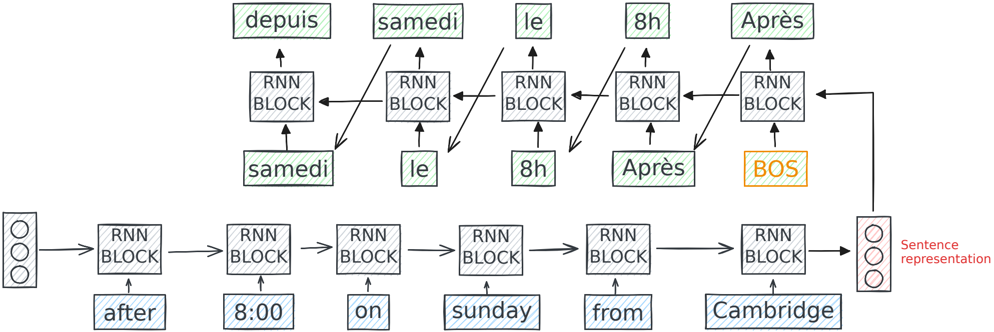
Sequence to Sequence

Model a sequence to sequence task with RNN :
→ Use *"many to one"* variant to encode the input sequence
→ Use *"one to many"* variant to generate a new sequence



Sequence to Sequence

Model a sequence to sequence task with RNN :
→ Use *"many to one"* variant to encode the input sequence
→ Use *"one to many"* variant to generate a new sequence

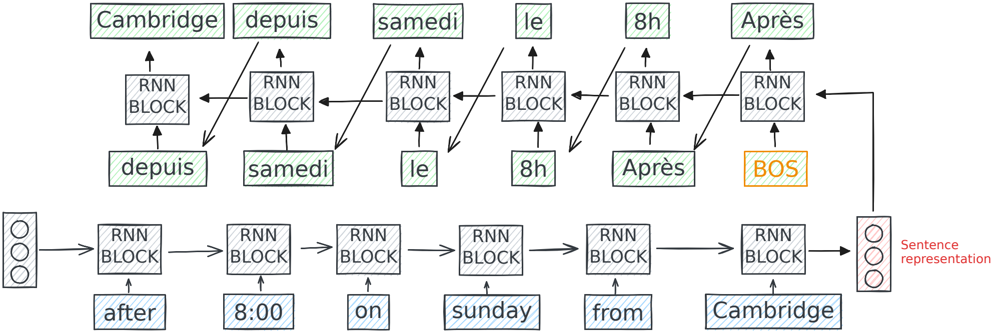


Sequence to Sequence

Model a sequence to sequence task with RNN :

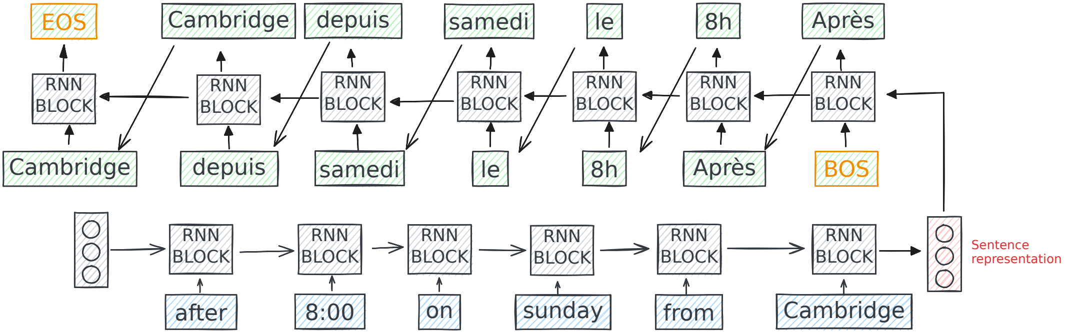
→ Use "*many to one*" variant to encode the input sequence

