Simultaneously Improving Utility and User Experience in Task-oriented Dialogue Systems

Phillip Lippe, Pengjie Ren, Hinda Haned, Bart Voorn, Maarten de Rijke
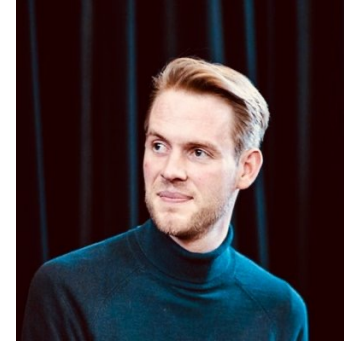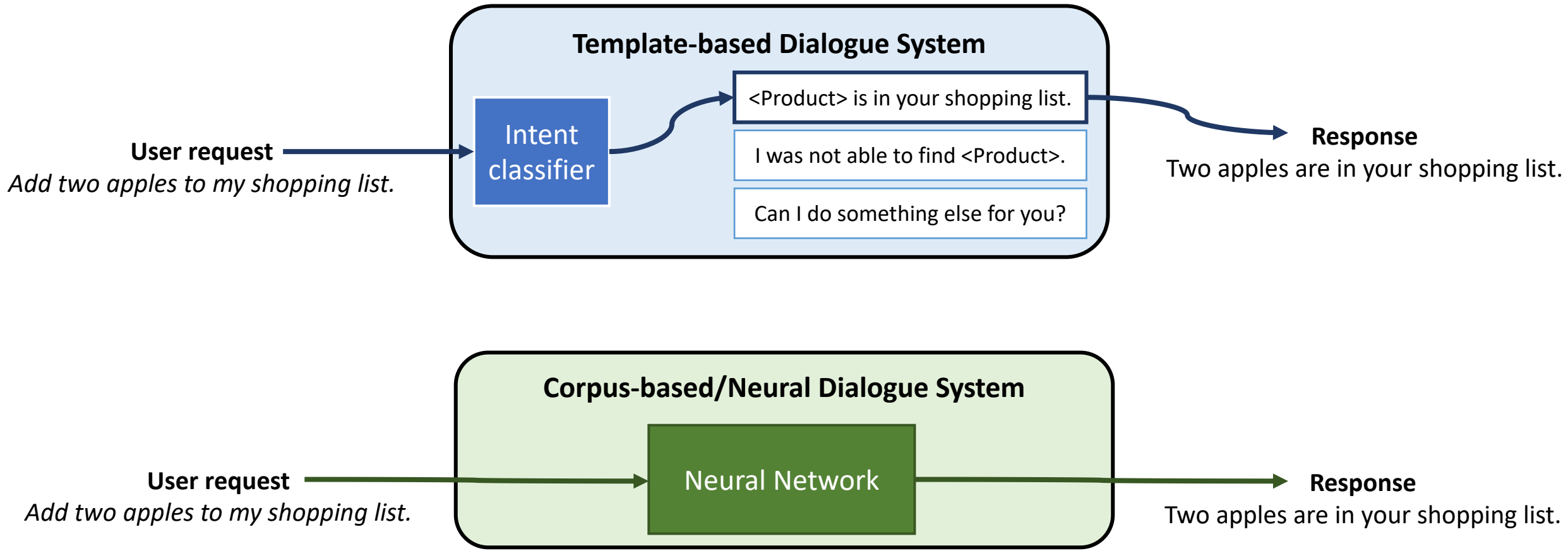
Introduction
Motivation

