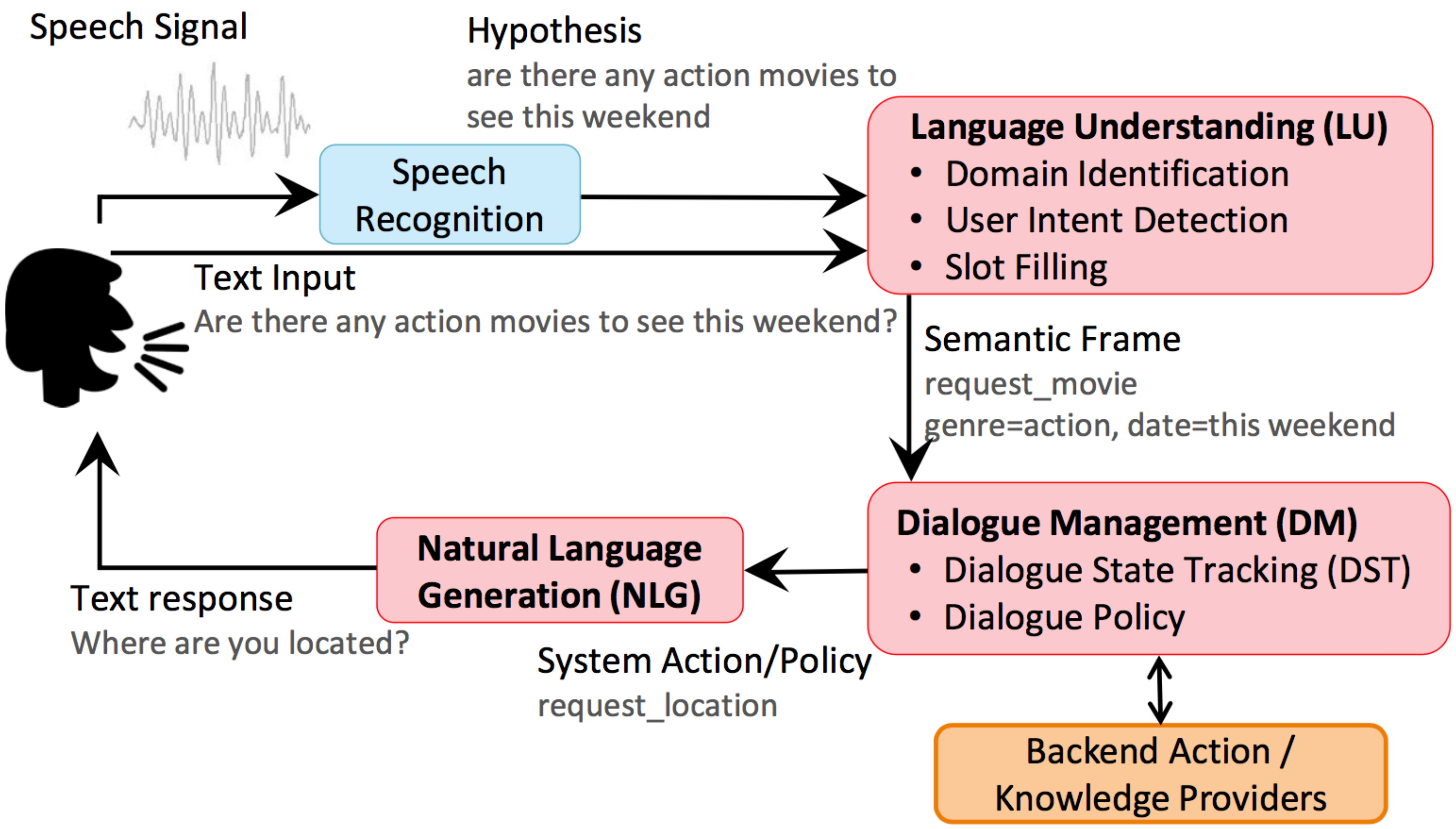
Task-oriented DS



Language Understanding

- Three modules:
  - Domain classification
  - Intent classification
  - Slot filling

