HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Hotels Time Dialogue Belief Location Restaurant Domain German

Breaking open Belief Tracking

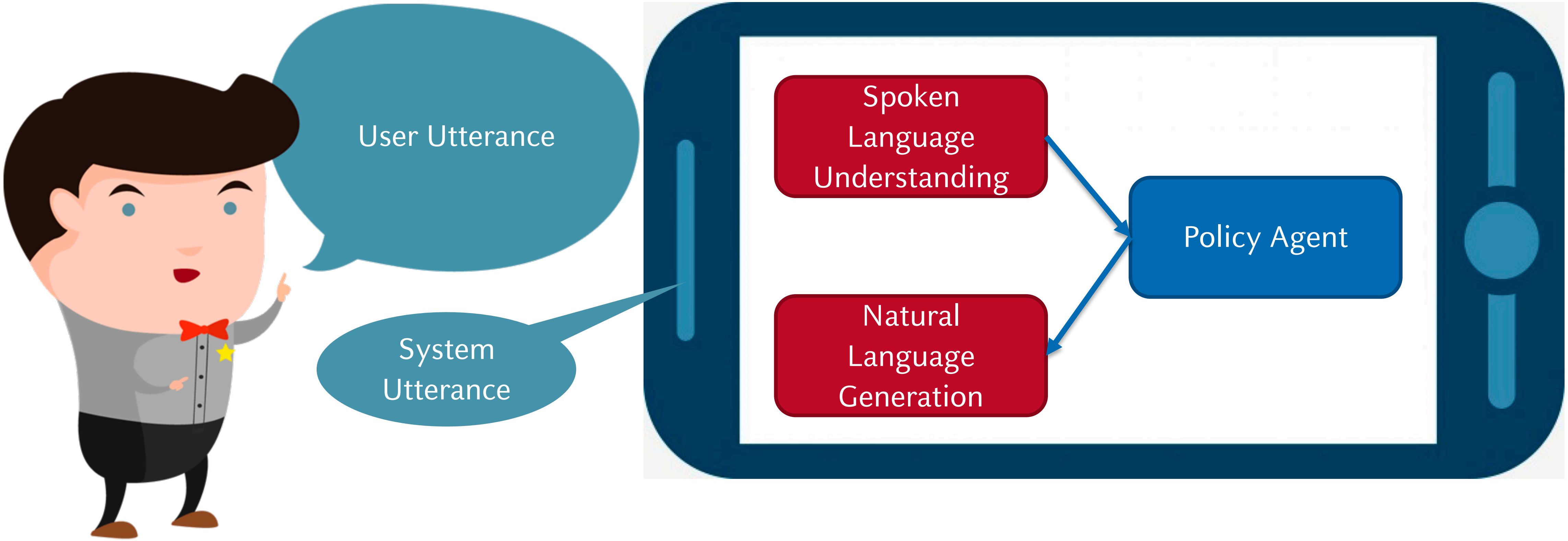
Carel van Niekerk

23 August 2019

Düsseldorf

Germany

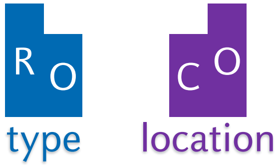
Dialogue System



Dialogue Systems without Belief Tracker

Hey. I need a restaurant near the city centre.

hello(type=restaurant)	0.6
inform(type=restaurant, location=centre)	0.4



Where would you like the restaurant?

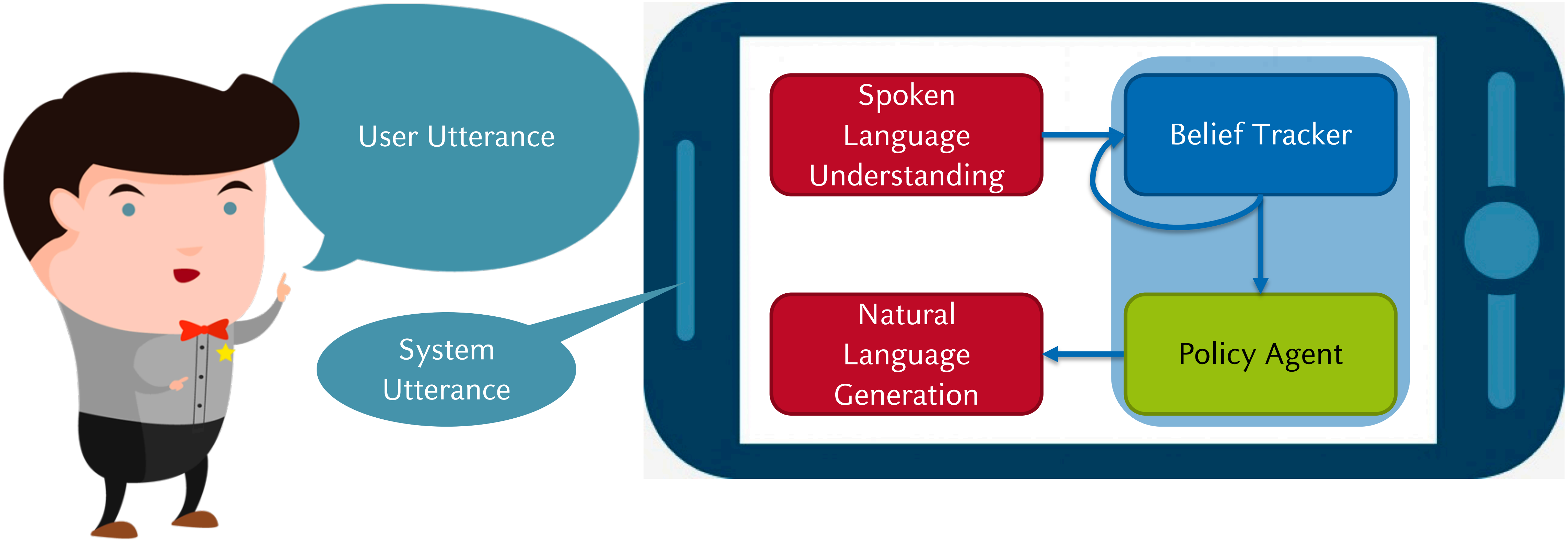
The City Centre!

inform(location=city)	0.6
inform(location=centre)	0.4

Do you want a restaurant?

Dialogue System

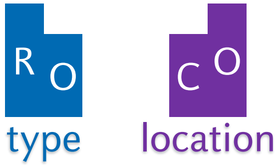
HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF



Dialogue System with Belief tracking

Hey. I need a restaurant near the city centre.

hello(type=restaurant)	0.6
inform(type=restaurant, location=centre)	0.4



Where would you like the restaurant?

The City Centre!

inform(location=city)	0.6
inform(location=centre)	0.4

To confirm you want a restaurant near the city centre?

Dialogue Belief Tracking

- Belief State – Internal **Distribution** over **states**
- State - **Information** the **agents** needs to make **decisions**
  - Capture **user intentions**
  - Capture **history** of dialogue
- Aim: **Predict** Belief State

Datasets

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

- Two main comparative sets:
  - WOZ 2.0
  - **MultiWOZ 2.1** (Most Challenging)

- Metrics
  - **Slot** accuracy – Proportion of **domain-slot-value** triplets **correctly** identified.
  - **Joint-goal** accuracy – Proportion of **turn** where **all** user goals **correctly** identified.

WOZ 2.0

- Single Domain – Restaurants
- 1200 Dialogues

Model	Slot Accuracy	Joint-goal Accuracy
NBT	-	84.8%
MDBT	96.4%	85.5%
GLAD	97.1%	88.1%
StateNET	-	88.9%
GCE	97.4%	88.5%
GLAD + RC + FS	97.4%	89.2%

MultiWOZ 2.0

- **Multiple** Domain – 7 domains
- **10000+** Dialogues
- **Richer & Noisier** Dialogues

Model	Slot Accuracy	Joint-goal Accuracy
MDBT	89.53%	15.57%
GCE	98.42%	36.57%
Neural Reading		41.10%
HYST		44.24%
SUMBT	96.44%	46.65%
TRADE	96.42%	48.62%

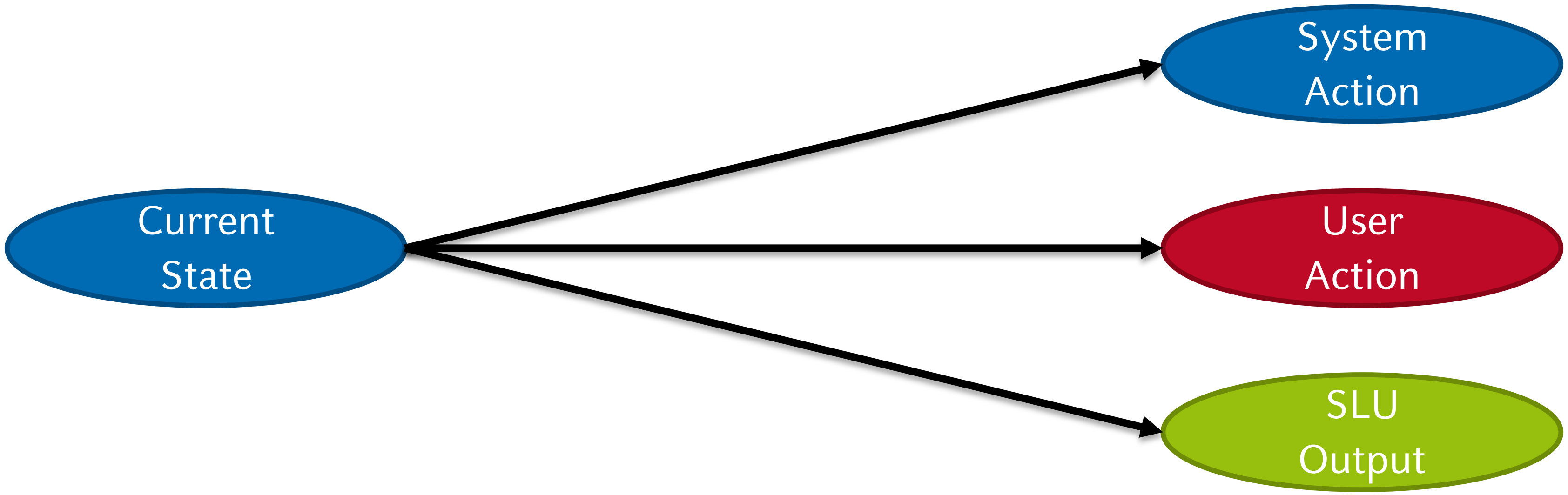
MultiWOZ 2.1

- Corrections:
  - **Delayed** annotation
  - **Incorrect** annotation
  - **Missed** annotations
  - **Spelling** errors in annotations

Model	2.0	2.1
Neural Reading	41.10%	36.40%
HYST	44.24%	38.10%
TRADE	48.62%	45.60%