→ Use "*one to many*" variant to generate a new sequence



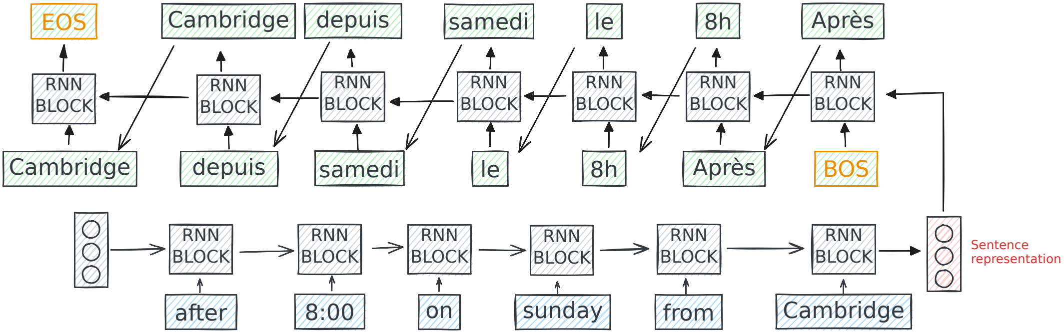
Sequence to Sequence

Model a sequence to sequence task with RNN :
→ Use “*many to one*” variant to encode the input sequence
→ Use “*one to many*” variant to generate a new sequence



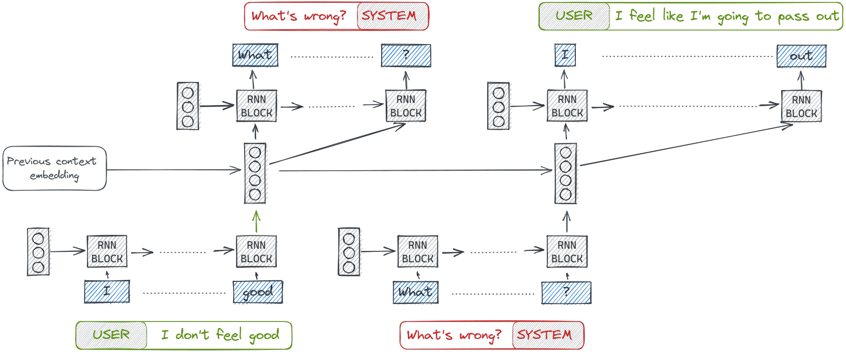
Sequence to Sequence

Model a sequence to sequence task with RNN :
→ Use "*many to one*" variant to encode the input sequence
→ Use "*one to many*" variant to generate a new sequence



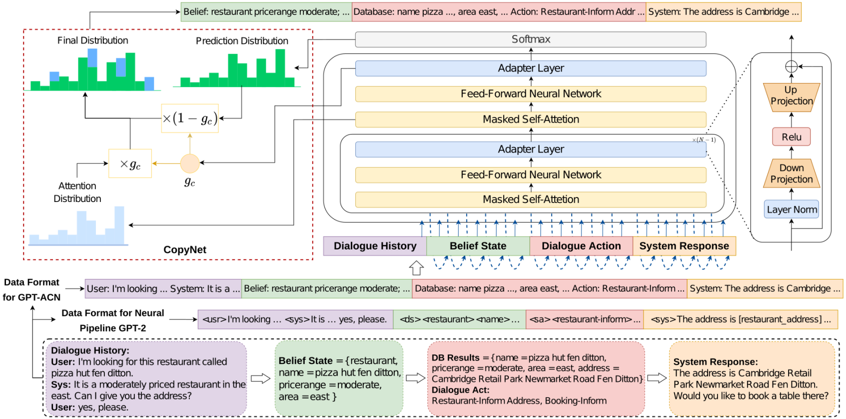
NB: The EOS word is a special word indicating the End Of the Sentence

Recurrent generative approach:
Essentially it is about learning to generate a sentence based on a given input sentence (and its context) [SSB+16]



Using generative approaches for dialogue systems ?

