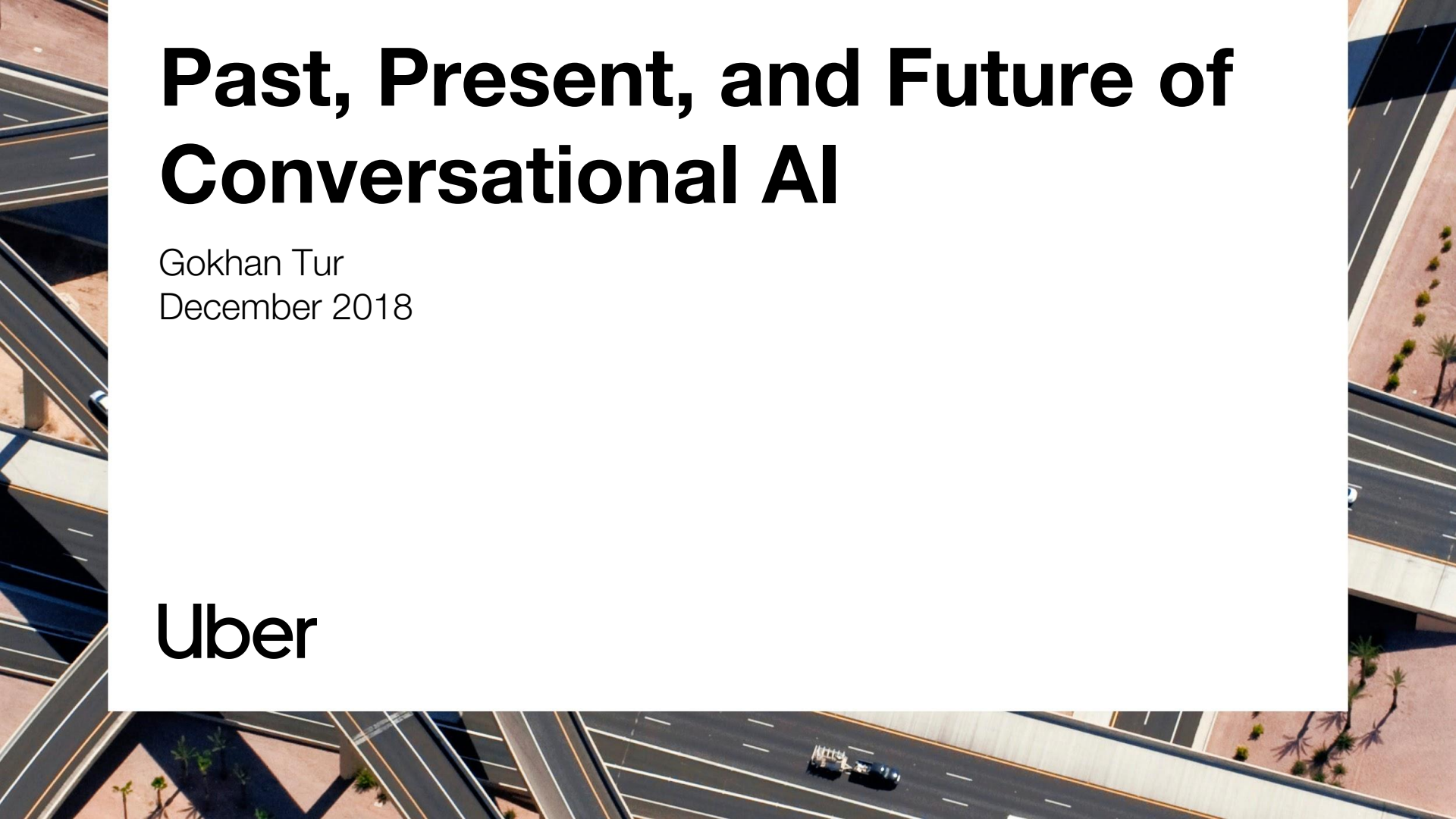


Past, Present, and Future of Conversational AI

Gokhan Tur
December 2018

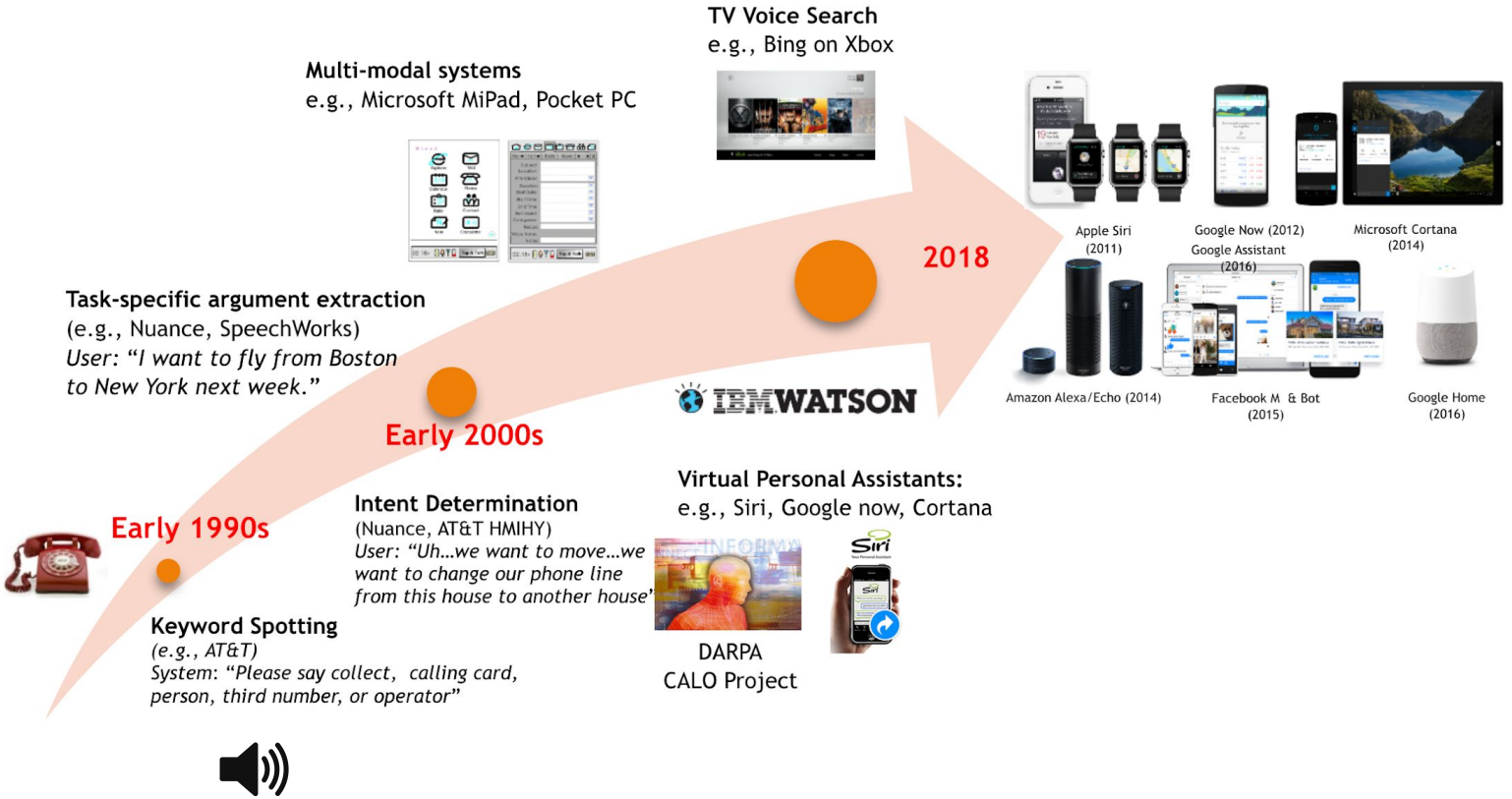
Uber

Language Understanding

"*At the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted*" (Turing 1950)