Generative vs Discriminative

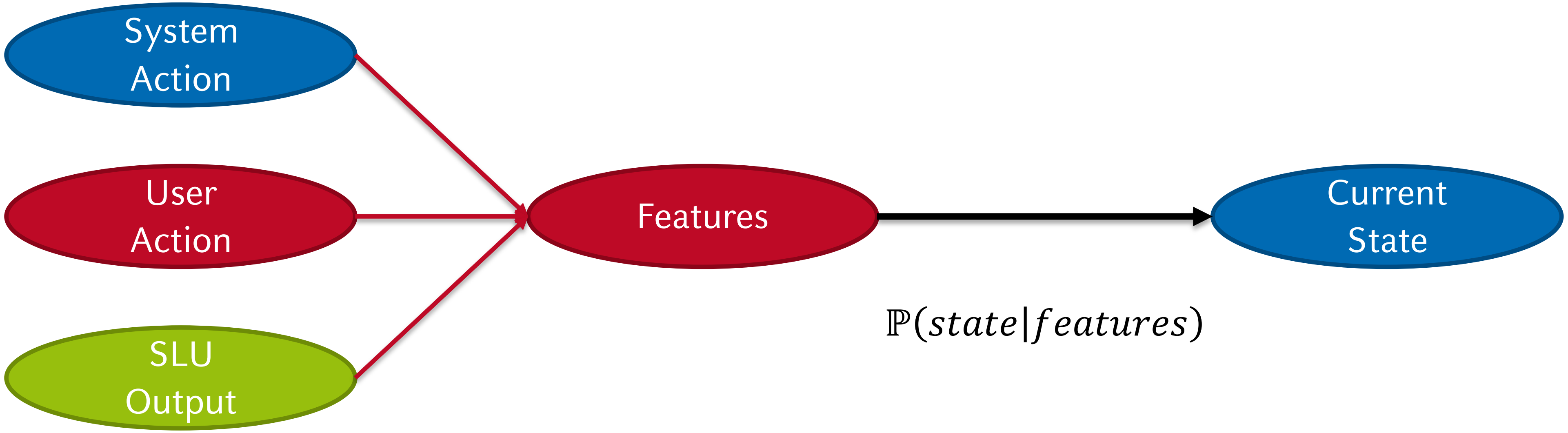
Generative



$$\mathbb{P}(sys, usr, slu | state)$$

The **future observations** are **generated** by the **current state**.

Discriminative



Discriminate between the possible **states** using **features** of the dialogue.

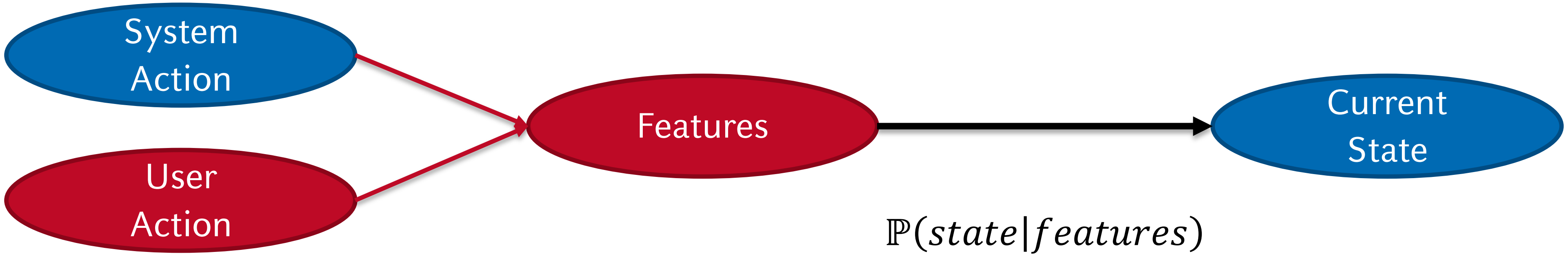
Generative vs Discriminative

- **Generative models:**
  - Assumes turns are **independent** given the state.

- **Discriminative models:**
  - **No assumptions** about the independence.
  - **Outperforms** Generative models.

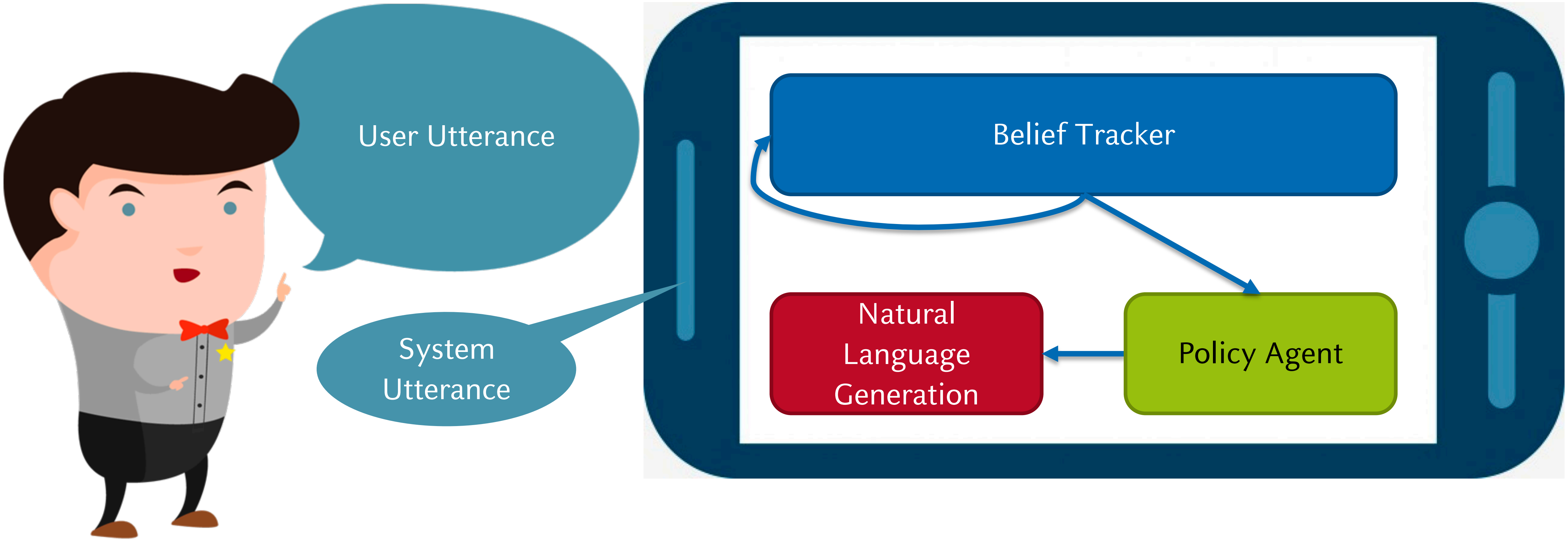
Models without a independent SLU

- Independent SLU Problems:
  - Accumulation of **errors**
  - Requires **additional** annotated training **data**.



Combining the **SLU** and **Feature extractors** into one.

Dialogue System



Single Unit SLU and DST

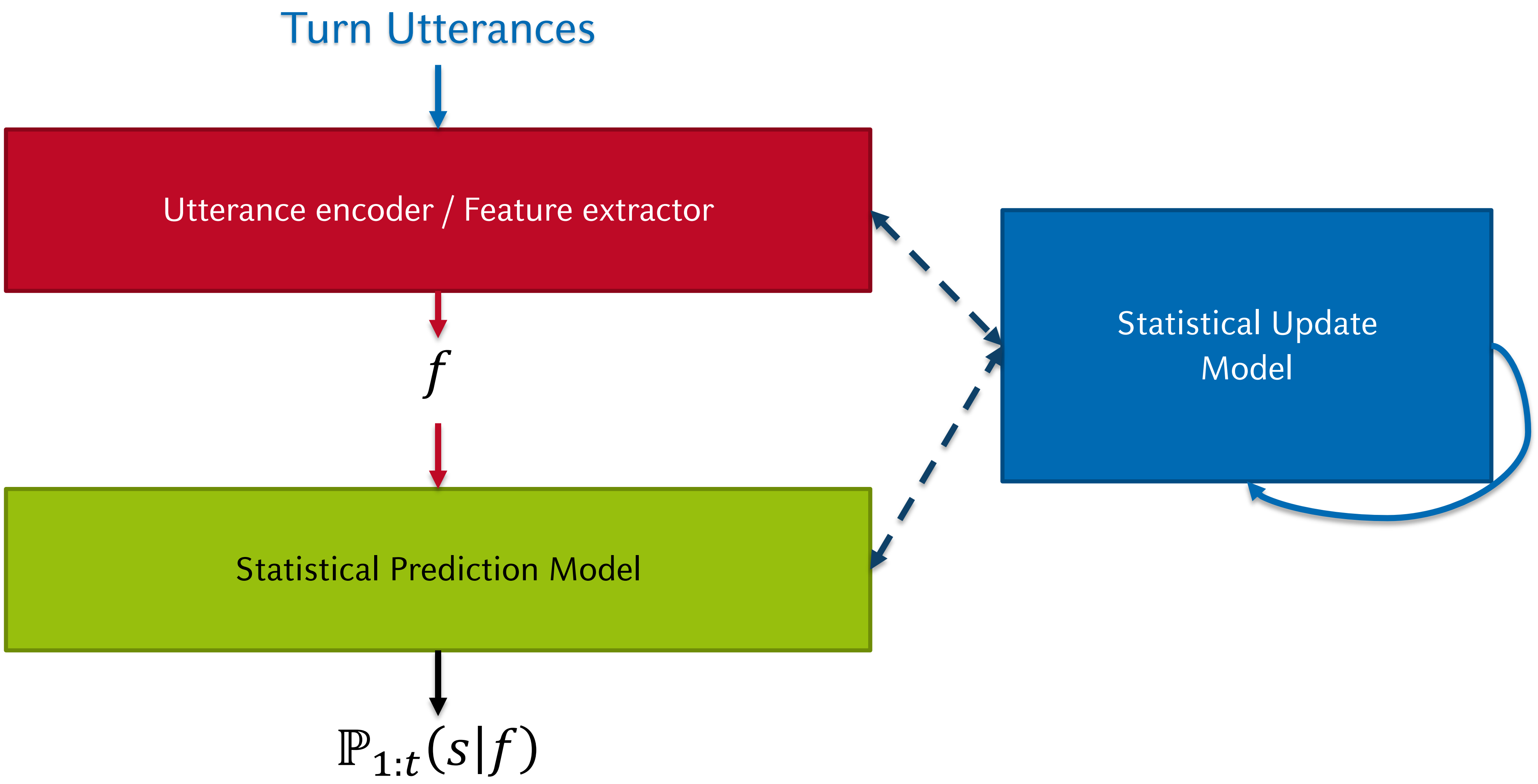
- **Delexicalisation** for single SLU and Belief Tracker model.
  - Requires **large dictionaries** of semantic lexicons.