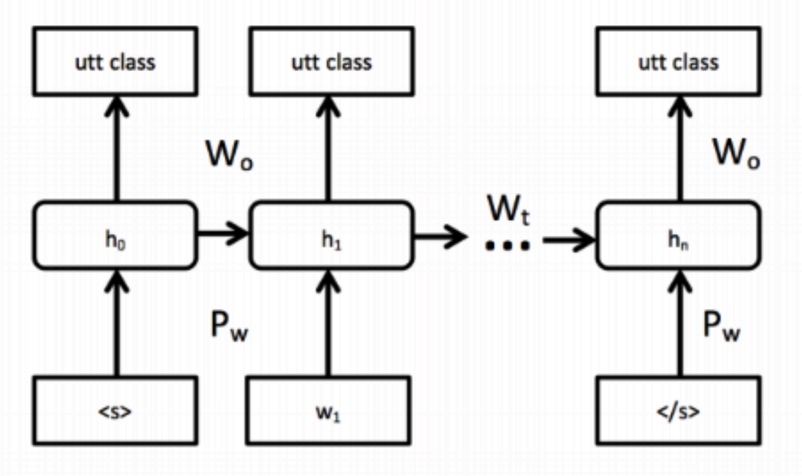
LU: Domain/Intent Classification

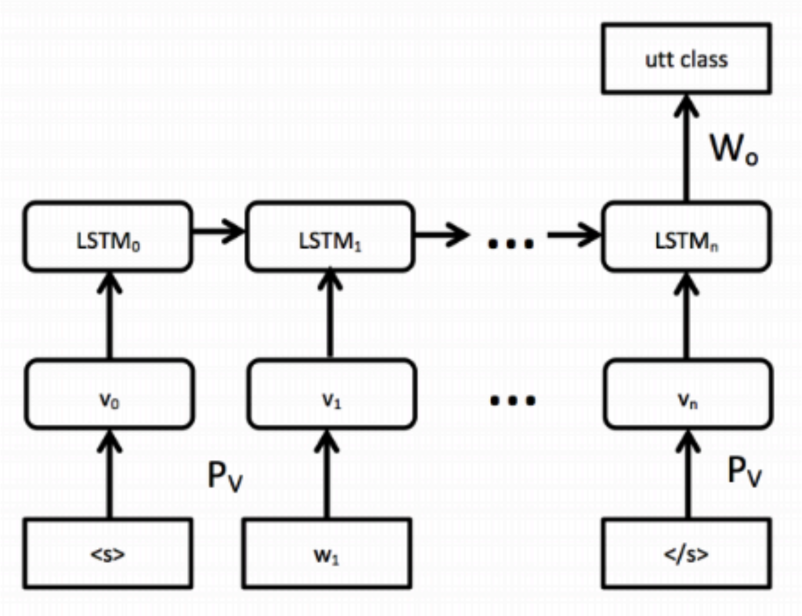
Find me a cheap Taiwanese restaurant in Oakland

Domain	Intent
Movies	find_movie, buy_tickets
Restaurants	find_restaurant, find_price, book_table
Music	find_lyrics, find_singer
Sports	…
…	

LU: Domain/Intent Classification

- Challenges:
  - addressee detection
  - Sparseness of n-gram
  - Large number of singletons
- Approaches:
  - RNN with LSTM to solve both domain/intent classification and addressee detection
  - Word-hashing to resolve singleton
    - Kat: #Ka, Kat, at#





LU: Slot Filling

Flights from Boston to New York today

	flights	from	Boston	to	New	York	today
Entity Tag	O	O	B-city	O	B-city	I-city	O
Slot Tag	O	O	B-dept	O	B-arrival	I-arrival	B-date

LU: Slot Filling

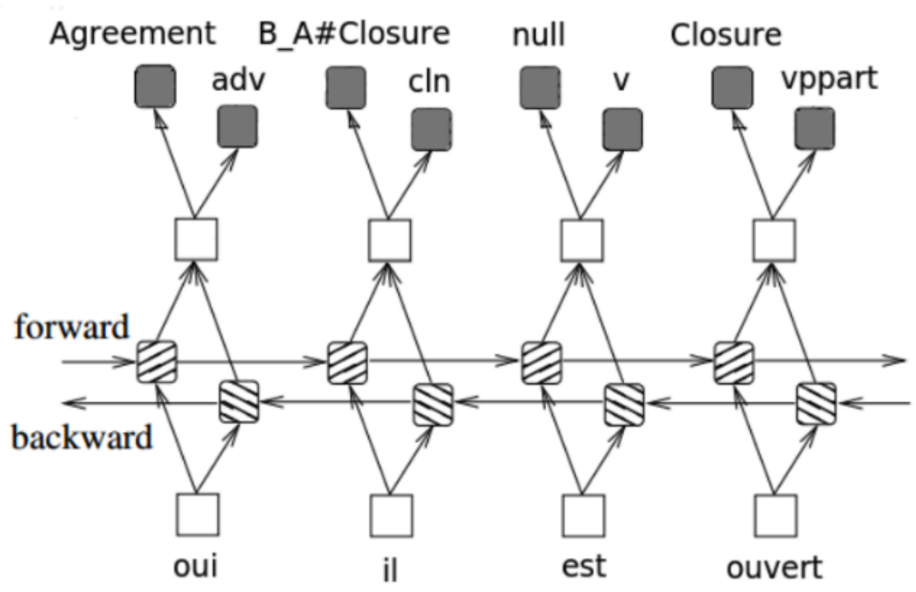
- Challenges:
  - to model dependencies between labels
  - To capture contextual information
- Approaches:
  - RNN with LSTM
  - LSTM-look around (the input is n-grams)
  - bLSTM
  - Encoder-decoder networks
  - Attention based encoder-decoder

LU: Joint-learning and Multi-domain

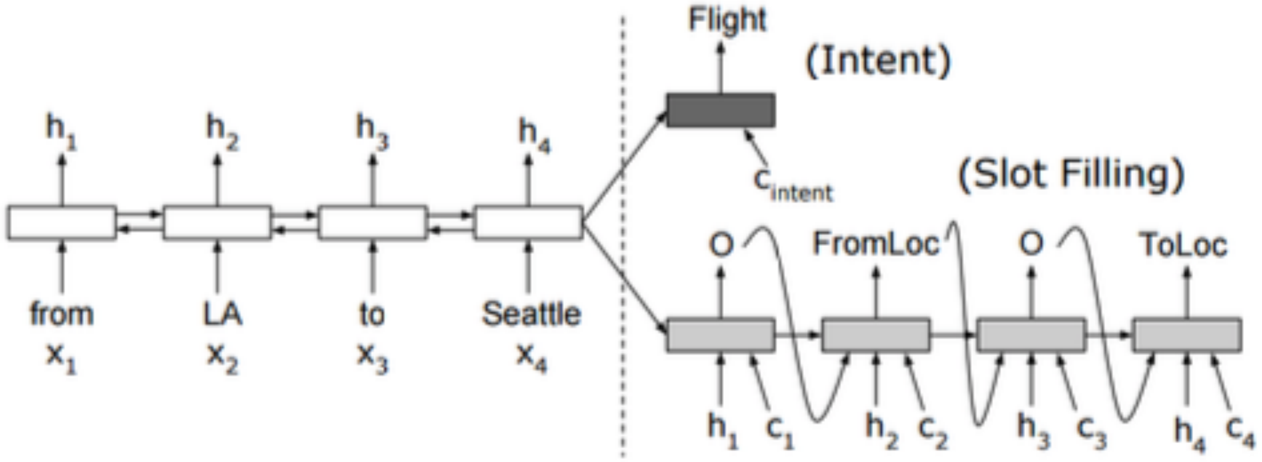
- Motivation:
    - to prevent error propagation in the pipeline approach
    - to reduce the number of training data required for each domain
- Challenges:
    - To use external information, e.g., dependency tree and parse tree
    - Unseen slot value