"(Semantic) Meaning is a holy grail for linguistics and philosophy" (Jackendoff 2002)

Origins of Conversational AI Systems



Origins of Language Understanding

- **Symbolic**:
  - **Symbolic AI:** semantic graph
  - **Classical NLP**: syntactic parsing with semantics
- **Data Driven**:
  - **Semantic Search**: query understanding with knowledge graph
  - **Speech**: call routing to dialogue
  - **Statistical NLP**: statistical semantic parsing
  - **ML**: yet another nail for my hammer

Case of ATIS

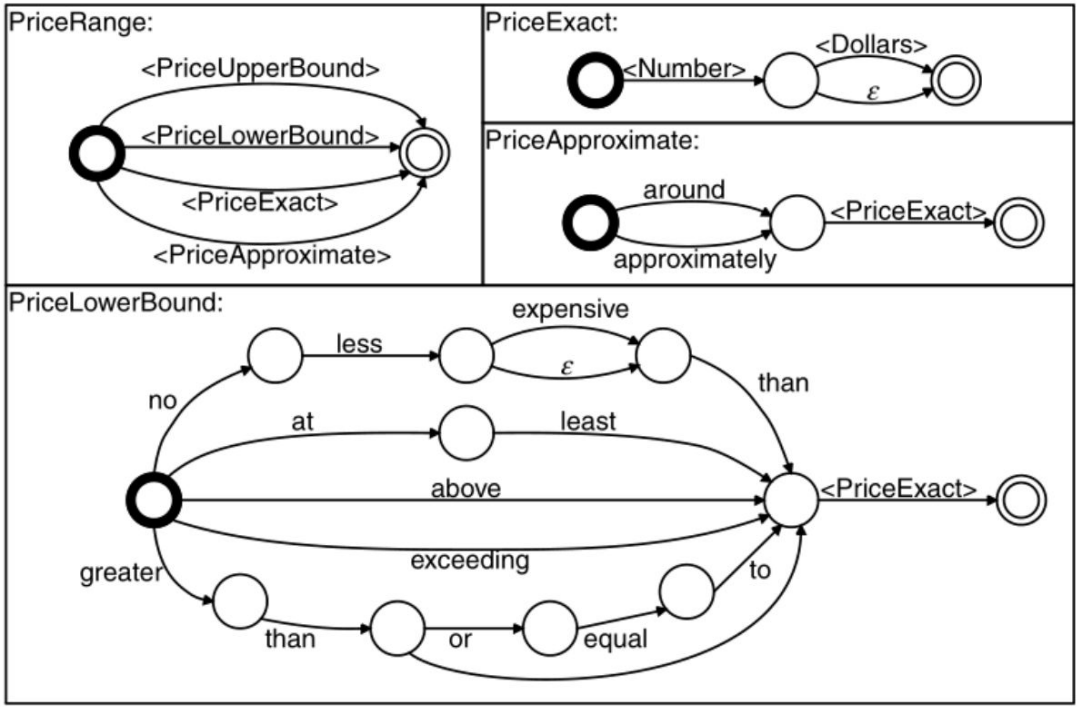
- Single turn flight information queries
- Task: Natural language input to SQL

"Please show me the flights to Boston on Monday."

Domain: Flights
Intent: Find Flights
Destination: Boston
Date: Monday

CMU Phoenix System (Ward 1991)

- Based on DYPAR grammar based parser
- 3.2K non-terminals, 13K grammar rules
- 1st place in the first ATIS evaluations



Classical NLP Approach

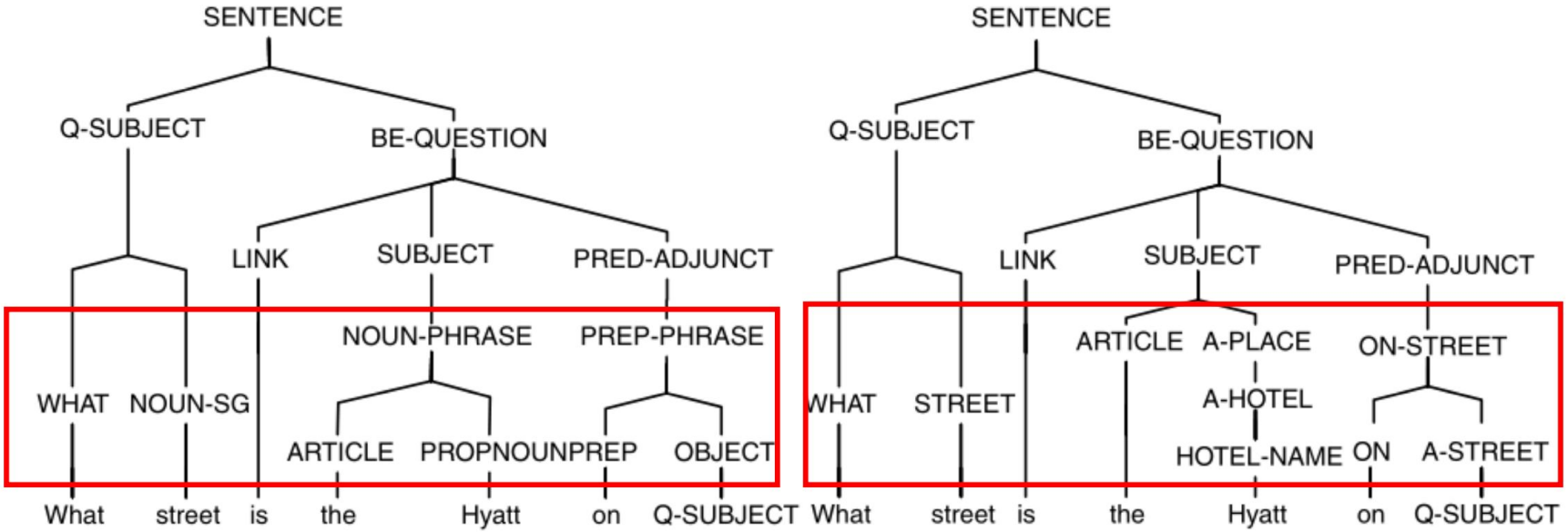


Figure 3.7 TINA parse tree with syntactic rules only (left) and with lower-level syntactic rules replaced by domain-dependent semantic rules (right) (The second tree is reproduced from Seneff (1992) (© 1992 Seneff))

Speech Origins of ConvAI:

Long Before Alexa (1922)

Speech Origins of ConvAI:

AT&T Router (1982)

This is AT&T. Please say collect, calling card, person to person, or third number.