- **Word Embeddings** and Feature Extractors
  - **Convolutional** Neural Networks
  - **Recurrent** cell NN
  - More **scalable**
  - **Equivalent** or better **performance**

Fully Statistical Trackers

- Statistical Dialogue Trackers:
  - **Recurrent** Cell
  - More **adaptable**
  - Superior **performance**

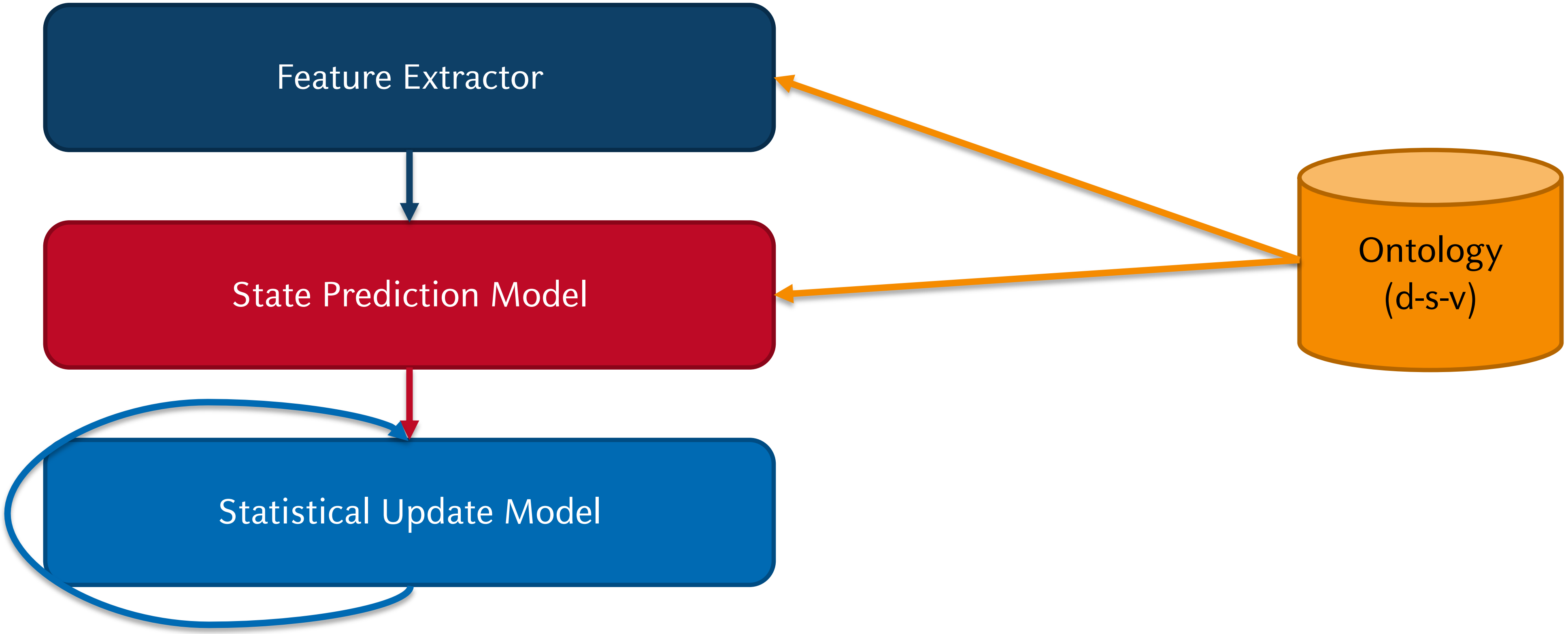
Statistical Discriminative Belief Tracking



Multi-Domain Belief Tracker

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Overview:



HEINRICH HEINE UNIVERSITÄT DÜSSELDORF

Utterance Encoder

- **Semantic similarity** to identifies the presence of a state in a utterance.
- **Slot-Value Features:**
  - **User confirm** (System **slot-value** + User **affirm**)

    *System: So you want a **restaurant** near the **centre** of town?*

    Restaurant-location-centre

    *User: **Yes***

Utterance Encoder

- **User request** (System **slot** + User **value**)

*System: Where would you like the **hotel** to be?*

Hotel-location-??

*User: Near the **Rhine** river.*

Rhine

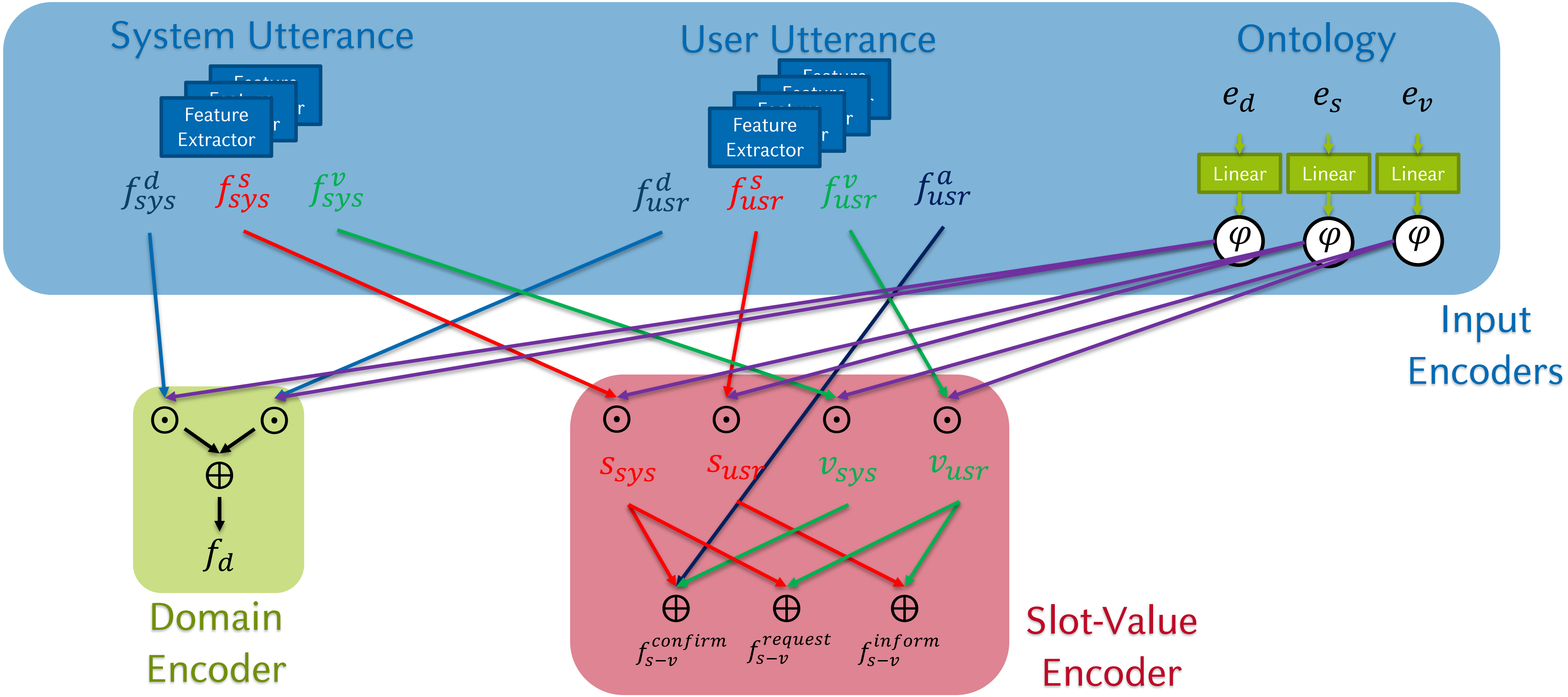
- **User inform** (User **slot-value**)

*User: I need a **taxi** to the **airport** at 10.*

Taxi-destination-airport

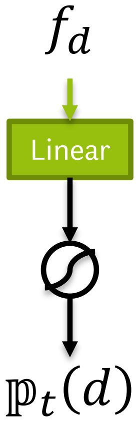
HEINRICH HEINE UNIVERSITÄT DÜSSELDORF

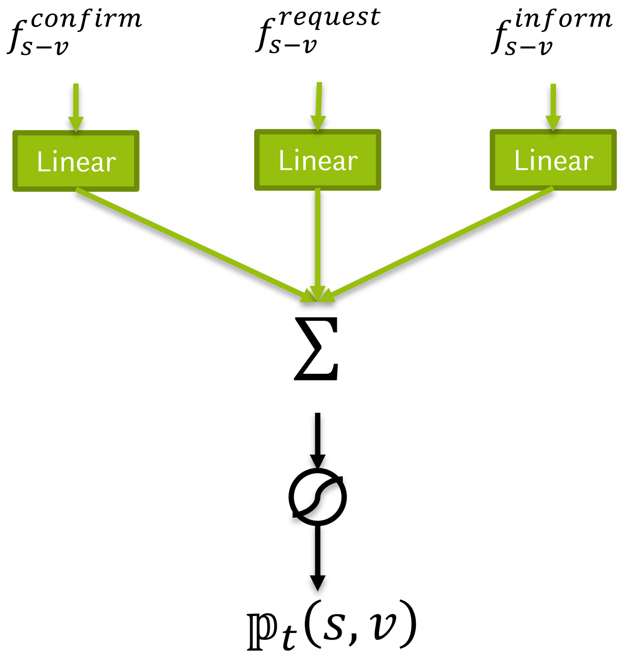
Utterance Encoder



HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Statistical Prediction Model