How can we adapt automated dialogue systems to have more natural conversations?
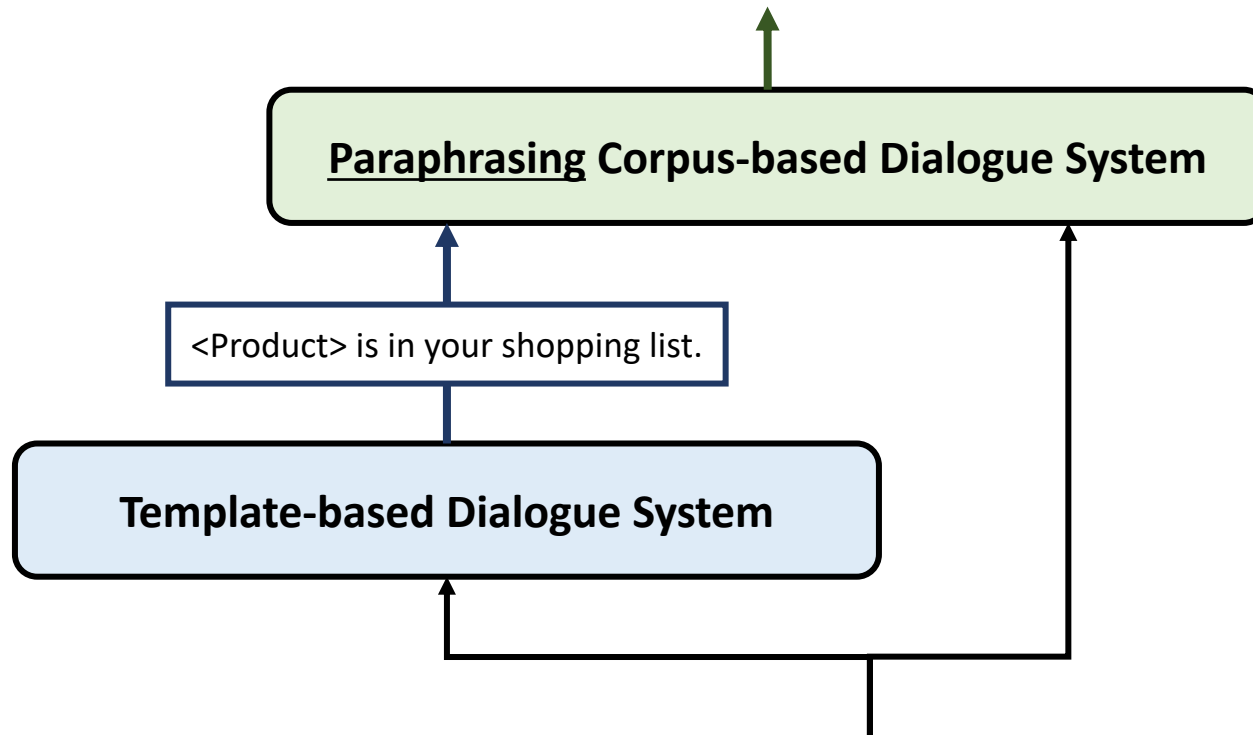
Introduction
Motivation

Introduction
Approach

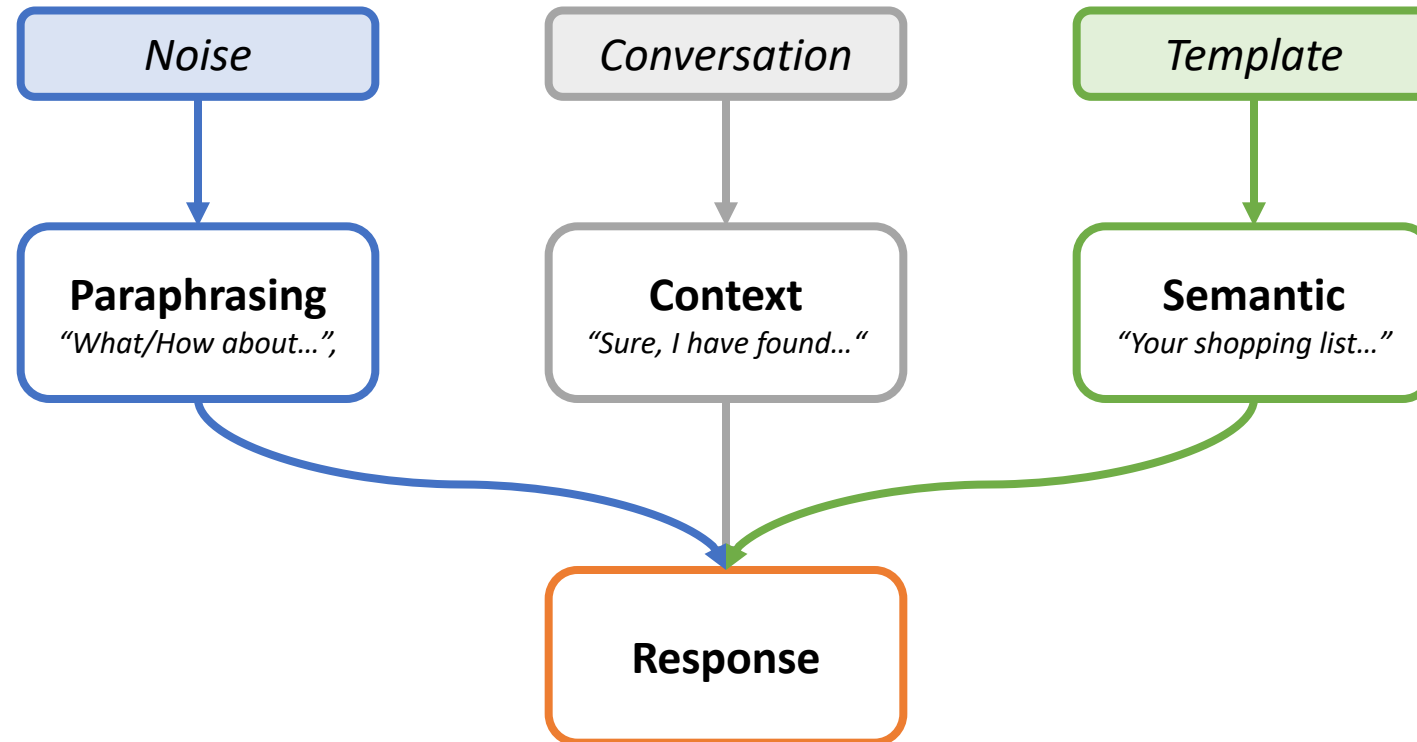**Responses**
- *Sure, two apples are added to your shopping list. How else I can help you with?*
- *I added two apples. Anything else?*