LU: Joint-learning and Multi-domain

- Multi-task bLSTM (POS, disfluency, NER, frame label)



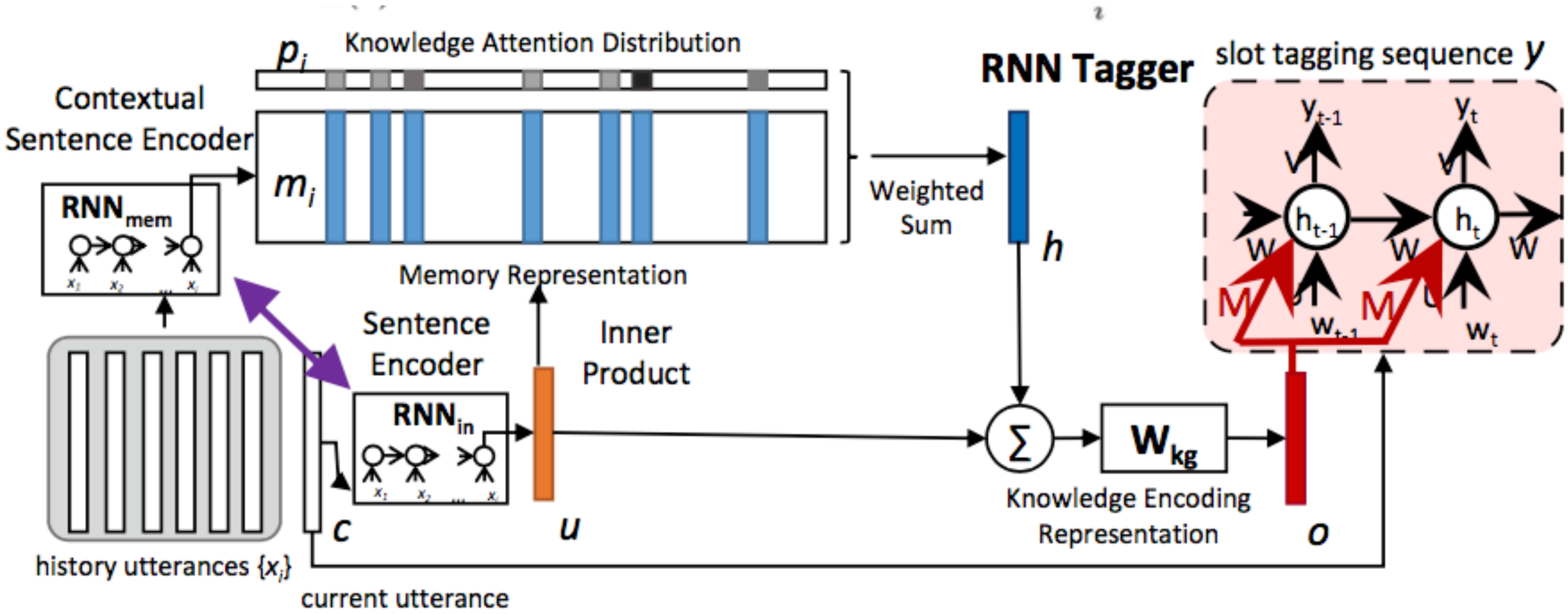
- Slot filling and intent prediction at the same times



LU: Contextual

- Motivation: many works exploit adjacency pair of utterances, not the history of dialogue
- Approaches:
    - LSTM over the whole dialogue
    - Knowledge guided attention network (memory network)

Knowledge guided attention network



Dialogue Management

- Dialogue State Tracking
- Dialogue Policy

DM: Dialogue State Tracking

S: Which part of town?
request(area)

U: A cheap place in the north
inform(area=north, pricerange=cheap)

0.8 inform(area=north), inform(pricerange=cheap)

0.1 inform(area=north)

area=north pricerange=cheap	0.7 area=north pricerange=cheap 0.1 area=north food=north_african 0.2 ()
method=byconstraints	0.9 byconstraints 0.1 none
requested=()	0.0 phone 0.0 address

DM: Dialogue State Tracking

- Challenges:
  - A DST that can work on many domains
- Approaches:
  - To train one generalized RNN model and then specialized it for each slot name

DM: Dialogue Policy

- Task: to guide what the system should say

- Motivation:
  - To develop a generic RL algorithm to learn dialogue policy for all domains
- Challenges:
  - Number of dialogues for training
  - Domain expertise
- Approaches:
  - RL algorithms with different reward (e.g. #turns maximized or minimized)
  - User simulation (to generate enough data using dialogue history)