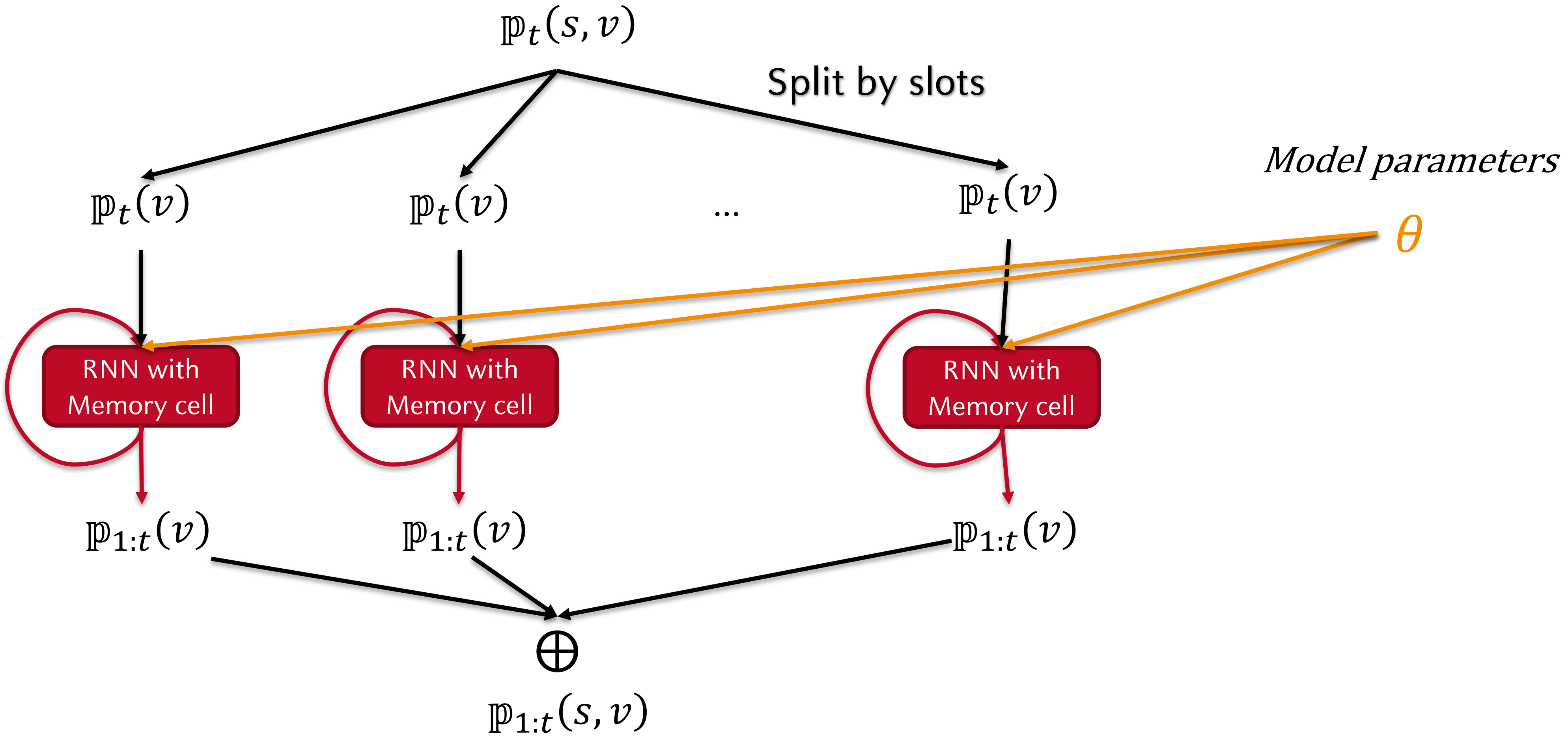


Statistical Prediction Model

- Two **independent** models. **Share knowledge** across domains.
- **Shares parameters** across all ontology terms -> **Scalable**
- **Multi-class** classification - individual **binary** classification
  - Allows **adaption** to new domains

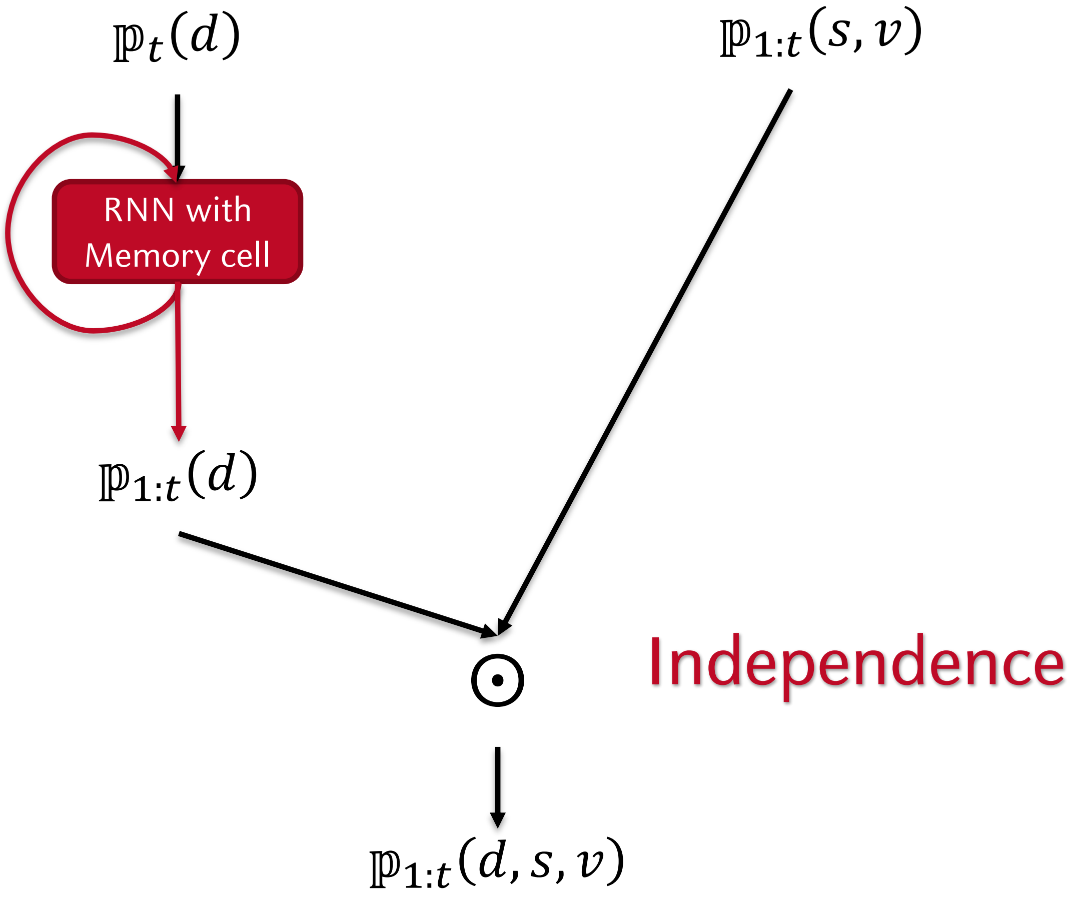
HEINRICH HEINE UNIVERSITÄT DÜSSELDORF

Statistical Update Model



HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Statistical Update Model



Overview

Dataset	Slot Accuracy	Joint-goal Accuracy
WOZ 2.0	96.4%	85.5%
MultiWOZ 2.0	89.53%	**15.57%**

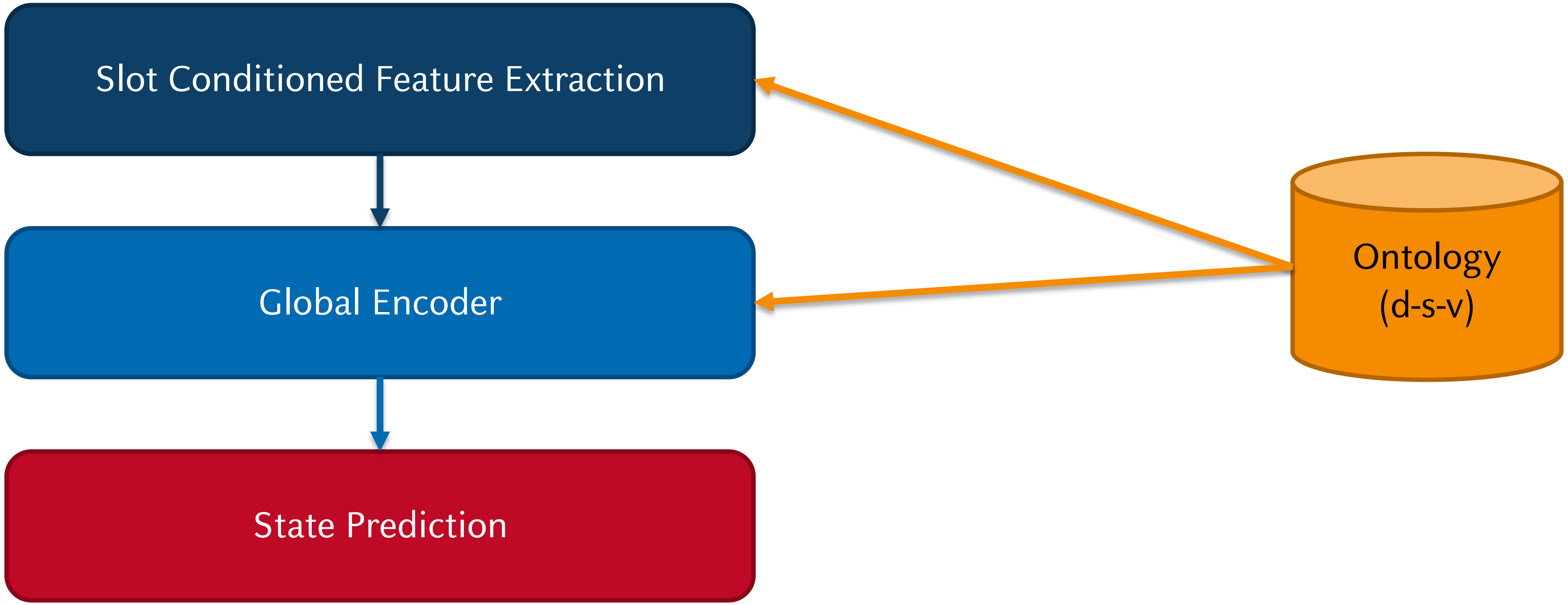
- Shortcomings **Adapting**
- Assumes Known Ontology – Scalability Issues

Improved Feature Extraction

- Slot **conditioned**
- **Global** parameter sharing
- Self-Attention **contextual** embeddings

Globally-Conditioned Encoder (GCE)

Overview:



Globally-Conditioned Encoder (GCE)

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Utterance Encoder

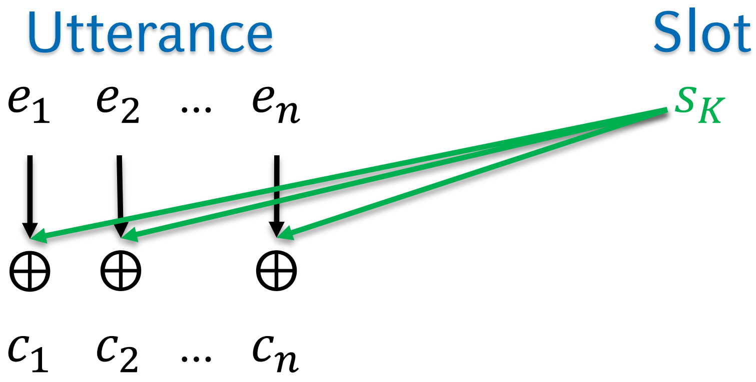
- Bidirectional **LSTM** model -> Contextual **token** embeddings

- **Convolutional** self-attention -> Contextual **utterance** embedding.