Speech Origins of Language Understanding

"*Airplanes don't flap their wings*" Fred Jelinek

Speech Recognition

$\text{argmax}_W P(W|A) = \text{argmax}_W P(A|W) P(W)$

Modeled by HMM

Language Understanding

$\text{argmax}_M P(M|W) = \text{argmax}_M P(W|M) P(M)$

Can be modeled by HMM

(Pieraccini and Levin, Eurospeech'91)

AT&T Chronus System

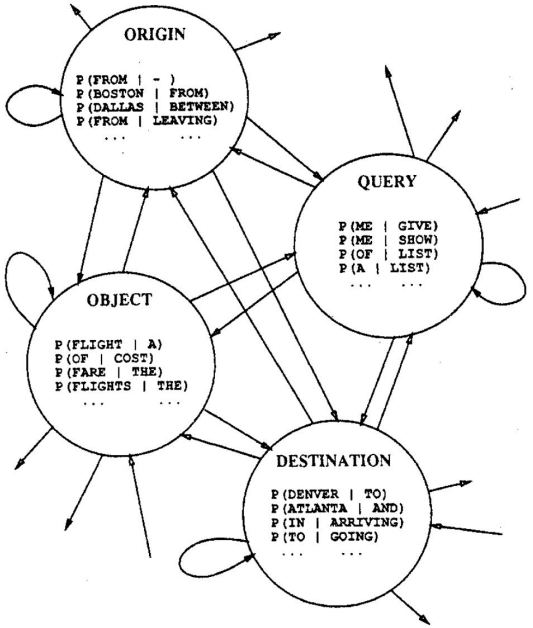


Figure 2: Language/conceptual model as an HMM

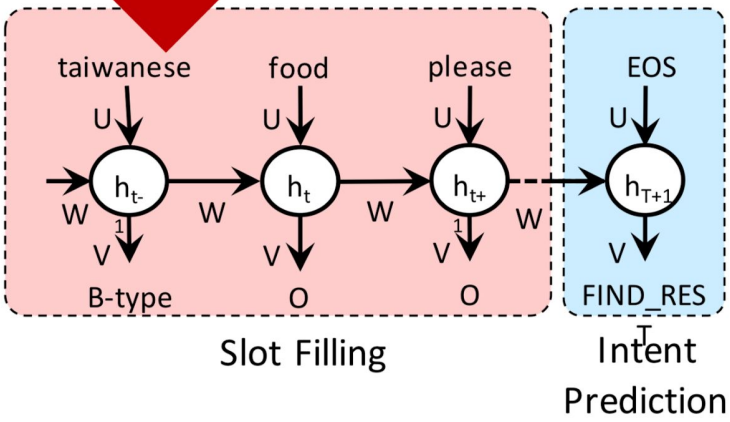
Language Understanding

- **Domain/Intent**
  - Naive Bayes (Gorin et al. 1997)
  - MaxEnt (Chelba, Mahajan, Acero; 2003)
  - SVM (Haffner, Tur, Wright; 2003)
  - Boosting (Gupta et al, 2005)
  - DNN (Sarikaya et al, 2011; Tur et al & Deng et al, 2012)
  - RNN (Ravuri & Stolcke, 2015)
- **Joint**
  - Recursive NN (Guo et al., 2014)
  - bLSTM (Hakkani-Tur et al., 2016)
  - Multi-Head Seq2Seq (Liu and Lane 2016)
- **Slots**
  - CFG (Ward, 1990)
  - pCFG (Seneff, 1992)
  - HMM (Pieraccini,Levin, 1991)
  - CRF (Wang et al, 2005)
  - NN-MM (Deoras et al., 2014)
  - Vanilla RNN (Mesnil et al 2013 & Yao et al, 2013)
  - LSTM/GRU-RNN (Yao et al, 2014)
  - Seq2Seq RNN (Kurata et al., 2016)

Language Understanding Modeling

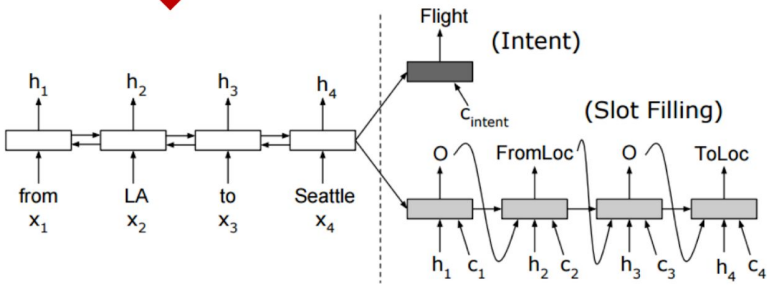
Sequence-based (Hakkani-Tur et al., 2016)

- Slot filling and intent prediction in the same output sequence

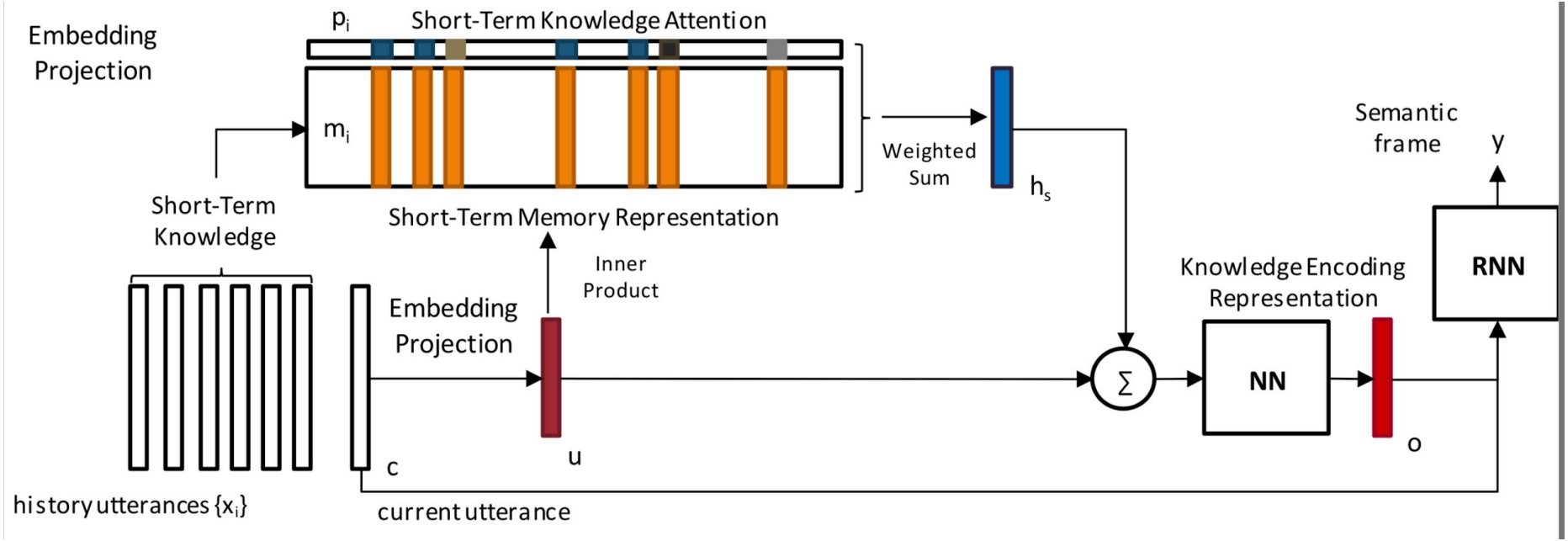


Parallel (Liu and Lane, 2016)