Language Generation

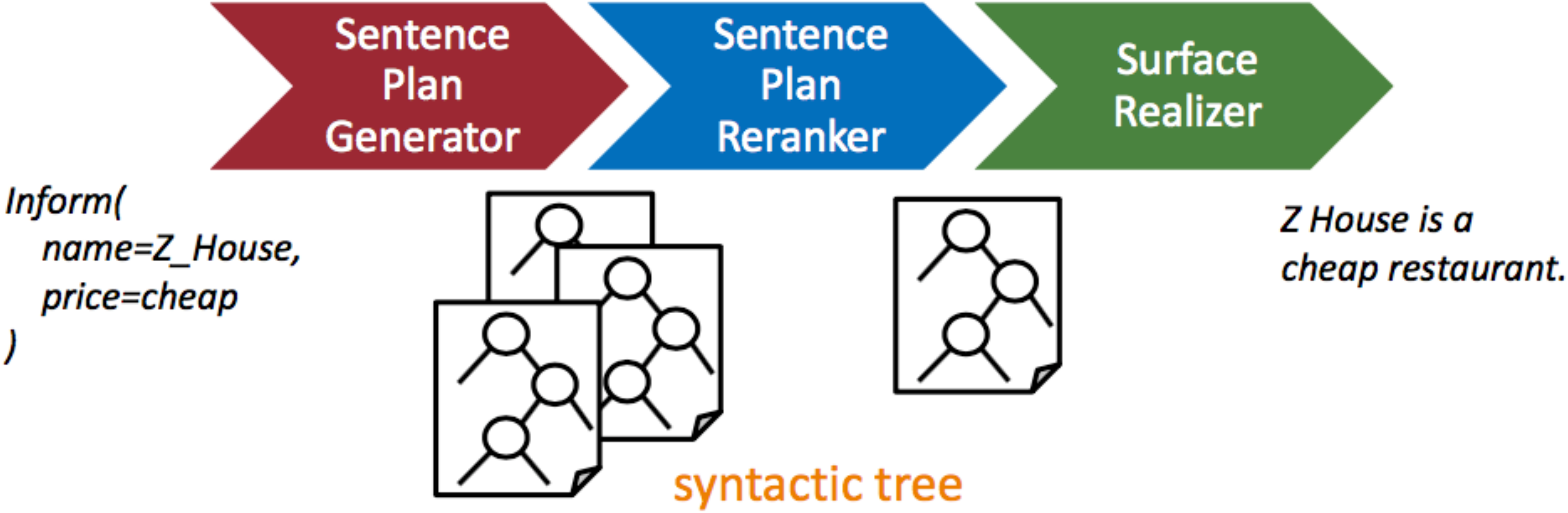
- To map dialogue acts into natural language

inform(name=Seven_Days, foodtype=Chinese)