- Embeddings:
  - Current User Utterance
  - Previous j System acts
  - Value candidates

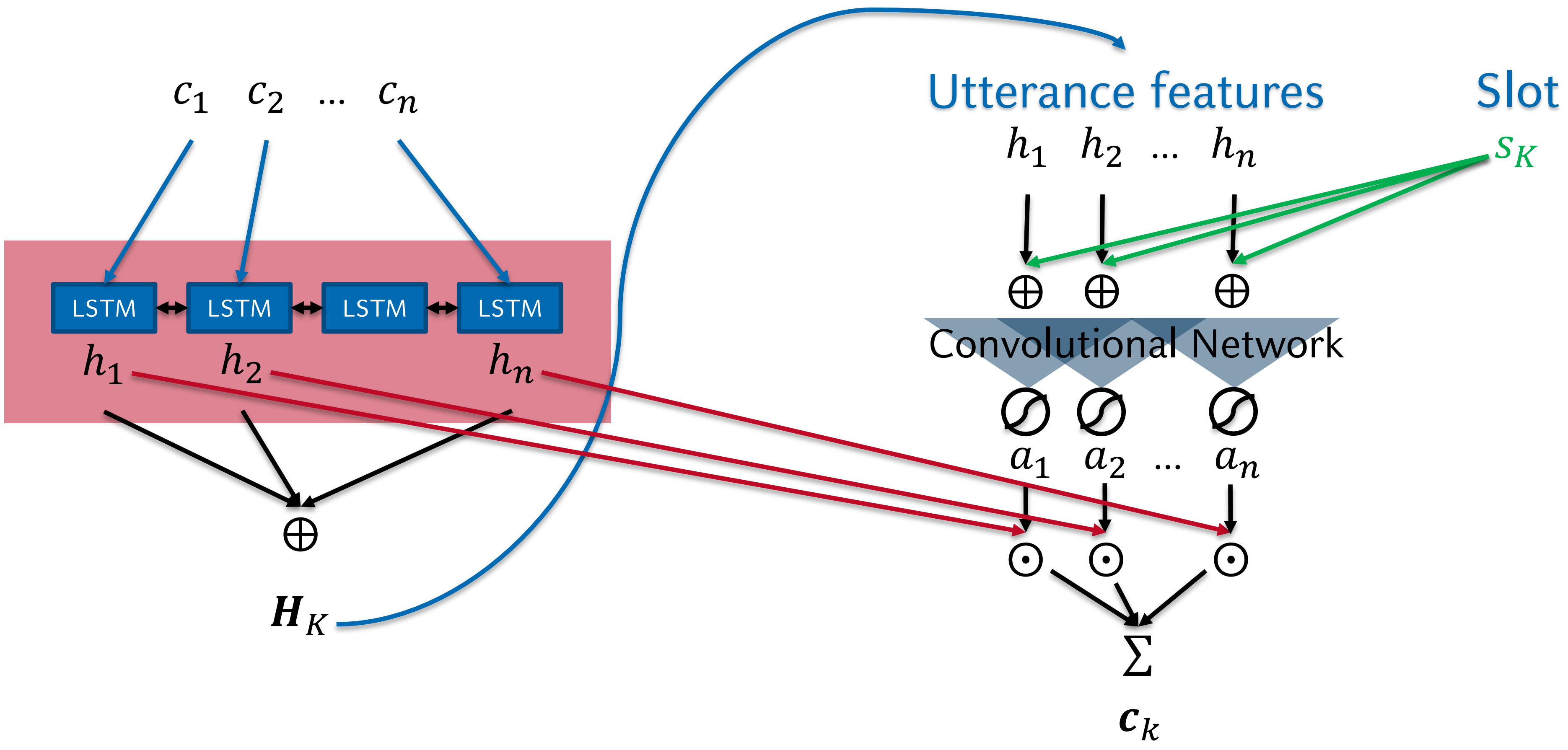
Globally-Conditioned Encoder (GCE)

Utterance Encoder – Slot Conditioned token embeddings



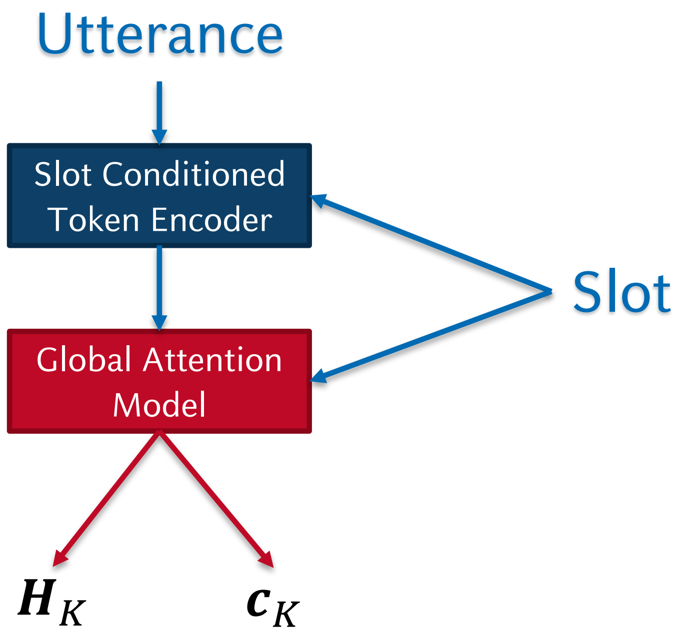
HEINRICH HEINE UNIVERSITÄT DÜSSELDORF

Utterance Encoder – Bi-directional LSTM with Self-Attention

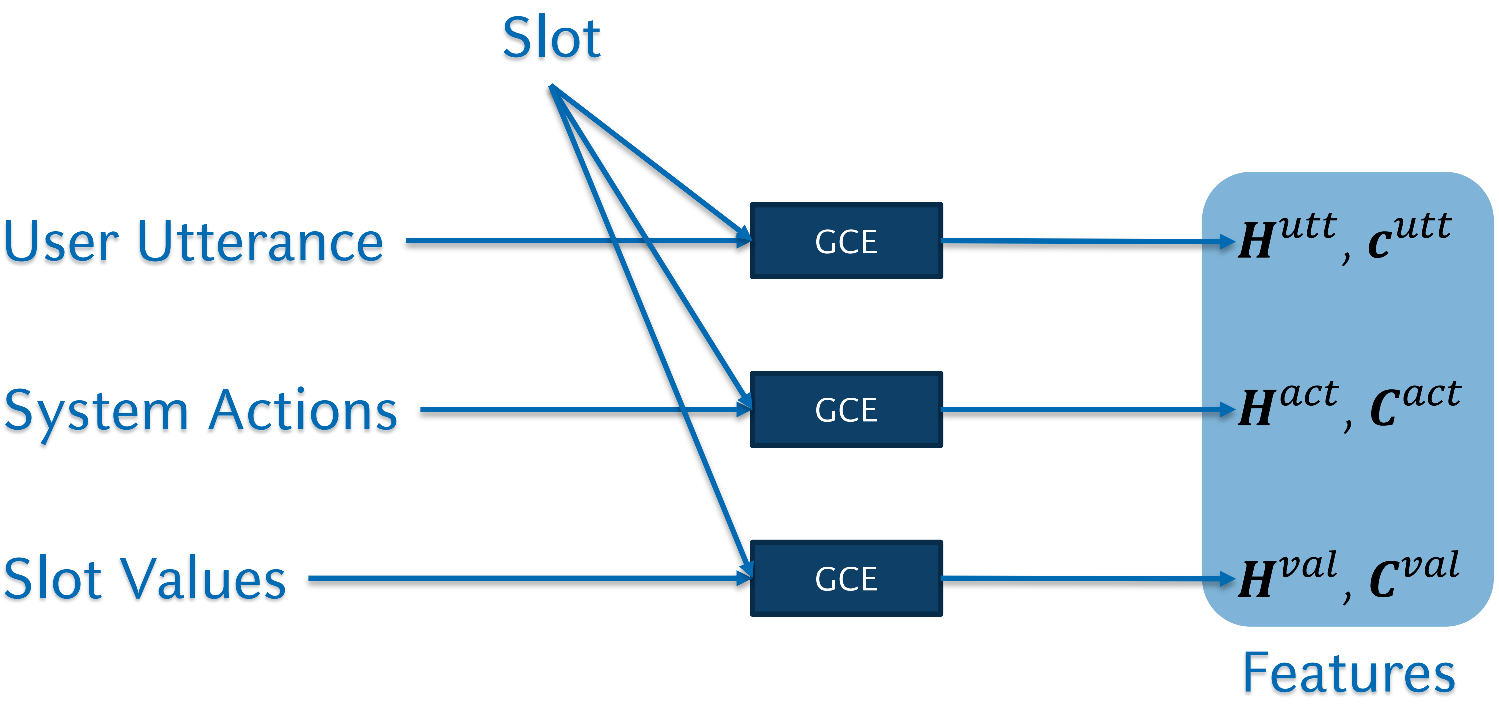


HEINRICH HEINE UNIVERSITÄT DÜSSELDORF

The Globally-Conditioned Encoder (GCE)



The Globally-Conditioned Encoder (GCE)



Statistical Prediction Model

- **Utterance scoring** model
  - **User token** embeddings + **value** embeddings
  - Degree to which the slot-value pair was **mentioned** by the **user**.

User: I want a ***Italian*** *restaurant.*

Statistical Prediction Model

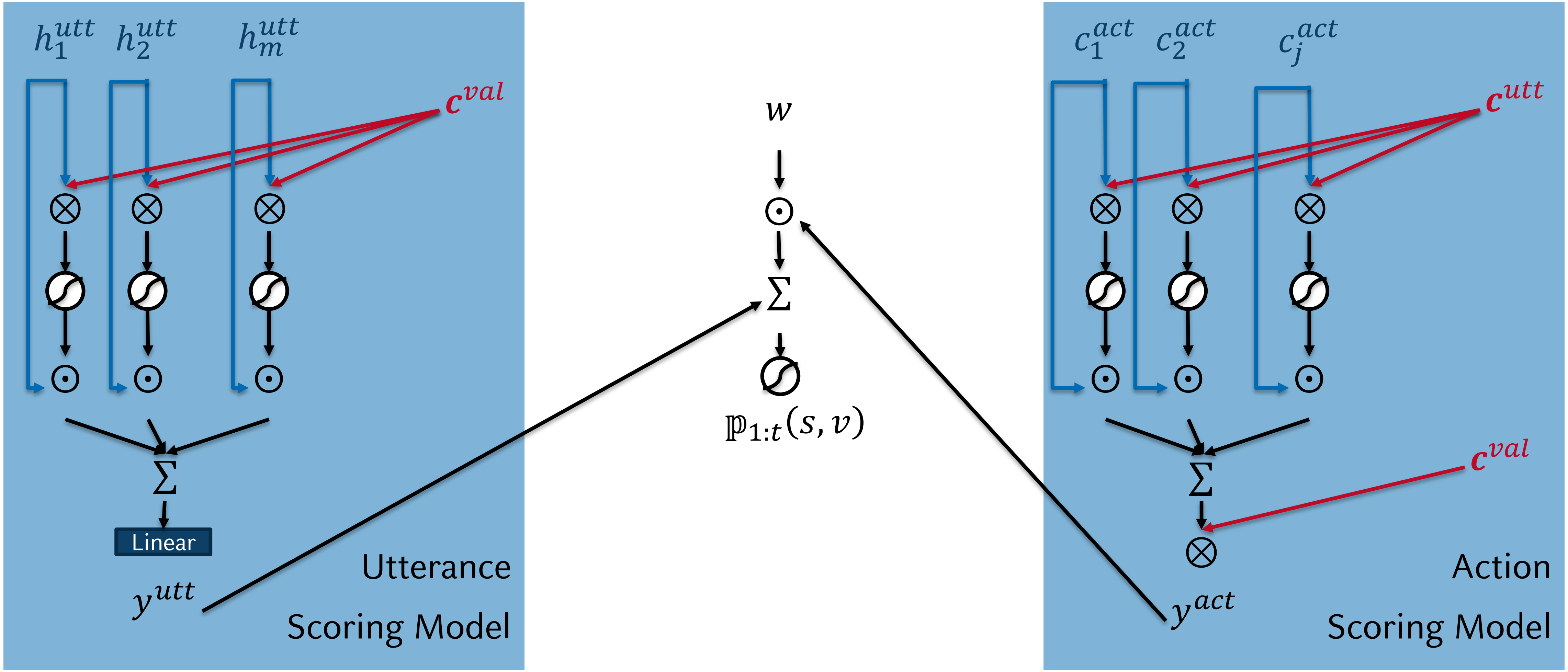
- **Action scoring** model
  - **System** utterance embeddings + **User** utterance + **value**
  - Degree to which the slot-value was **mentioned** by the **system**.