- Intent prediction and slot filling are performed in two branches

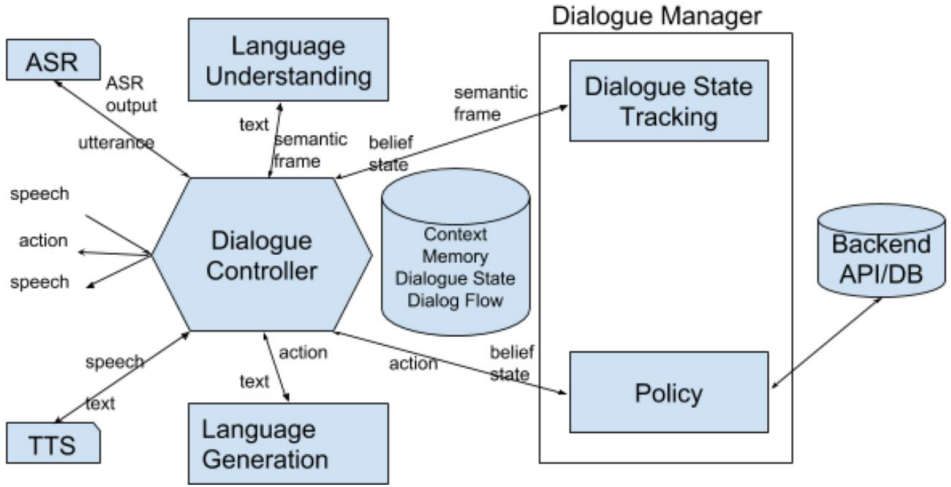


Contextual Understanding (Chen et al., 2016)

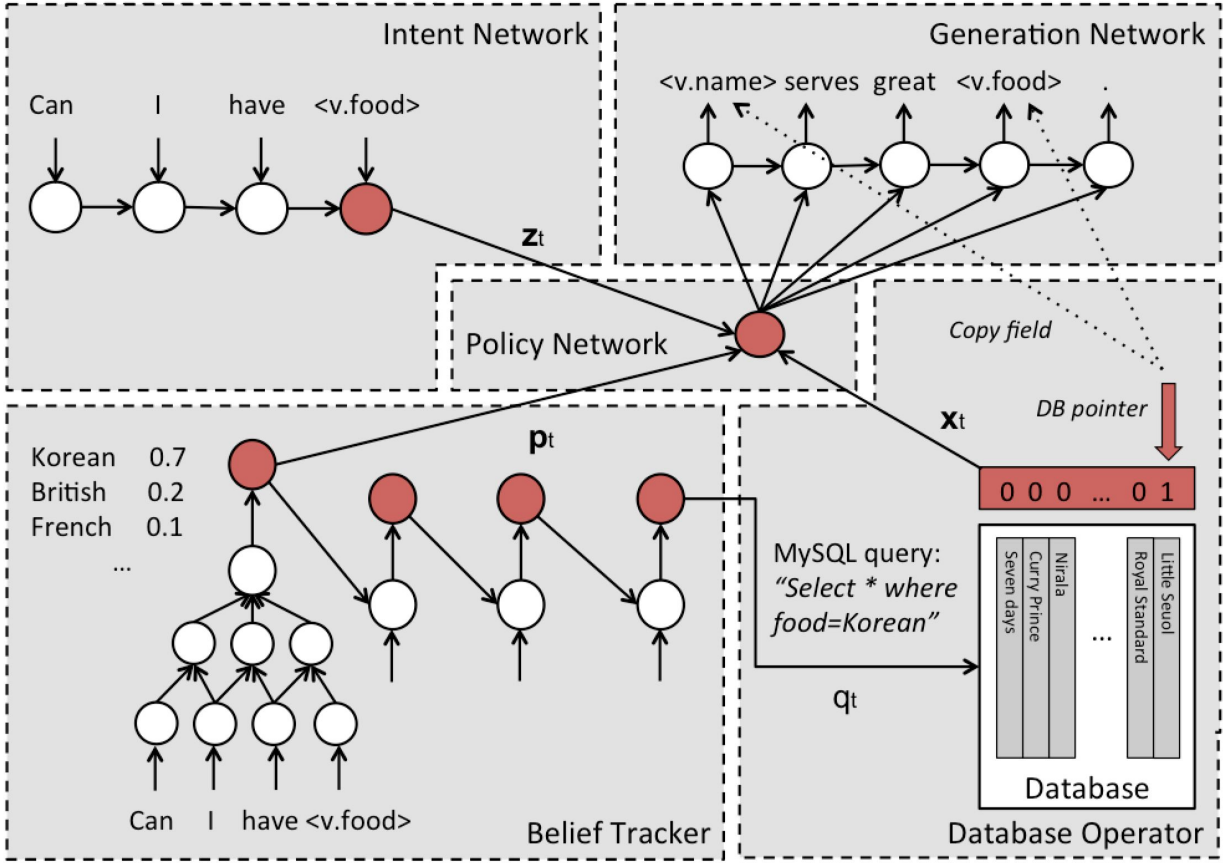


Dialogue Systems

- Turn 1: *Book a table at Il Fornaio for tomorrow*
  - **NLU**: reserve(restaurant_name: Il Fornaio, date:tomorrow)
  - **DST**: reserve(restaurant_name: Il Fornaio, date: tomorrow)
  - Backend requires number of people
  - **Policy**: request_info(num_people)
  - **NLG**: *For how many people?*
  -
- Turn 2: *2 people please for 6pm*
  - **NLU**: reserve(num_people: 2, time: 6pm)
  - **DST**: reserve(restaurant_name: Il Fornaio, date: tomorrow, time: 6pm, num_people: 2, time: 6pm)
  - Backend confirms
  - **Policy**: report(success)
  - **NLG**: *Successfully booked*