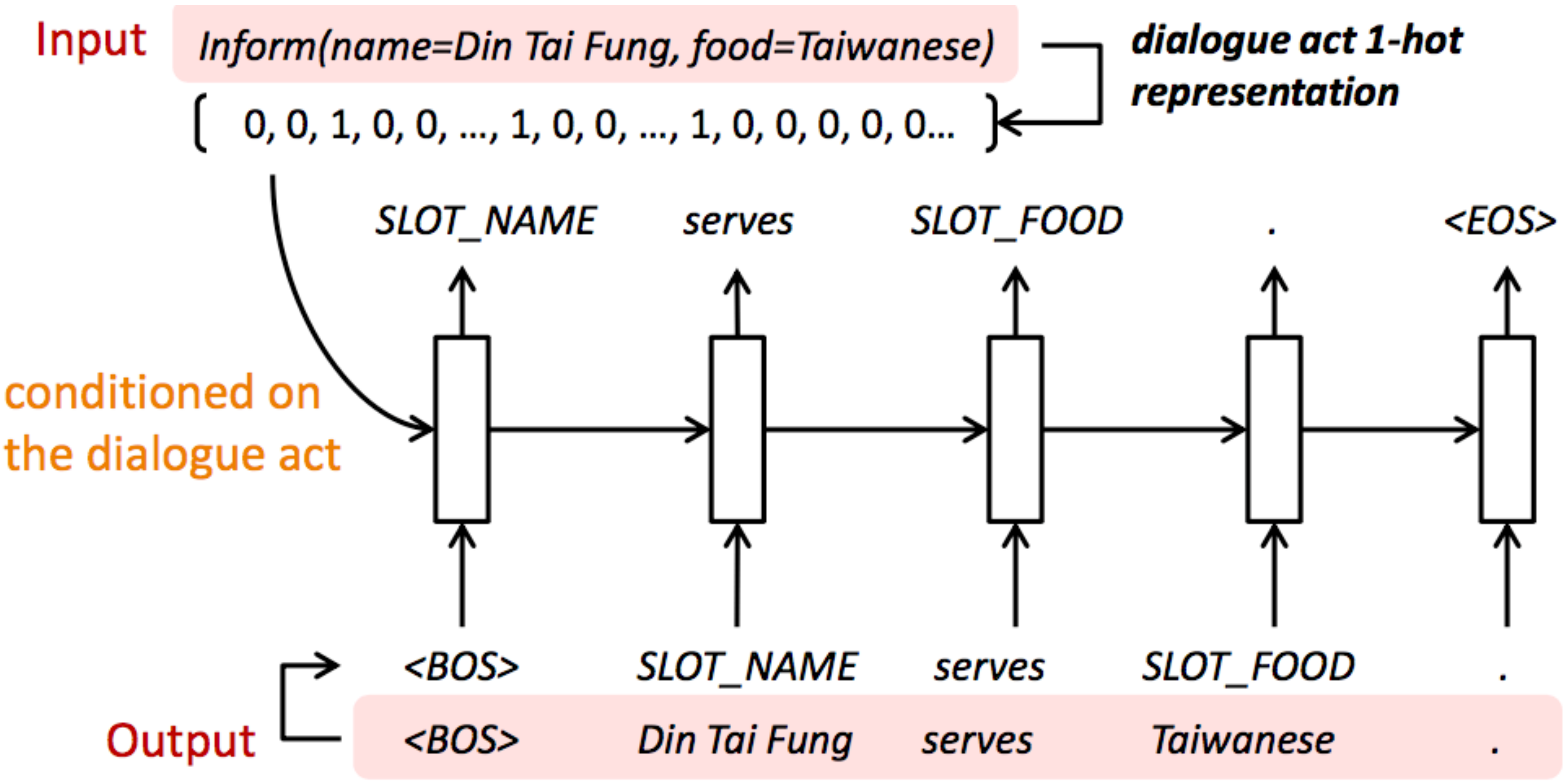
↓

Seven Days is a nice Chinese restaurant

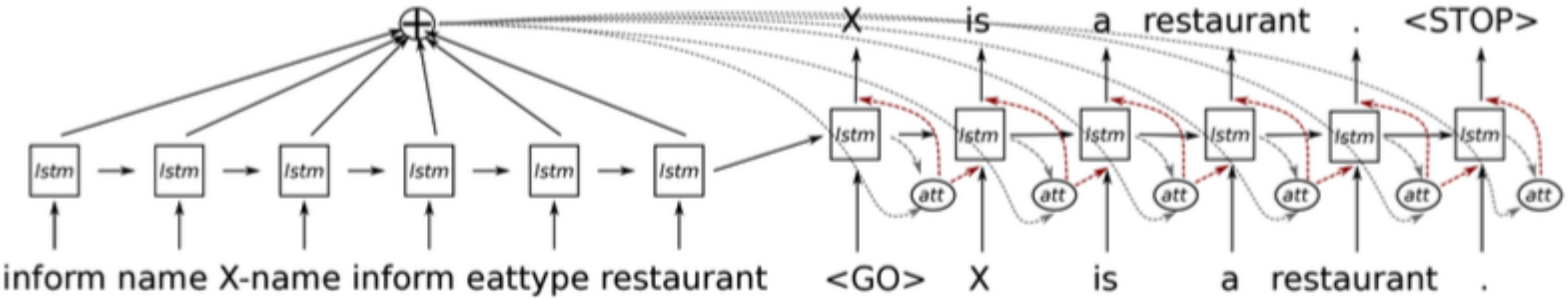
LG: Statistical NLG



LG: NN

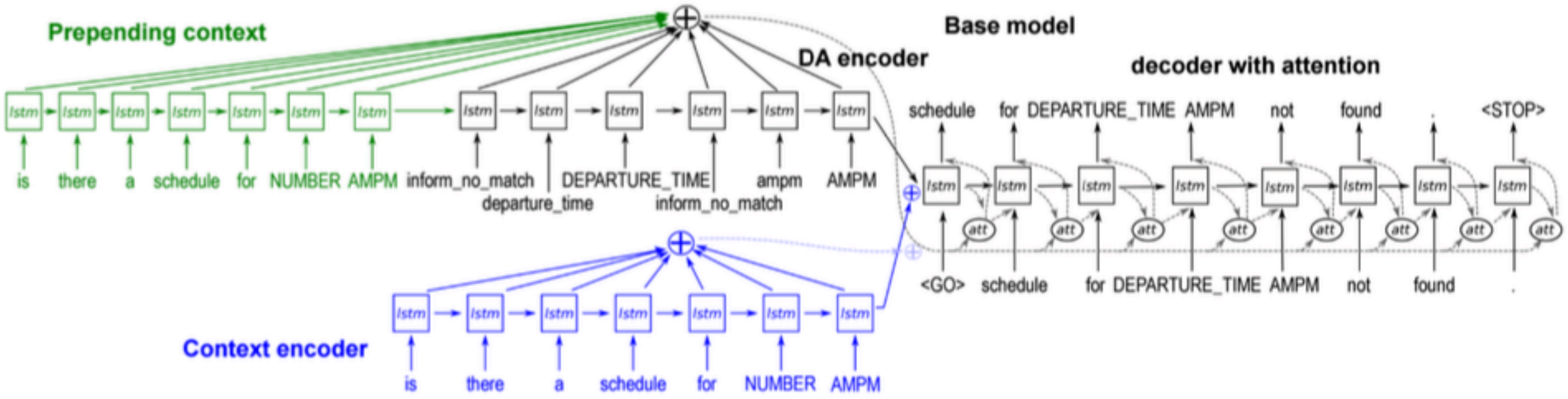

Input
Inform(name=Din Tai Fung, food=Taiwanese)
dialogue act 1-hot representation
0, 0, 1, 0, 0, …, 1, 0, 0, …, 1, 0, 0, 0, 0, 0…
SLOT_NAME
serves
SLOT_FOOD
.
<EOS>
conditioned on the dialogue act
<BOS>
SLOT_NAME
serves
SLOT_FOOD
.
Output
<BOS>
Din Tai Fung
serves
Taiwanese
.

LG: NN



X is a restaurant . <STOP>
lstm
att
inform name X-name inform eattype restaurant
<GO> X is a restaurant .