*System: Would you like the restaurant to be in the **east** of town?*

Location not east

*User: **No**.*

Statistical Prediction Model



Globally-Conditioned Encoder (GCE)

HEINRICH HEINE UNIVERSITÄT DÜSSELDORF

Overview

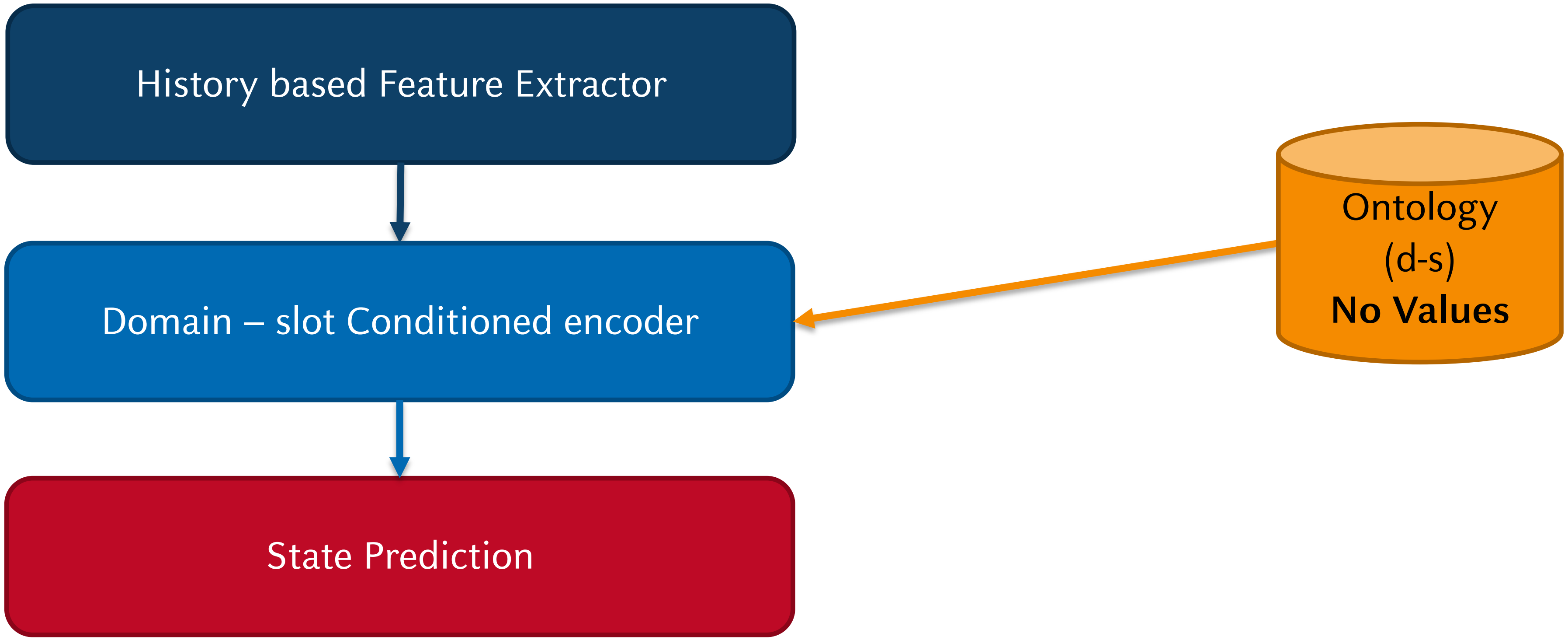
Dataset	Slot Accuracy	Joint-goal Accuracy
WOZ 2.0	97.38%	88.51%
MultiWOZ 2.0	98.42%	**36.57%**

- Limitations:
  - **Past J** system utterances used.
  - Assumes Known Ontology – Scalability Issues

Dynamic Ontology

- Value **generation** models

Overview:


History based Feature Extractor
Domain – slot Conditioned encoder
State Prediction
Ontology
(d-s)
No Values

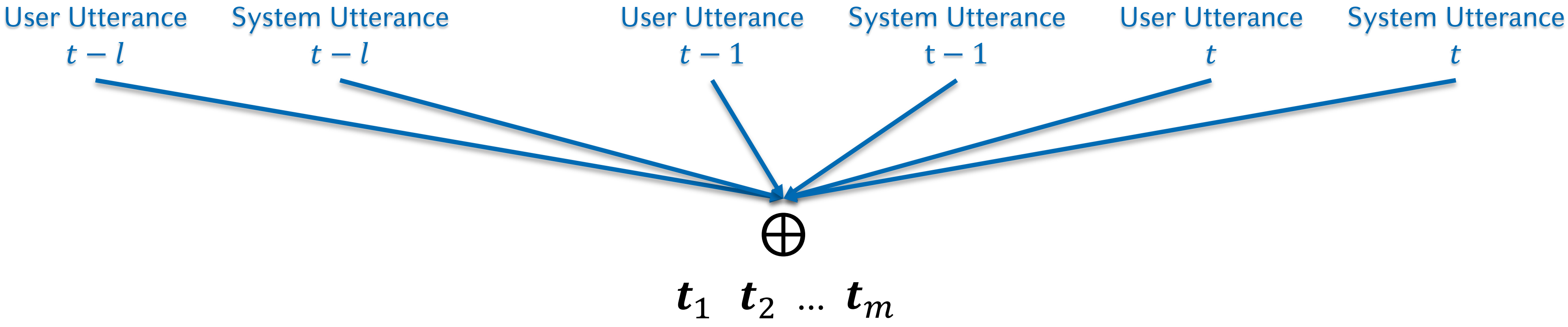
Transferable Dialogue State Generator (TRADE)

Utterance Encoder

- Bidirectional **GRU** model -> Contextual **token** embeddings
- Domain-slot **conditioned** GRU -> Contextual **history** embedding.
- Encodes past l turns **jointly**.