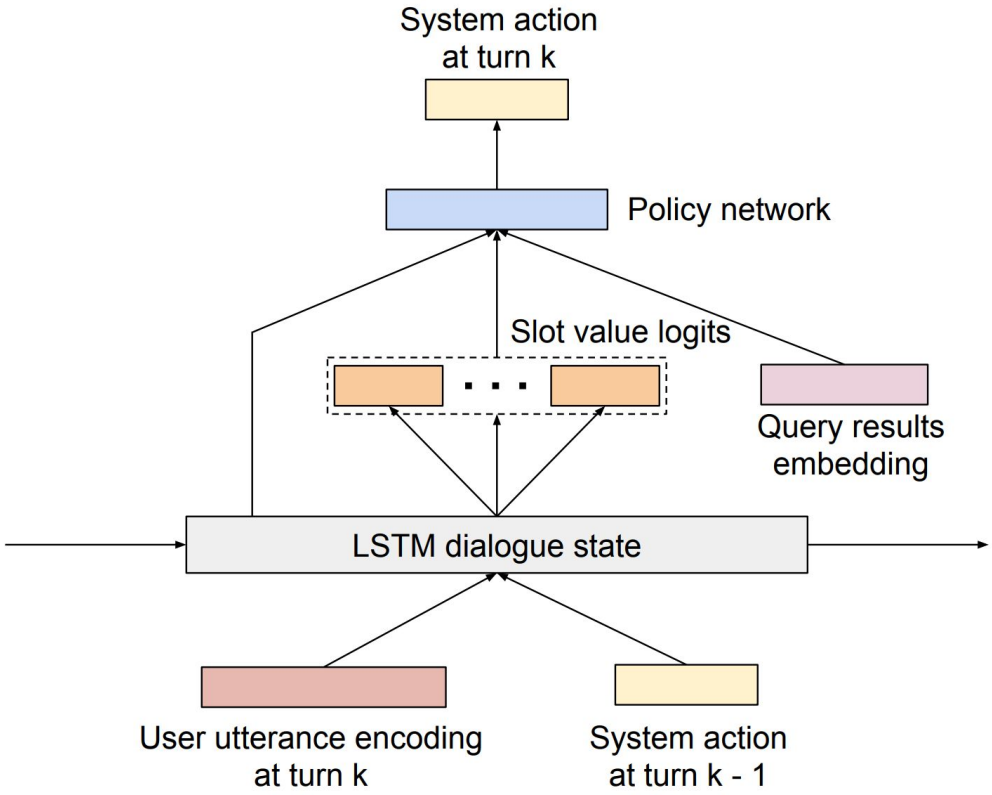


U. of Cambridge End-to-End (Wen et al., 2016)



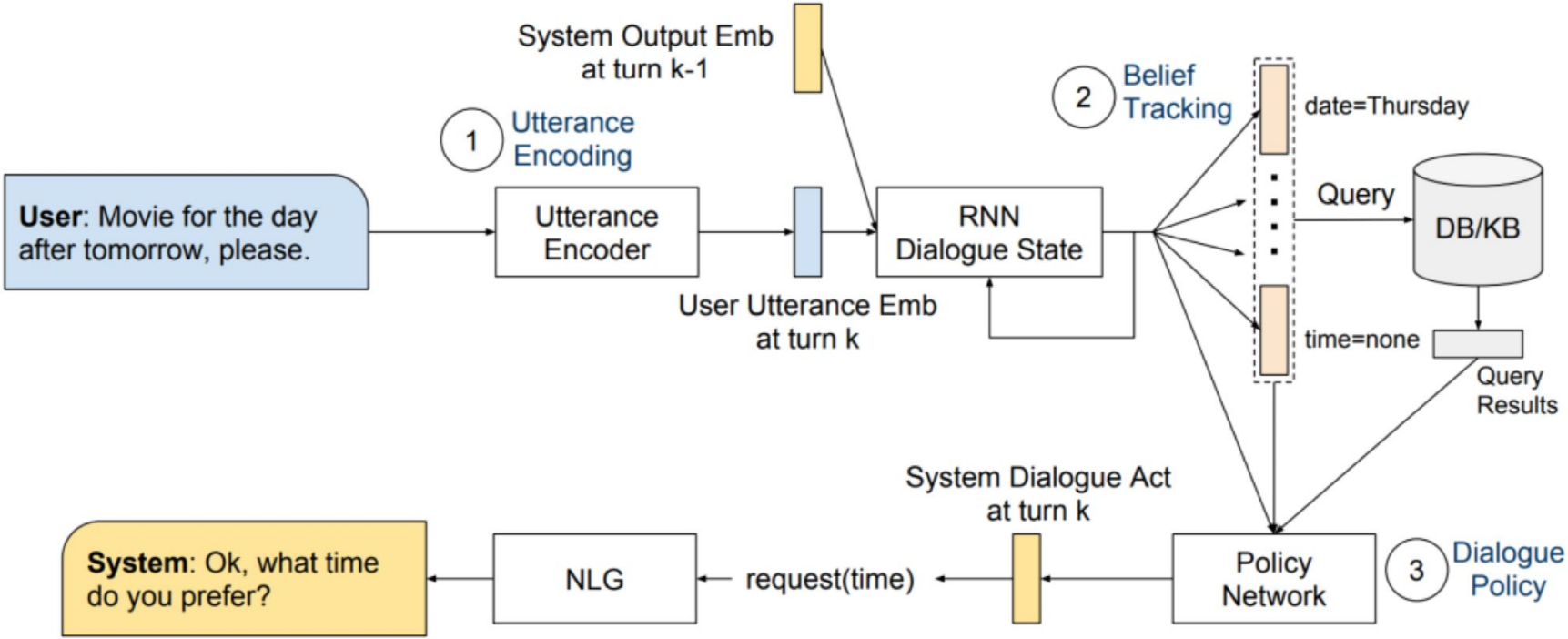
CMU End-to-End

(Liu and Lane, 2018)



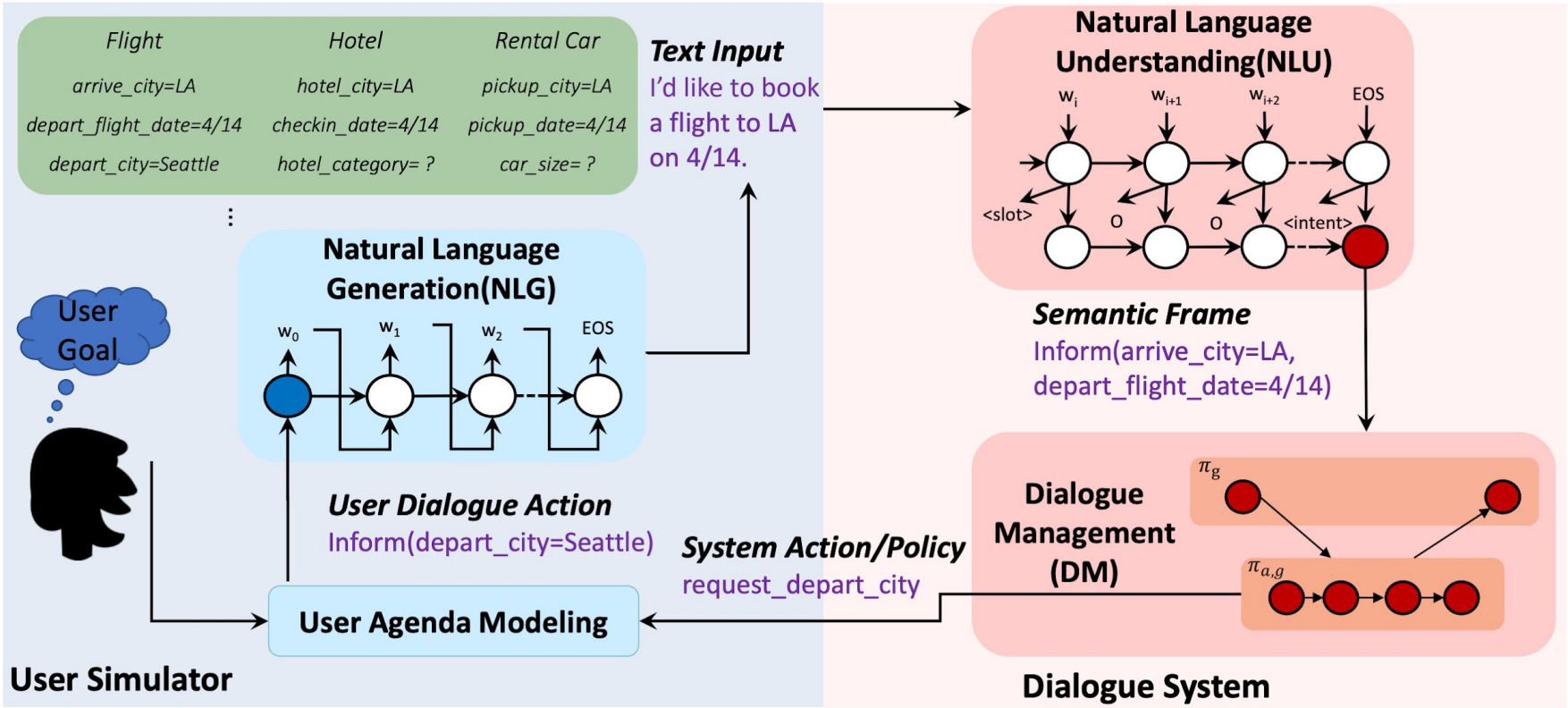
Google End-to-End

(Liu et al., 2018)