LG: Contextual

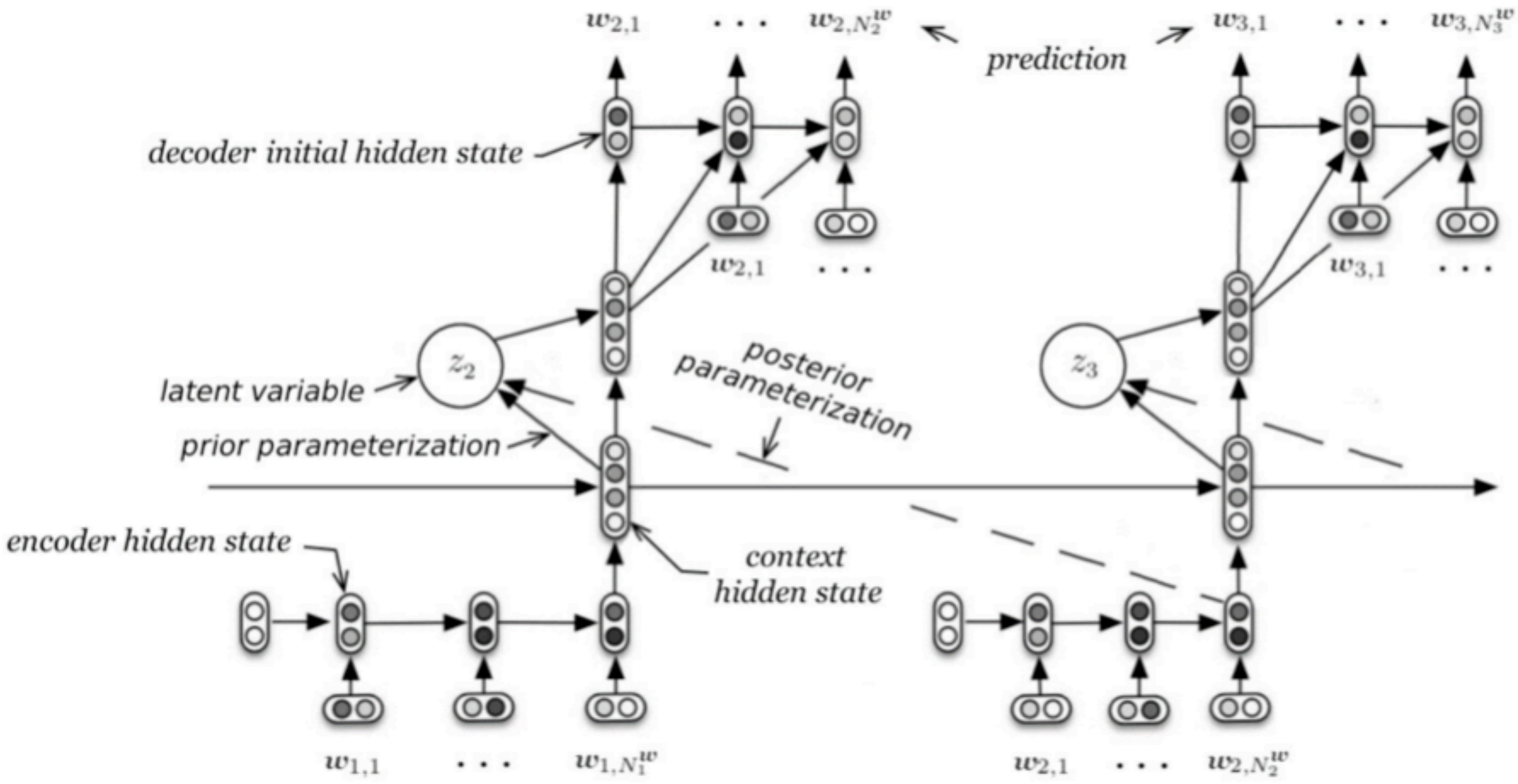


End-to-End

- ChitChat
- Task-oriented

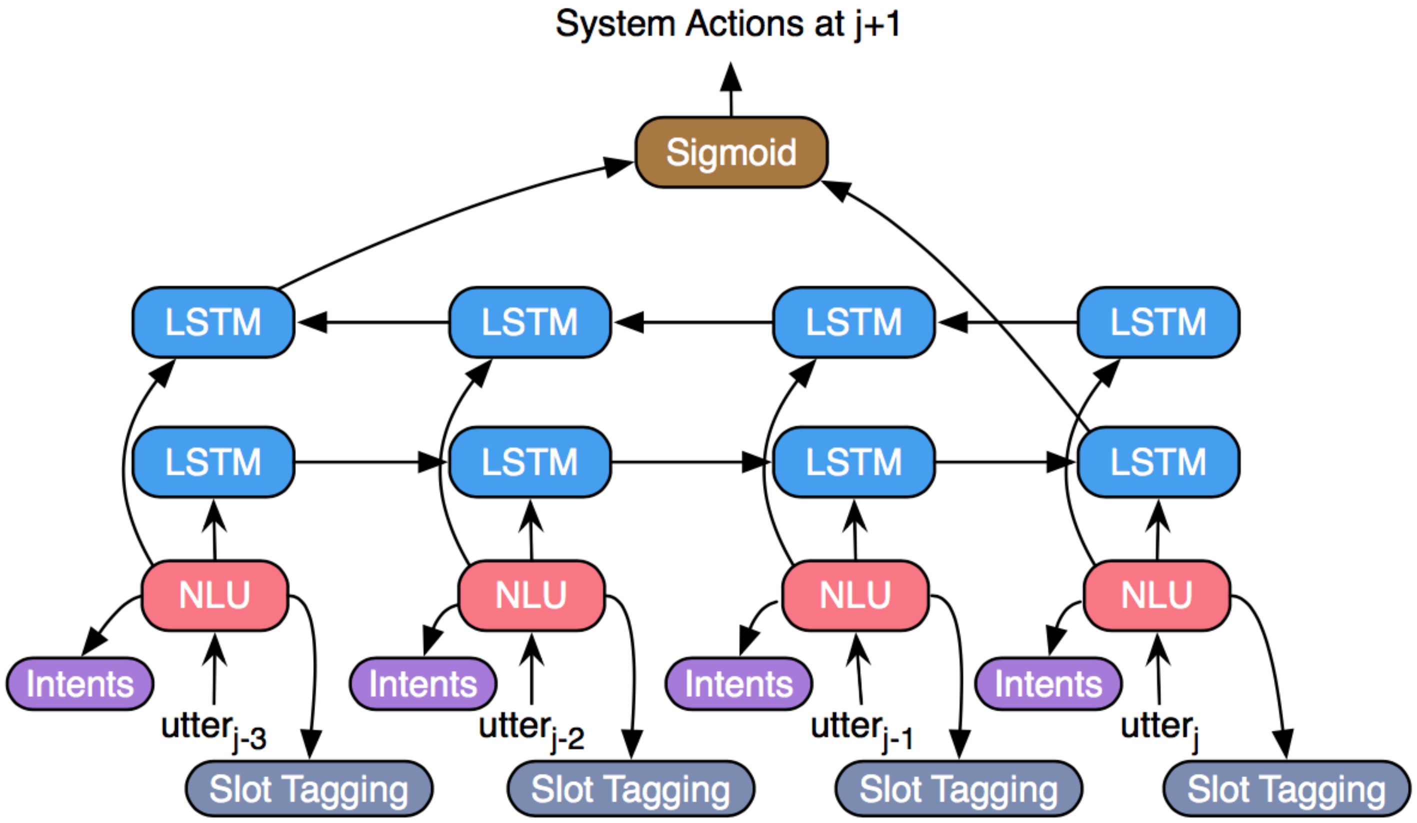
ChitChat

- Motivation:
  - To model dialogue without directly measurable goals
  - To train task-less DS using task-oriented data to obtain task-oriented DS
- Challenges:
  - To model topic in the DS
  - Dull response
- Approaches:
  - Seq2seq with MMI, deep RL, personalized DS (using user's personal history),