$\rightarrow$ Feeding the generator with context



NLU/State/DA evaluation:
Does retrieved domain, intents and slots type/values correspond to what is expected?

- **Domain, Intent** Multiclass classification (mono-label) :

$$accuracy = \frac{correctly\ predicted}{number\ of\ predictions}$$

- **slot type/values** Multiclass classification (multi-label)
  - **precision/recall**

$$precision = \frac{TP}{TP + FP}$$

$$recall = \frac{TP}{TP + FN}$$

  - **F1-score**

$$F1 = 2\frac{precision \times recall}{precision + recall}$$

for non categorical slot value use exact match? /use partial match?

Comparing the ground truth to generated text

$\rightarrow$ Evaluate the number similar words/tokens

$\rightarrow$ Evaluate the number of words tuples (n-grams approaches)

Comparing the ground truth to generated text

$\rightarrow$ Evaluate the number similar words/tokens

$\rightarrow$ Evaluate the number of words tuples (n-grams approaches)

ROUGE Approaches

ROUGE approaches relies on different n-grams comparisons.

ROUGE-1 $\rightarrow \frac{common\_word}{total\_unigrams}$

ROUGE-2 $\rightarrow \frac{common\_bigrams}{total\_bigrams}$

ROUGE-L $\rightarrow \frac{longuest\_common\_subsequence}{total\_unigrams}$

Comparing the ground truth to generated text

→ Evaluate the number similar words/tokens

→ Evaluate the number of words tuples (n-grams approaches)

Limit

→ "*listen to what we say*"

→ "*Just hear us out*"

Both sentences should be similar !

Comparing the ground truth to generated text

→ Evaluate the number similar words/tokens

→ Evaluate the number of words tuples (n-grams approaches)

Limit
→ "*listen to what we say*"

→ "*Just hear us out*"

Both sentences should be similar ! → no common words

Comparing the ground truth to generated text
$\rightarrow$ Take into account semantic

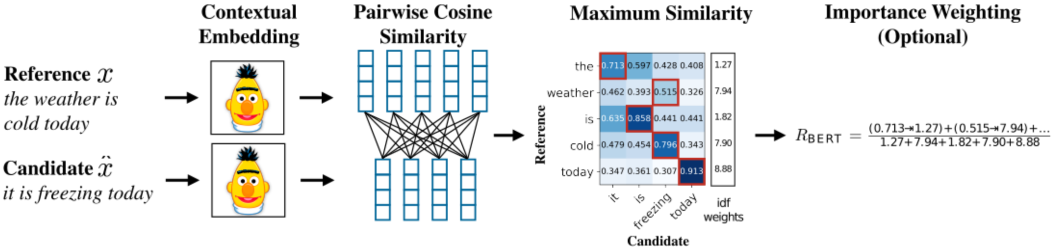
Evaluate Generation

Comparing the ground truth to generated text
$\rightarrow$ Take into account semantic

$\rightarrow$ Use embedding based approaches (BERTscore)

Comparing the ground truth to generated text
→ Take into account semantic

→ Use embedding based approaches (BERTscore)



Automatic evaluation
Evaluating simalirity with the ground truth :

- **On tokens/words level:** Counting the number of similar words or "N-grams" (e.g BLEU, ROUGE, METEOR)
- **On embeddings level** : Embed text and compare embeddings (may takes into account synonyms) such as the BERTScore
- **Entity** (between ground truth and generation)

Human evaluation
An example of evaluations approaches [Tho21]

- Fluency
- adequacy/informativness
- grammatical correctness

Exercise

Exercise

NLU System

- ATIS/SNIPS/MEDIA dataset
- Detecting intent
- Slot filling

Generative System on MULTIWoZ

- Test a transformer model for generation
- Propose improvements

Create a joint Intent/slot filling system
Develop an RNN architecture for both classification of intent and the classification/extraction of slots