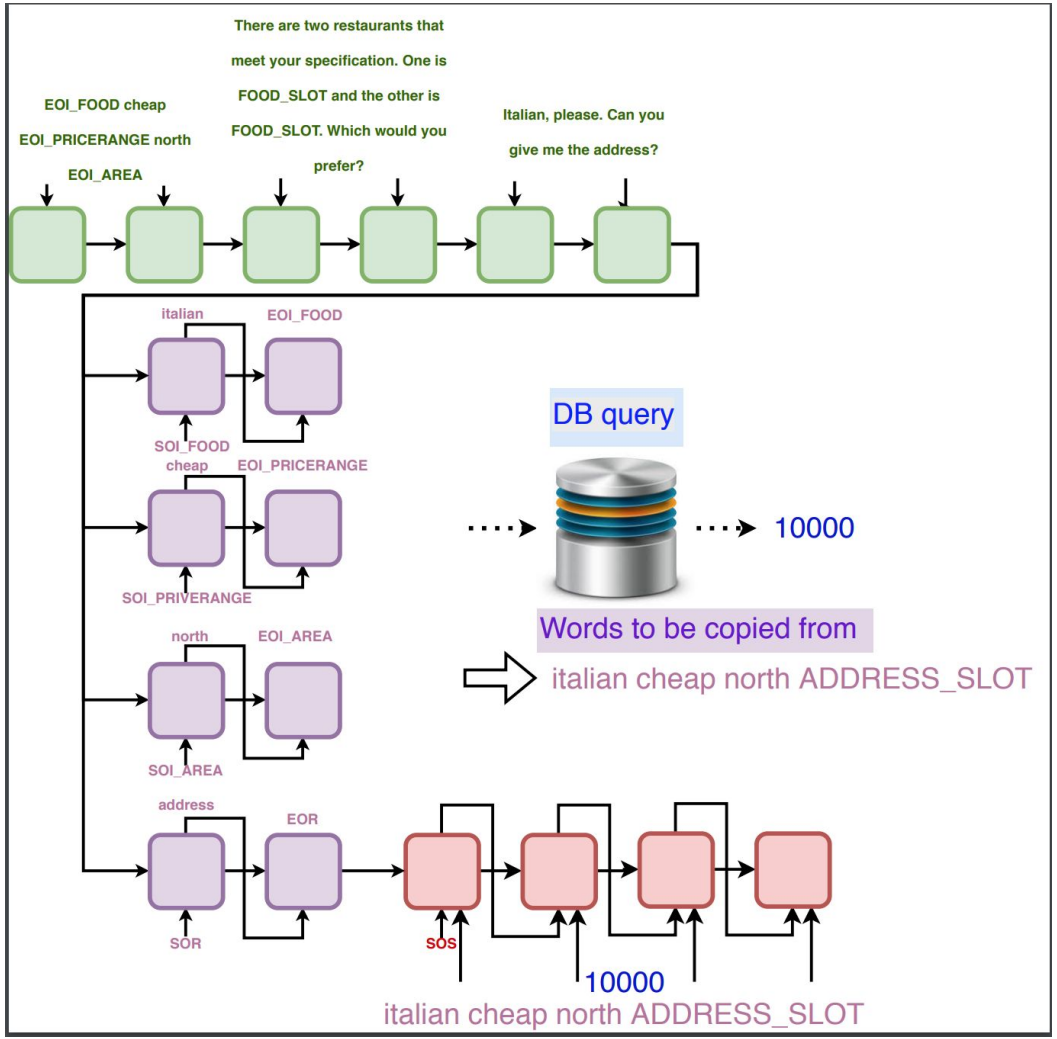
MSR End-to-End

(Li et al., 2018)



Uber End-to-End

(Shu et al. 2018)



What is the Number One Challenge in Conversational AI?

- Robustness to NL variability

Welcome to the Real World #1

What you build:

U: “I want to order a pizza”
S: “What is the size?
U: “Small”
S: “What is the topping?
U: “Cheese”
S: “All done”

What they actually expect:

U: “I want to order a pizza”

S: “What is the size?
U: “Do you make deep pan?”

S: “What is the topping?
U: “What kind of olives do you have?”

Welcome to the Real World #2

What you build:

U: “show me a cat picture”
U: “spell beautiful”
U: “what does modem mean”
U: “knock knock”
U: “set alarm for 7am”
U: “tell me a joke”
U: “tell me a story”
U: “show flights to Boston”
U: “show me directions to LAX”
U: “show tapa places nearby”

What they actually expect:

U: “show me a picture”
U: “can you spell a word for me”
U: “what does my name mean”
U: “knock knock knock”
U: “delete this alarm”
U: “tell this joke again to my wife”
U: “then what happened to the prince?”
U: “check-in to my flight”
U: “directions to my daughter’s music school”
U: “aren’t there any other tapa places nearby?”

Welcome to the Real World #3

What you build:

U: “on the way to my brother’s house I need to pick up some cheap wine that goes well with lasagna”

What they actually expect:

U: “My brother invited me to his house. I am thinking of bringing a bottle of wine”
S: “Sure, red or white?”
U: “Seems like we are having lasagna”
S: “White it is. Here are some choices along your way”

Welcome to the Real World #4

Colbert: … I don't want to search for anything! I want to write the show!
Siri: Searching the Web for "search for anything. I want to write the shuffle."
Colbert: … For the love of God, the cameras are on, give me something?
Siri: What kind of place are you looking for? Camera stores or churches?

What is the Problem with Existing Systems?

- They all mimic “understanding”, while learning the in-domain concepts towards “targeted understanding”.
- No world knowledge, no common sense, no reasoning, no grounding...