  - IR-based technique (using twitter data)



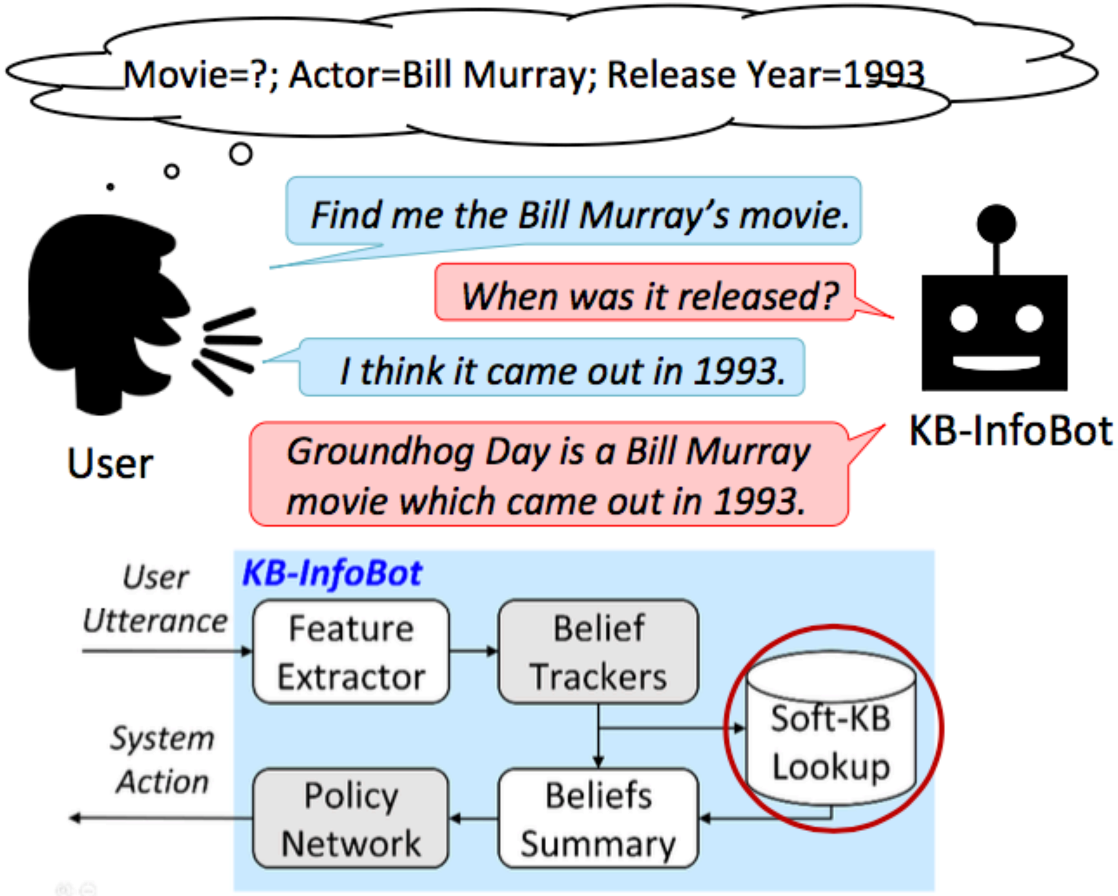
$w_{2,1}$ . . . w_{2,N_2^w}
prediction
$w_{3,1}$ . . . w_{3,N_3^w}
decoder initial hidden state
$w_{2,1}$ . . .
$w_{3,1}$ . . .
z_2
z_3
posterior parameterization
latent variable
prior parameterization
encoder hidden state
context hidden state
$w_{1,1}$ . . . w_{1,N_1^w}
$w_{2,1}$ . . . w_{2,N_2^w}