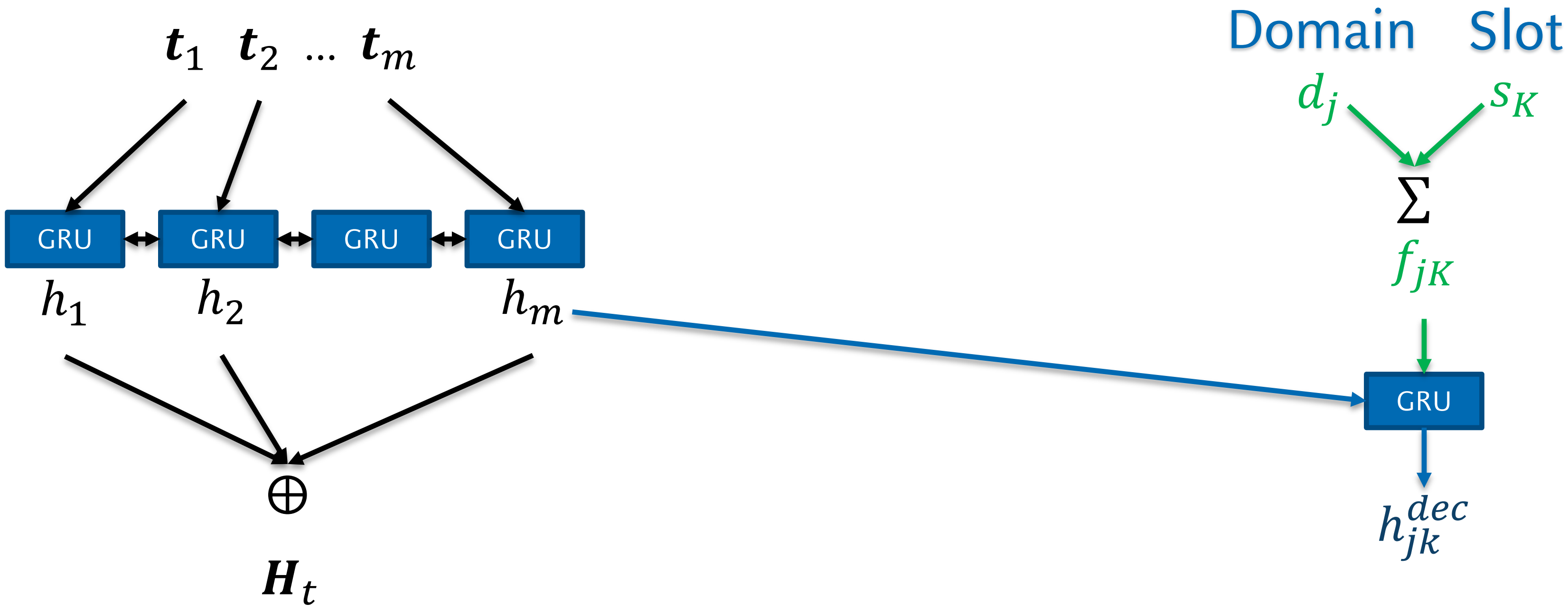
HEINRICH HEINE UNIVERSITÄT DÜSSELDORF

Utterance Encoder



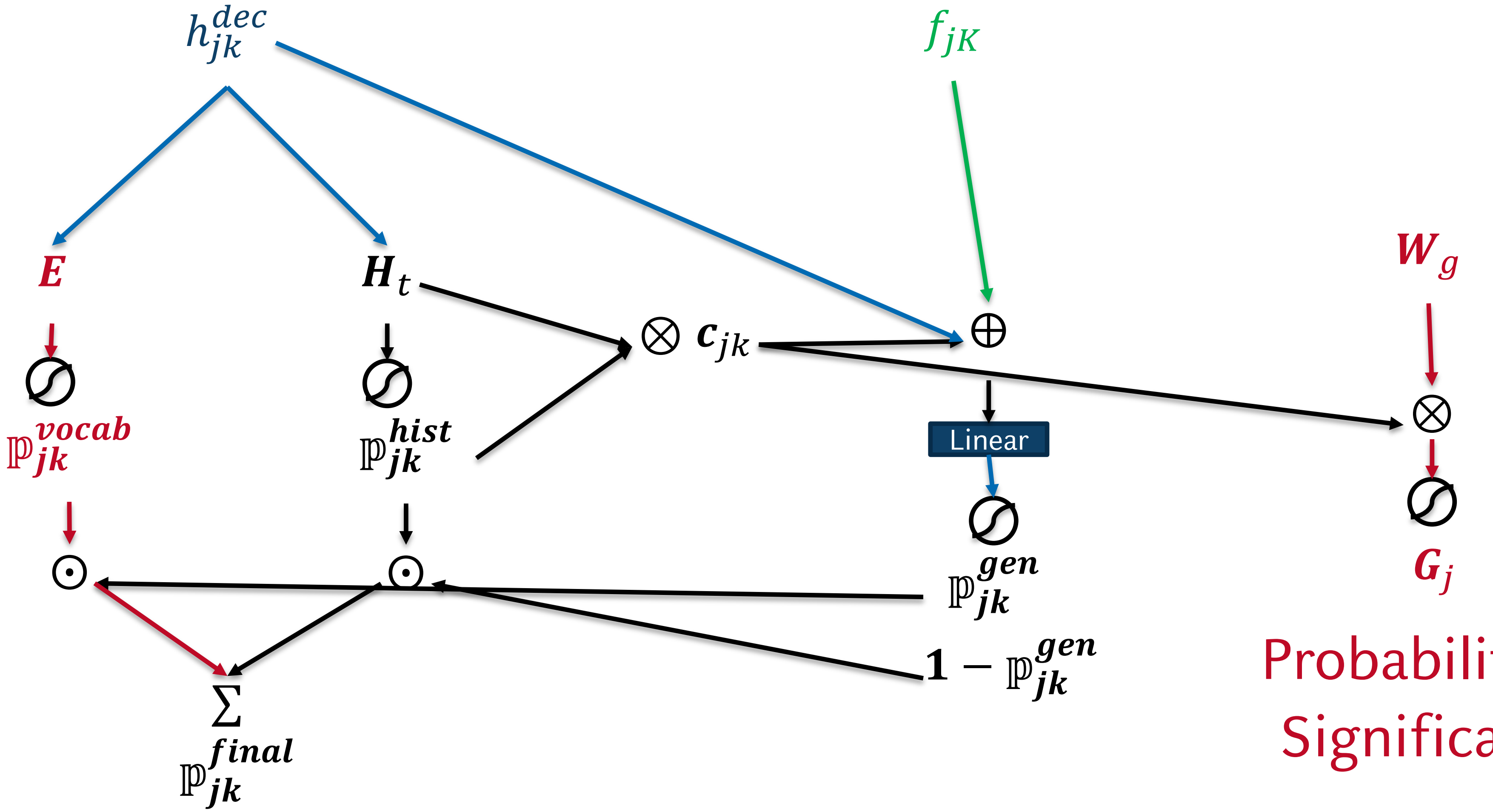
Transferable Dialogue State Generator (TRADE)

Utterance Encoder



HEINRICH HEINE UNIVERSITÄT DÜSSELDORF

Statistical Prediction Model – The TRADEMARK!



Probability of Significance

Transferable Dialogue State Generator (TRADE)

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Overview

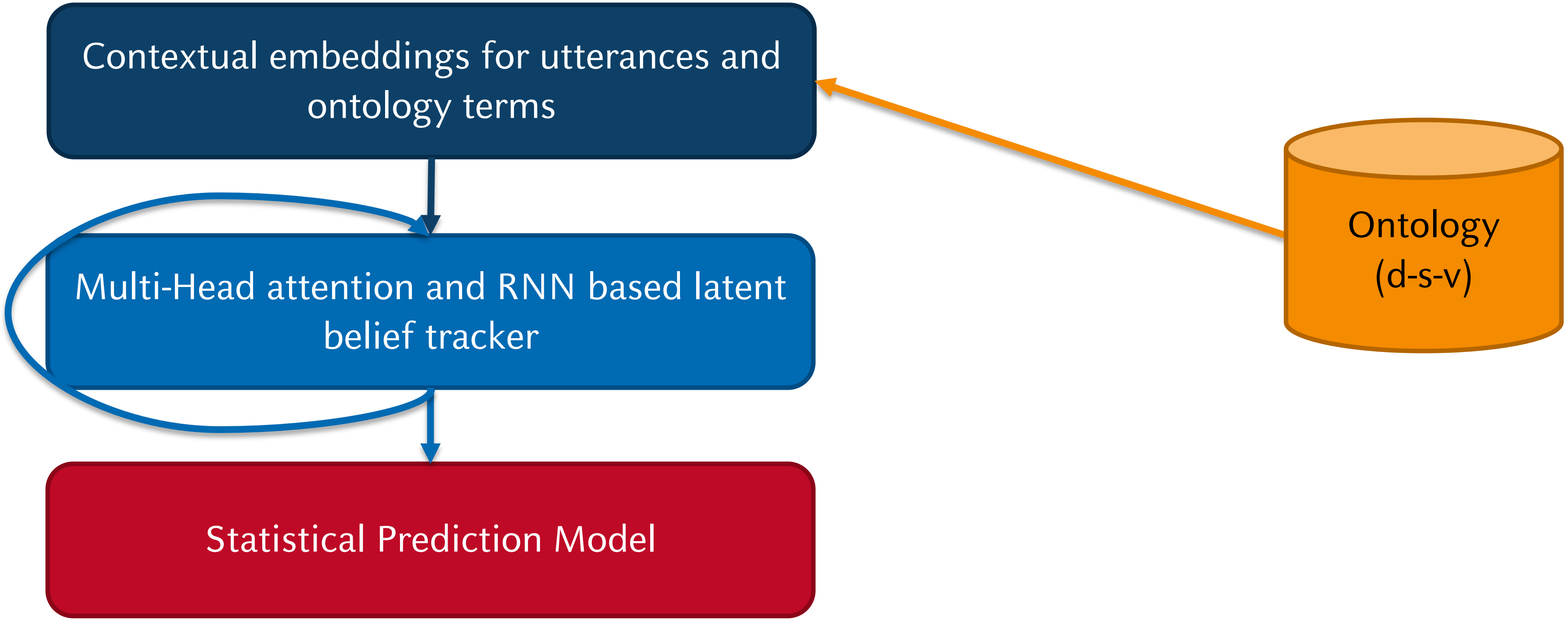
Model	Slot Accuracy	Joint-goal Accuracy
GCE	98.42%	**36.57%**
TRADE	96.42%	**48.62%**

- Positives:
  - **Generates values** with great success.
  - Shows promise with **few-shot** learning.
- Limitation:
  - **Zero-shot** performance not great.
  - **Past L** turns used. (Inefficient)
  - Requires domain-slots to be defined

Improved latent space mappings

- **Transformer** based **contextual** mappings
- Truly **statistical latent** space belief tracker

Overview:


Contextual embeddings for utterances and ontology terms
Multi-Head attention and RNN based latent belief tracker
Statistical Prediction Model
Ontology
(d-s-v)

Utterance Encoder

- Two fine-tuned BERT models:
  - **Utterance** embedding
  - **Domain-slot-value** embedding
- Use of **contextual** embeddings

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Utterance Encoder

[CLS] [User Utterance] [SEP]
[System Utterance] [SEP]

$\boldsymbol{h}_0$ $\boldsymbol{h}_1$ $\boldsymbol{h}_2$ … $\boldsymbol{h}_m$

$\oplus$

$\boldsymbol{H}_t$

[CLS] [Domain-Slot] [SEP]

[CLS] [Value] [SEP]

$\boldsymbol{h}_{[CLS]}$ → $\boldsymbol{q}_{slot}$

$\boldsymbol{h}_{[CLS]}$ → $\boldsymbol{y}_t^S$

Multi-head Attention

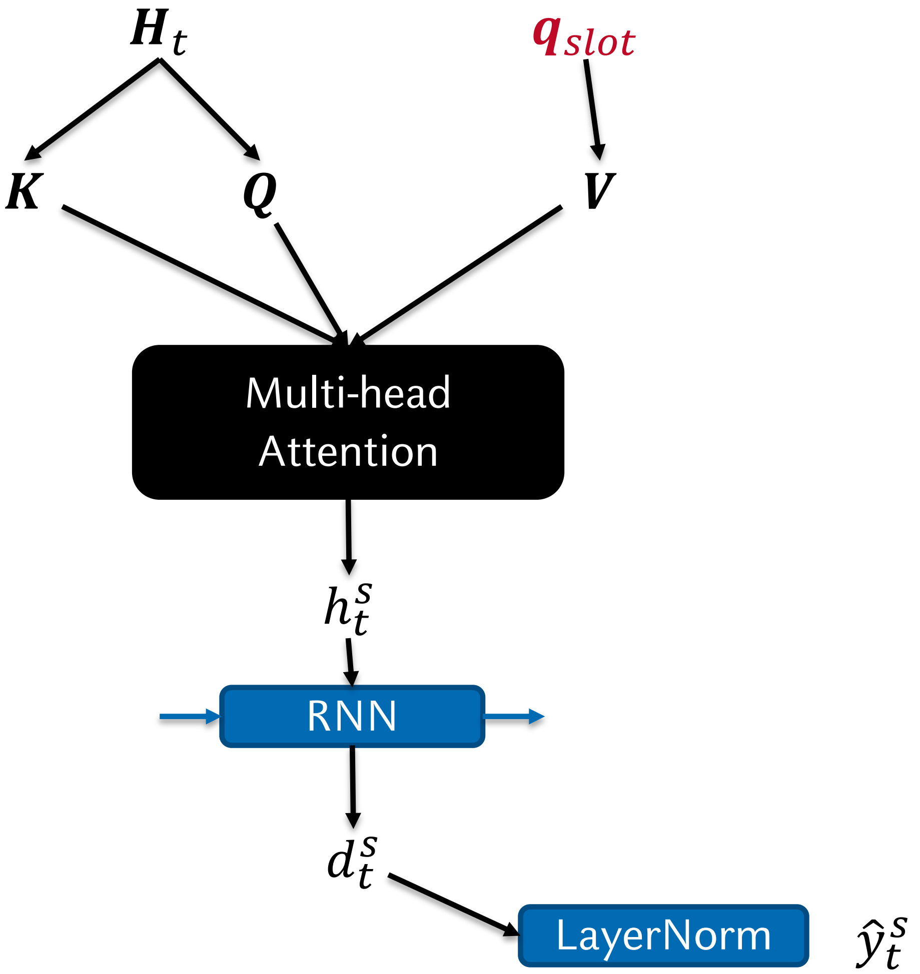
- Input:
  - **Query** – What is the encoder **asking**?
  - **Key** – The state of the encoder. Key **unlocks** the **answer**.
  - **Value** – How much **attention** should we give?
- Passed through **multiple attention heads**.
- Returns **context** embedding.