What is the Solution?

from Alyosha Efros

Machines Talking to Machines

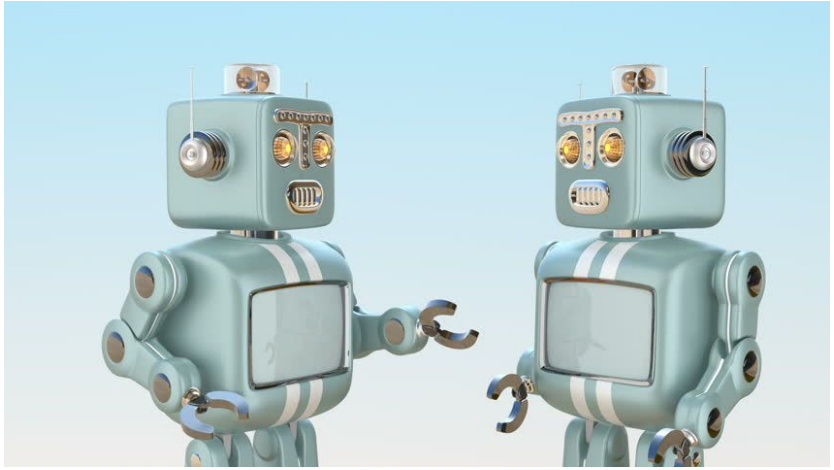
- Goal: generate natural and reasonable conversations for exploring the space using a user simulator (Hakkani-Tur 2016)
  - e.g., Google Duplex talking to Google Assistant
- Apply RL end to end!
- Approach: train a user simulator (e.g., Asri et al., 2016, Schatzmann et al., 2007)

User Goal: *date*=Friday, *num_tickets*=2, *theatre_name*=DontCare, *movie*=Sully, *time*=DontCare

0 -	SYSTEM	Hi, how can I help you? `greeting()`

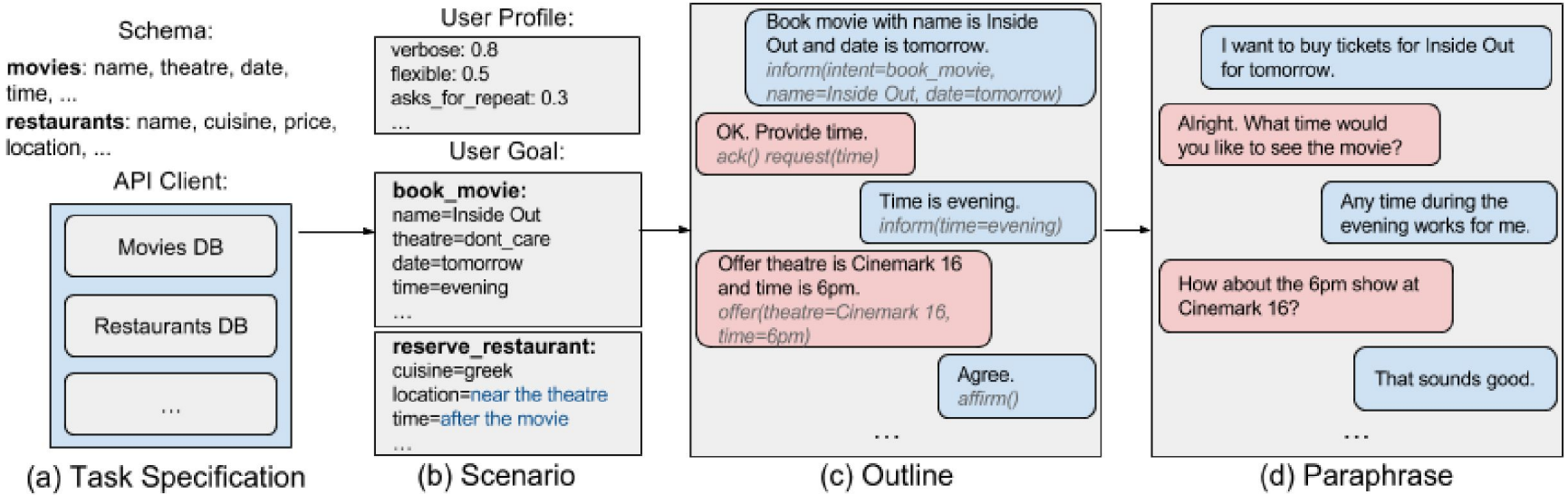
	USER	Hi, I'd like to buy tickets to see the Sully movie.
		`greeting() intent(buy_movie_tickets)`
		`inform(movie=Sully)`

1 -	SYSTEM	You wanna see Sully. How many tickets would you need? `confirm(movie=Sully) request(num_tickets)`
	USER	2 tickets please. `inform(num_tickets=2)` ...



Google Agenda Based Simulator

(Shah et al., 2018)



Alexa Speech Based User Simulator

(Fazel-Zarandi et al., 2017)

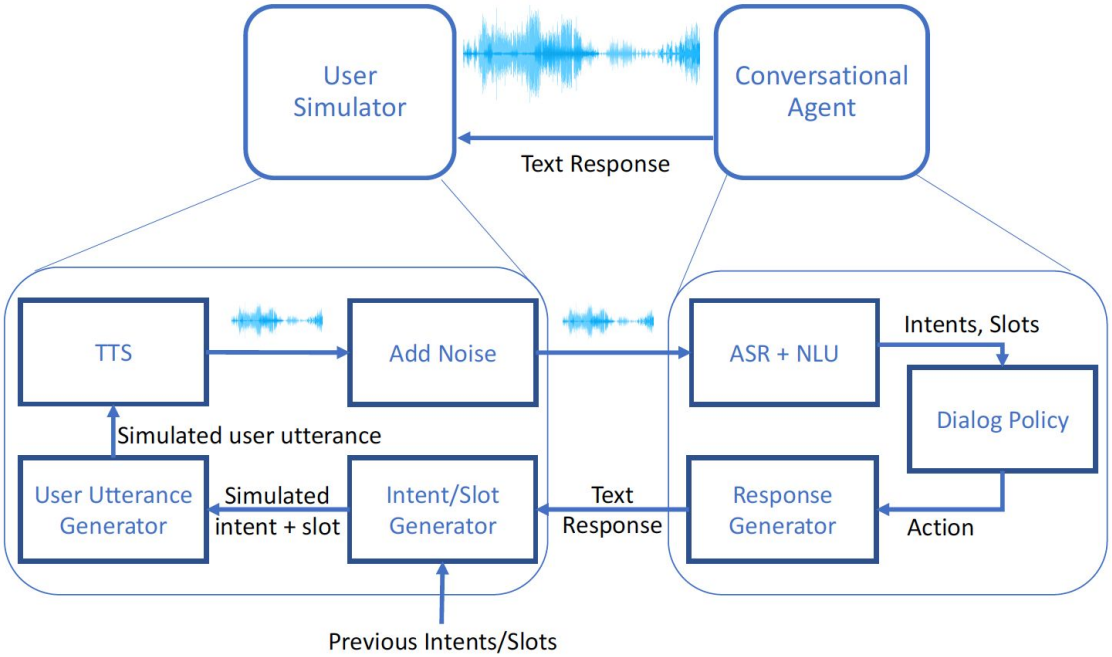
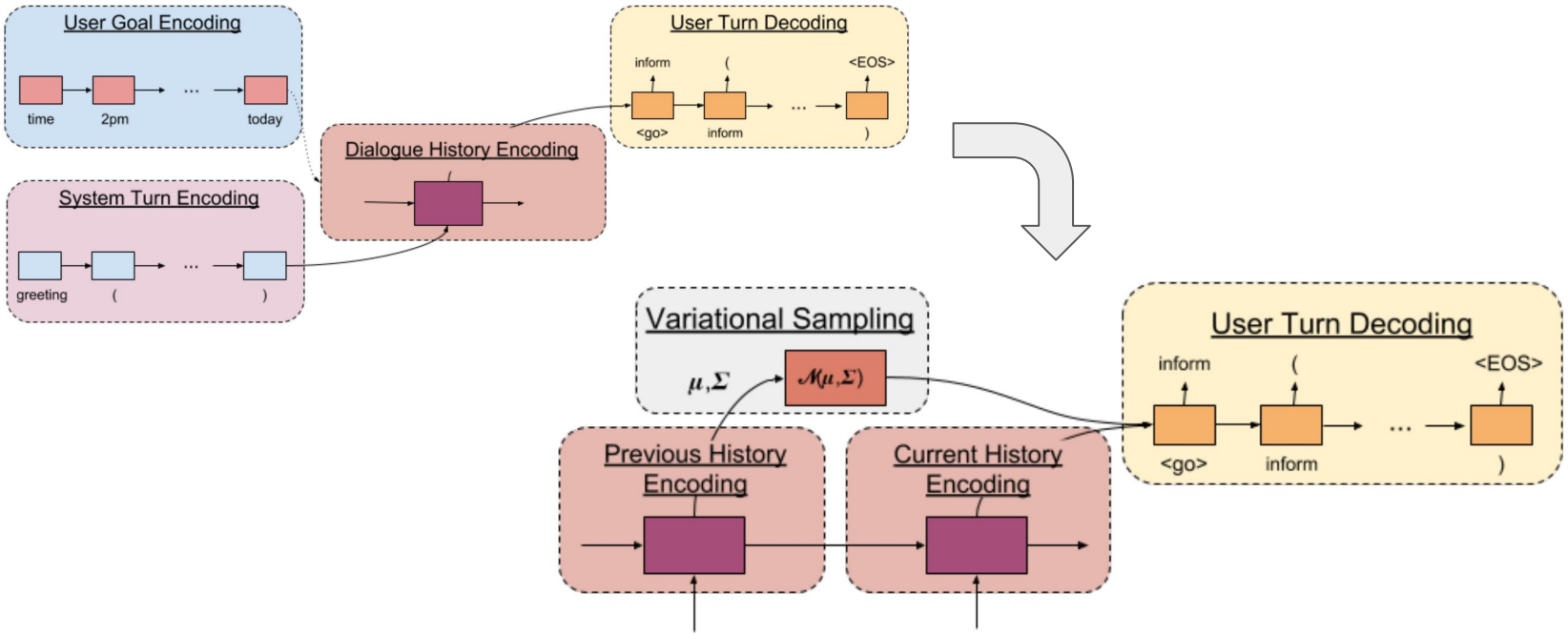