E2E: Task Oriented (Supervised Learning)

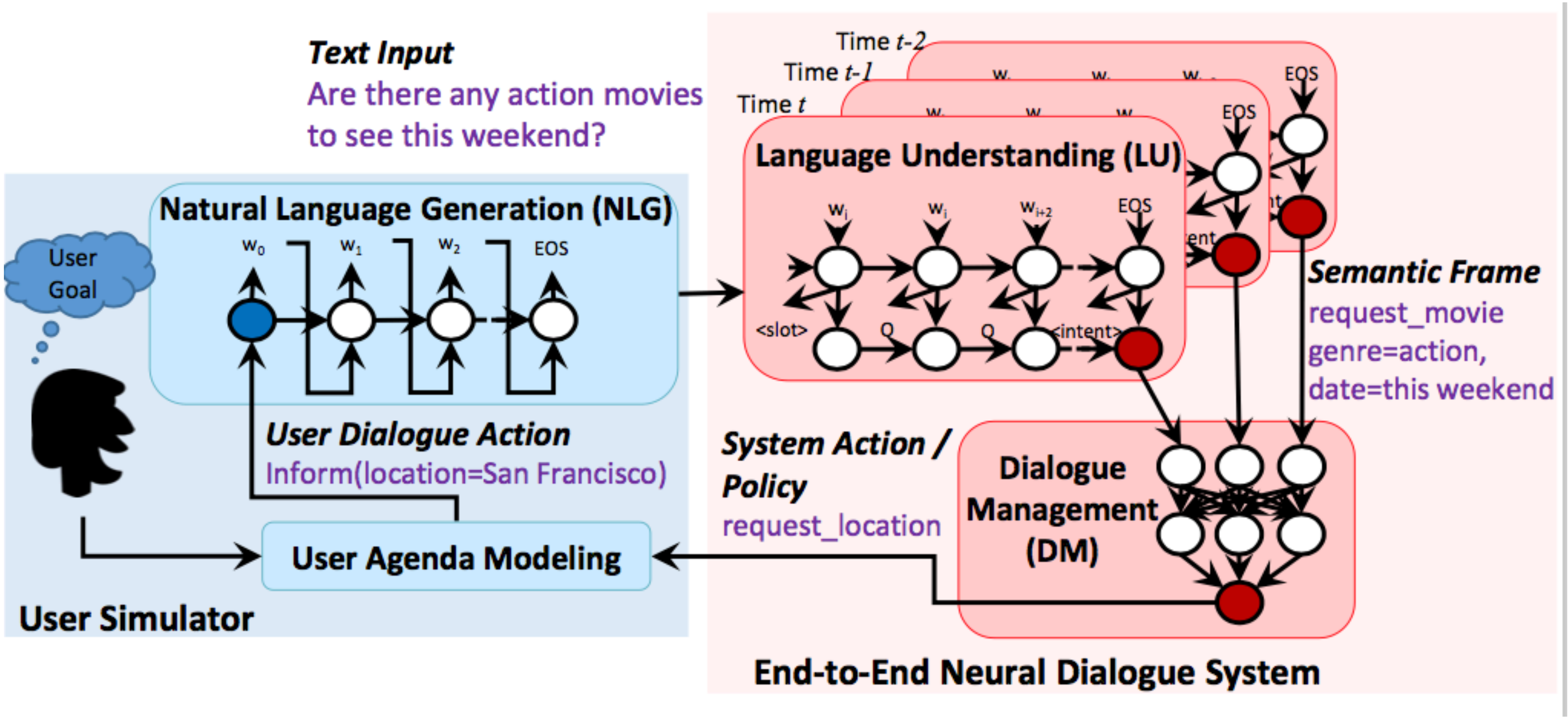


E2E: Task Oriented (Reinforcement Learning)

- To traverse knowledge base



E2E Task-Oriented (RL)



Breakdown

- Kyoto Institute of Technology
  - Poly kernel SVM
  - Features: word vector both in the system utterance and the previous user utterance (modified combinations of these utterances)
- Shizuoka University
  - Handcrafted rules
    - I: if there is no shared keywords between system’s and user’s utterance, then it is a breakdown
    - II: system’s utterance after user’s question is a breakdown
    - III: system’s utterance which is a question is a breakdown

- Tohoku University and PFI
    - Encode a pair of user’s and system’s utterance using NCM, LSTM encoder, BOW, or extended NCM

- NAIST
    - LSTM-RNN
    - Features: word frequency vector (user and system), frequency vector of co-occurrence words, doc2vec (user, system, co-occurrence)

- <u>NTT Communication</u>
  - 6-layer perceptron
  - Features: word vector (user and system), word class vector (user and system), perplexity, cosine similarity (system and previous system), personality question, dialogue acts (SVM)

- Hiroshima City University
  - LSTM-RNN
  - Features: word2vec of user’s and system’s utterances