Statistical Update Model

- **Query** = **System** utterance
- **Key** = **User** utterance
- **Value** = **Domain-slot**
- The **attention heads** provides the **context** of the dialogue.
- **RNN** tracks context over dialogue.
- Provides a **estimated contextual value** embedding.

HEINRICH HEINE UNIVERSITÄT DÜSSELDORF

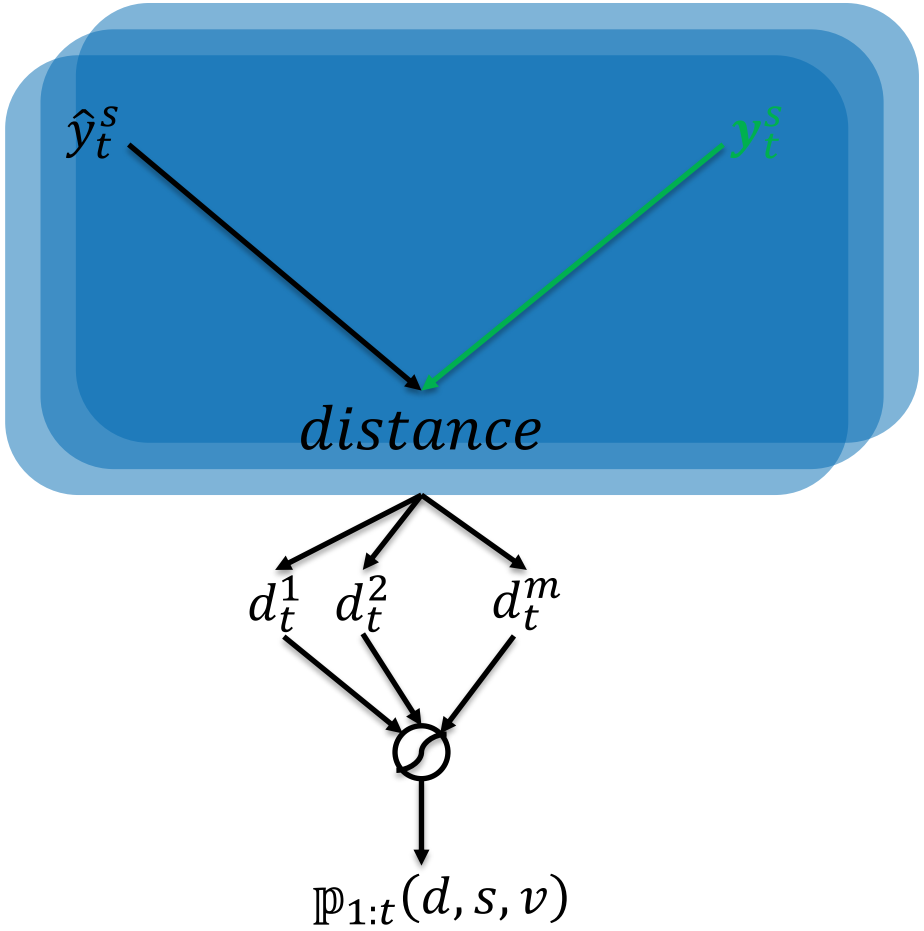
Statistical Update Model



Estimated value embedding

HEINRICH HEINE UNIVERSITÄT DÜSSELDORF

Statistical Prediction Model



Slot-Utterance Matching for Universal BT(SUMBT)

Overview

Model	Slot Accuracy	Joint-goal Accuracy
TRADE	96.42%	**48.62%**
SUBMT	96.44%	**46.65%**

- Positives:
  - True **latent space** fully statistical belief tracker.
- Limitation:
  - Very large model. **Expensive** to train.
  - Requires **ontology** to be defined.

Slot-Utterance Matching for Universal BT(SUMBT)

Overview

Dialog Example

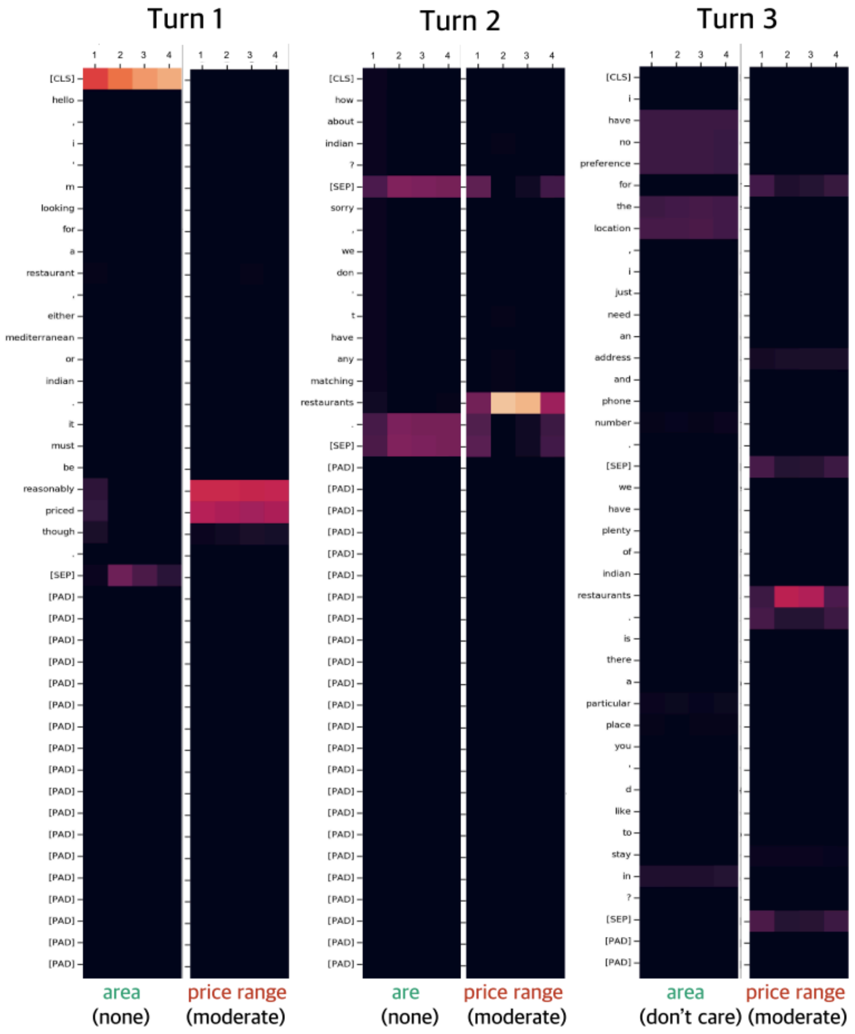
Turn 1, U: Hello, I'm looking for a restaurant, either Mediterranean or Indian, it must be reasonably priced though.

Turn 2, S: Sorry, we don't have any matching restaurants.

U: How about Indian?

Turn 3, S: We have plenty of Indian restaurants. Is there a particular place you'd like to stay in?

U: I have no preference for the location, I just need an address and phone number.



Success stories

- Word embeddings
  - Improved **performance**
  - Better **Scalability**
  - Successes of **contextual** embeddings
- Recurrent models
  - **Fully statistical**
  - Learns cross-turn dependencies
  - **No rules** needed

Success stories

- Semantic similarity
  - Leveraged from word **embeddings**
  - Does the user/system **mention** a concept?
- Projecting **dialogue history** onto latent **representation**.
- Knowledge sharing
  - **Parameter** sharing
  - Domains **share** slots, Slots **share** values
  - Improved **performance**
  - **Adaptability**
  - **Scalability**

Success stories

- Value generation methods
  - More **scalable**
  - Improved **performance**
  - **Adaptability**
  - Ontology only needs **domains and slots**
- **Joint** Belief Tracking and Policy Learning
  - Promises to improve **performance**

Shortcomings

- Predefined **ontology**
  - Not scalable
  - Not possible - **new values** can constantly be added (Restaurant names)
- **Zero-shot** adaption
  - Very little success
- **Rare** slot-value combinations
  - Difficulty accurately predicting these
  - Negatively impacts **joint goal accuracy**
  - Limits **adaptability**

Shortcomings

- Utilising **non-dialogue** data
  - Utilising non dialogue data through word embeddings.
- **Representation** of states
  - How to represent states
  - Is **domain-slot-value** sufficient
  - Could **graph** structures states be **embedded**
  - **Efficient** use of data for **rare states**
- Joint goal on **rich and noisy** datasets

Resources

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

- HyST: A Hybrid Approach for Flexible and Accurate Dialogue State Tracking
  R Goel, S Paul and D Hakkani-Tür, 2019
- Neural Belief Tracker: Data-Driven Dialogue State Tracking
  N Mrkšić, D Séaghdha, T Wen, B Thomson and S Young 2016
- The Dialog State Tracking Challenge: A Review
  JD Williams, A Raux, D Ramachandran, and A Black 2013
- SUMBT: Slot-Utterance Matching for Universal and Scalable Belief Tracking
  H Lee, J Lee and T Kim 2019
- Dialog State Tracking: A Neural Reading Comprehension Approach
  S Gao, A Sethi, S Aggarwal, T Chung and D Hakkani-Tür 2019
- Improving Dialogue State Tracking by Discerning the Relevant Context
  S Sharma, PK Choubey and R Huang 2019
- Large-Scale Multi-Domain Belief Tracking with Knowledge Sharing
  O Ramadan, P Budzianowski and M Gašić 2018

Resources

HEINRICH HEINE UNIVERSITÄT DÜSSELDORF

- Toward Scalable Neural Dialogue State Tracking Model

  E Nouri and E Hosseini-Asl 2018

- Transferable Multi-Domain State Generator for Task-Oriented Dialogue Systems

  A Madotto, E Hosseini-Asl and C Xiong 2018

- Towards Universal Dialogue State Tracking

  L Ren, K Xie, L Chen and K Yu 2019

- BERT-DST : Scalable End-to-End Dialogue State Tracking with Bidirectional Encoder Representations from Transformer

  G Chao and I Lane 2019

- Global-Locally Self-Attentive for Dialogue State Tracking

  V Zhong, C Xiong and R Socher 2019

- Fully Statistical Neural Belief Tracking

  N Mrkšić Nikola and I Vulić 2018

- Word-Based Dialog State Tracking with Recurrent Neural Networks

  M Henderson, B Thomson and S Young 2015