Figure 1: User simulator and conversational agent interaction. We use the text response generated by the agent before sending it to TTS to eliminate the need for ASR/NLU on the user simulator side.

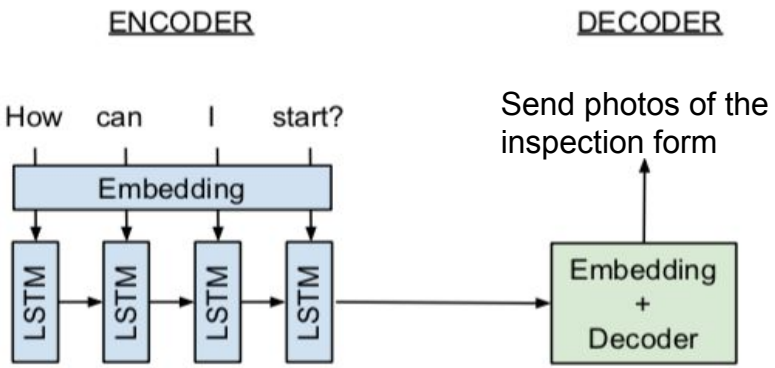
Model Based User Simulator

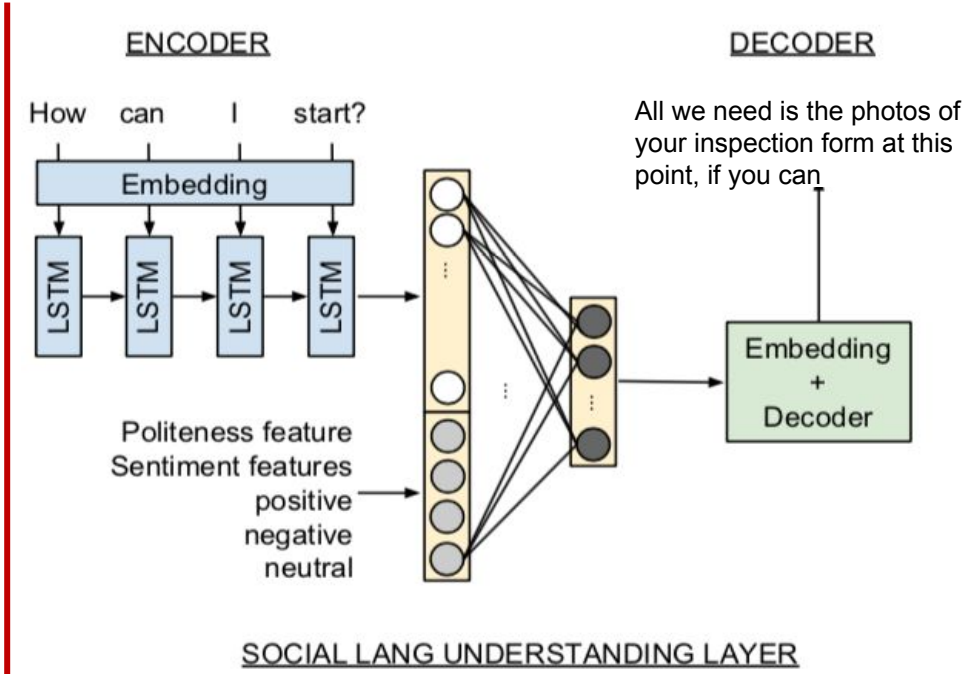
- How do achieve NL variability? (Gur et al., SLT 2018)



Social Language Generation

(Wang et al., 2018)





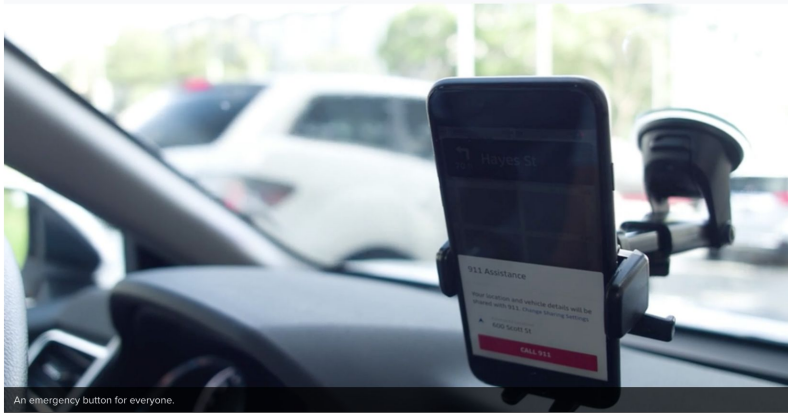
Thanks!

Mashable VIDEO ENTERTAINMENT CULTURE TECH SCIENCE SOCIAL GOOD SHOP MORE

Tech FOLLOW MA

Uber gives drivers voice control so they can keep their hands on the wheel

Share on Facebook Share on Twitter +



An emergency button for everyone.

Uber