- Using word embeddings train with the model
- Using pretrained words embedding

`https://gitlab.dsi.universite-paris-saclay.fr/thomas.gerald/textminingandchatbot/-/tree/main/tp/TP-1?ref_type=heads`

References i

- Tom Ayoola, Shubhi Tyagi, Joseph Fisher, Christos Christodoulopoulos, and Andrea Pierleoni, *Refined: An efficient zero-shot-capable approach to end-to-end entity linking*.
- Florian Baud and Alex Aussem, *Répondre aux requêtes des étudiants avec un agent conversationnel à mémoire supervisée*, 2023.
- Yun-Nung Vivian Chen, Dilek Hakkani-Tür, Gokhan Tur, Jianfeng Gao, and Li Deng, *End-to-end memory networks with knowledge carryover for multi-turn spoken language understanding*, Proceedings of The 17th Annual Meeting of the International Speech Communication Association (INTERSPEECH 2016), ISCA, June 2016.
- Guan-Lin Chao and Ian R. Lane, *BERT-DST: scalable end-to-end dialogue state tracking with bidirectional encoder representations from transformer*, CoRR **abs/1907.03040** (2019).

- Oralie Cattan, Sophie Rosset, and Christophe Servan, *On the cross-lingual transferability of multilingual prototypical models across NLU tasks*, Proceedings of the 1st Workshop on Meta Learning and Its Applications to Natural Language Processing (Online), Association for Computational Linguistics, August 2021, pp. 36–43.
- Emily Dinan, Varvara Logacheva, Valentin Malykh, Alexander Miller, Kurt Shuster, Jack Urbanek, Douwe Kiela, Arthur Szlam, Iulian Serban, Ryan Lowe, Shrimai Prabhumoye, Alan W Black, Alexander Rudnicky, Jason Williams, Joelle Pineau, Mikhail Burtsev, and Jason Weston, *The second conversational intelligence challenge (convai2)*, 2019.
- Renato De Mori, *Spoken language understanding: Systems for extracting semantic information from speech*, pp. 9 – 40, 03 2011.
- James Glass, Giovanni Flammia, David Goodine, Michael Phillips, Joseph Polifroni, Shinsuke Sakai, Stephanie Seneff, and Victor Zue, *Multilingual spoken-language understanding in the mit voyager system*, Speech Communication **17** (1995), no. 1, 1–18.

Sahar Ghannay, Christophe Servan, and Sophie Rosset, *Neural networks approaches focused on French spoken language understanding: application to the MEDIA evaluation task*, Proceedings of the 28th International Conference on Computational Linguistics (Barcelona, Spain (Online)), International Committee on Computational Linguistics, December 2020, pp. 2722–2727.

Matthew Henderson, *Machine learning for dialog state tracking: A review*, Proceedings of The First International Workshop on Machine Learning in Spoken Language Processing, 2015.

Chiori Hori, Takaaki Hori, and Shinji Watanabe, *Context sensitive spoken language understanding using role dependent lstm layers*, 2015.

Jiyeon Ham, Soohyun Lim, Kyeng-Hun Lee, and Kee-Eung Kim, *Extensions to hybrid code networks for fair dialog dataset*, Computer Speech & Language **53** (2019), 80–91.

- Sameer Khurana, Antoine Laurent, and James Glass, *SAMU-XLSR: Semantically-aligned multimodal utterance-level cross-lingual speech representation*, IEEE Journal of Selected Topics in Signal Processing **16** (2022), no. 6, 1493–1504.
- Leonardo Campillos Llanos, Dhouha Bouamor, Pierre Zweigenbaum, and Sophie Rosset, *Managing linguistic and terminological variation in a medical dialogue system*, Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16) (Portorož, Slovenia), European Language Resources Association (ELRA), May 2016, pp. 3167–3173.
- Chia-Hsuan Lee, Hao Cheng, and Mari Ostendorf, *Dialogue state tracking with a language model using schema-driven prompting*, Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (Online and Punta Cana, Dominican Republic), Association for Computational Linguistics, November 2021, pp. 4937–4949.

- Stefan Larson and Kevin Leach, *A survey of intent classification and slot-filling datasets for task-oriented dialog*, 2022.
- E. Levin, R. Pieraccini, and W. Eckert, *A stochastic model of human-machine interaction for learning dialog strategies*, IEEE Transactions on Speech and Audio Processing **8** (2000), no. 1, 11–23.
- Gaëlle Laperrière, Valentin Pelloin, Mickaël Rouvier, Themos Stafylakis, and Yannick Estève, *On the use of semantically-aligned speech representations for spoken language understanding*, 2022.
- Grégoire Mesnil, Yann Dauphin, Kaisheng Yao, Yoshua Bengio, Li Deng, Dilek Hakkani-Tur, Xiaodong He, Larry Heck, Gokhan Tur, Dong Yu, and Geoffrey Zweig, *Using recurrent neural networks for slot filling in spoken language understanding*, IEEE/ACM Transactions on Audio, Speech, and Language Processing **23** (2015), no. 3, 530–539.

- Baolin Peng, Chenguang Zhu, Chunyuan Li, Xiujun Li, Jinchao Li, Michael Zeng, and Jianfeng Gao, *Few-shot natural language generation for task-oriented dialog*, Findings of the Association for Computational Linguistics: EMNLP 2020 (Online), Association for Computational Linguistics, November 2020, pp. 172–182.
- Iulian V. Serban, Alessandro Sordoni, Yoshua Bengio, Aaron Courville, and Joelle Pineau, *Building end-to-end dialogue systems using generative hierarchical neural network models*, Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, AAAI'16, AAAI Press, 2016, p. 3776–3783.
- Thórhildur Thorleiksdóttir, *Understanding human potentials for evaluating generative models*, 2021.
- Jeanne villaneau and Antoine Jean-Yves, *Categorial grammars used to partial parsing of spoken language*, 2004.

- Mathilde Veron, Sophie Rosset, Olivier Galibert, and Guillaume Bernard, *Evaluate on-the-job learning dialogue systems and a case study for natural language understanding*, 2021.
- Jason D. Williams, Kavosh Asadi, and Geoffrey Zweig, *Hybrid code networks: practical and efficient end-to-end dialog control with supervised and reinforcement learning*, Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (Vancouver, Canada), Association for Computational Linguistics, July 2017, pp. 665–677.
- Di Wu, Liang Ding, Fan Lu, and Jian Xie, *SlotRefine: A fast non-autoregressive model for joint intent detection and slot filling*, Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP) (Online), Association for Computational Linguistics, November 2020, pp. 1932–1937.

- Chien-Sheng Wu, Steven C.H. Hoi, Richard Socher, and Caiming Xiong, *TOD-BERT: Pre-trained natural language understanding for task-oriented dialogue*, Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP) (Online), Association for Computational Linguistics, November 2020, pp. 917–929.
- Tsung-Hsien Wen, David Vandyke, Nikola Mrksic, Milica Gasic, Lina M. Rojas-Barahona, Pei-Hao Su, Stefan Ultes, and Steve Young, *A network-based end-to-end trainable task-oriented dialogue system*, 2016.
- Yunyi Yang, Yunhao Li, and Xiaojun Quan, *UBAR: towards fully end-to-end task-oriented dialog system with GPT-2*, AAAI, AAAI Press, 2021, pp. 14230–14238.
- Steve J. Young, *Using pomdps for dialog management*, 2006 IEEE Spoken Language Technology Workshop (2006), 8–13.

Kaisheng Yao, Baolin Peng, Yu Zhang, Dong Yu, Geoffrey Zweig, and Yangyang Shi, *Spoken language understanding using long short-term memory neural networks*, 2014 IEEE Spoken Language Technology Workshop (SLT), 2014, pp. 189–194.