Paraphrasing Corpus-based Dialogue System

<Product> is in your shopping list.

Template-based Dialogue System

User request
- *Can you add two apples to my shopping list?*
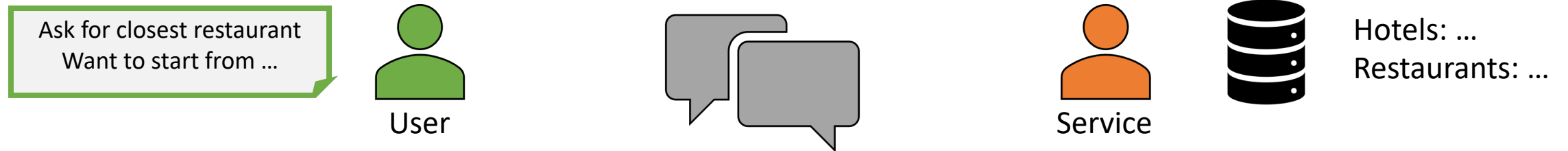- *Two apples please.*

Introduction
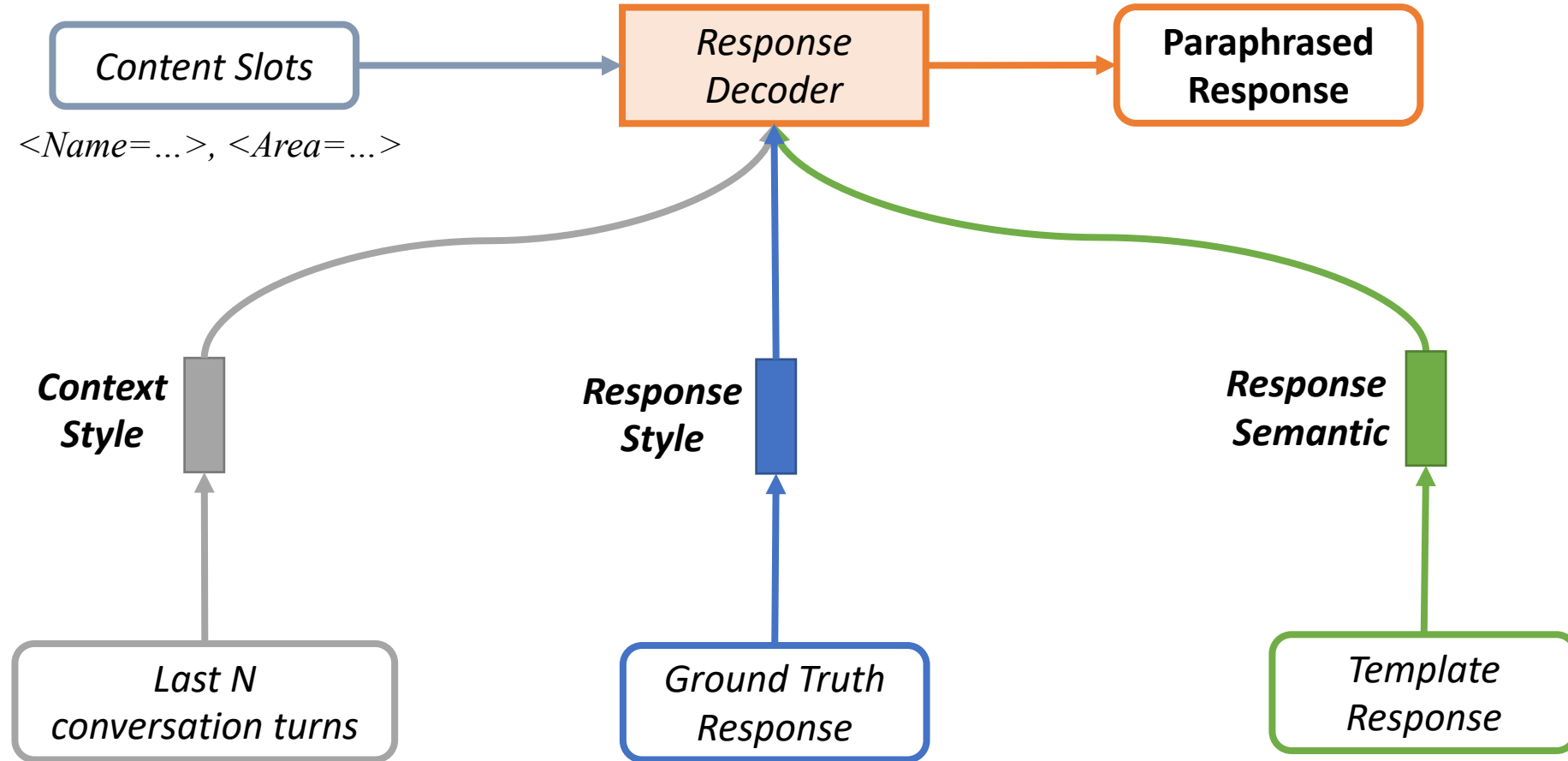## Approach

Model
Dataset

- MULTIWOZ (Multi-domain Wizard of Oz) dataset

- Human-to-Human dialogues

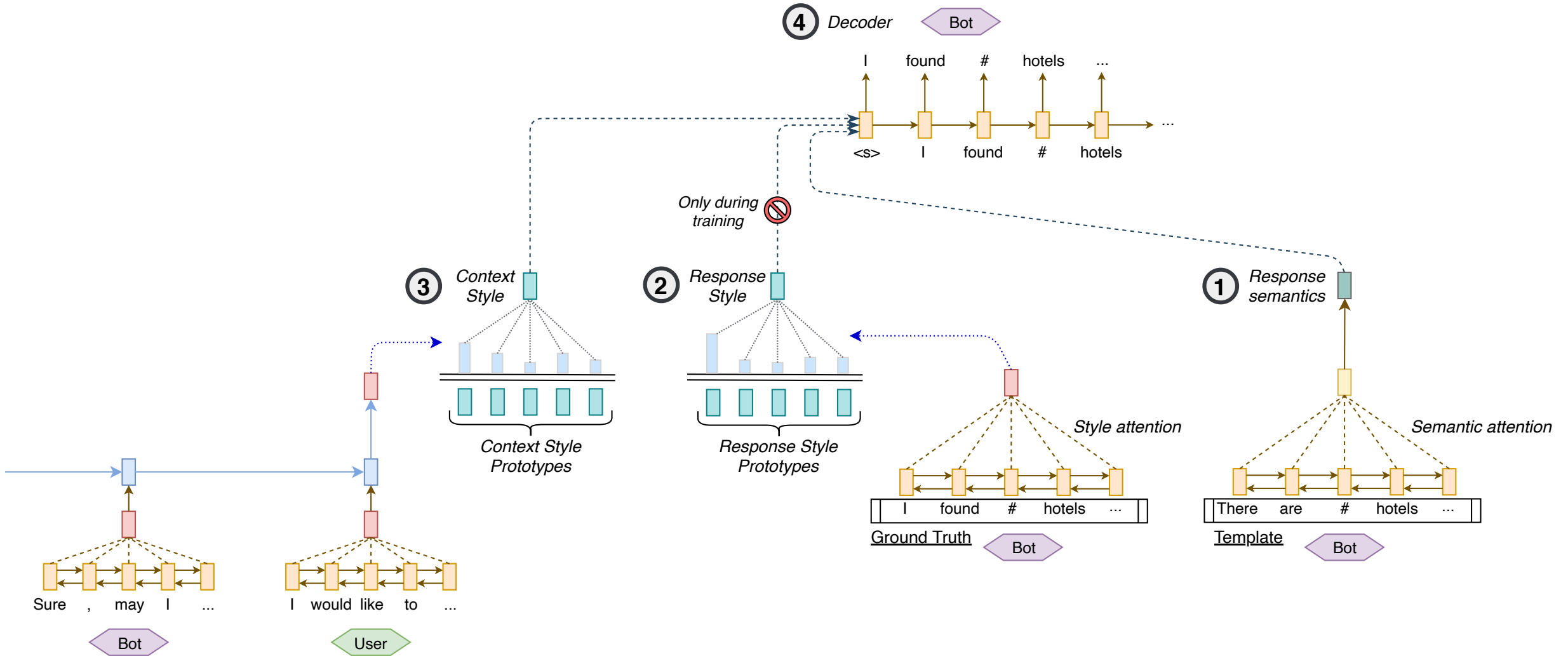- Assumption: Two responses with the same dialogue actions are paraphrases in different contexts

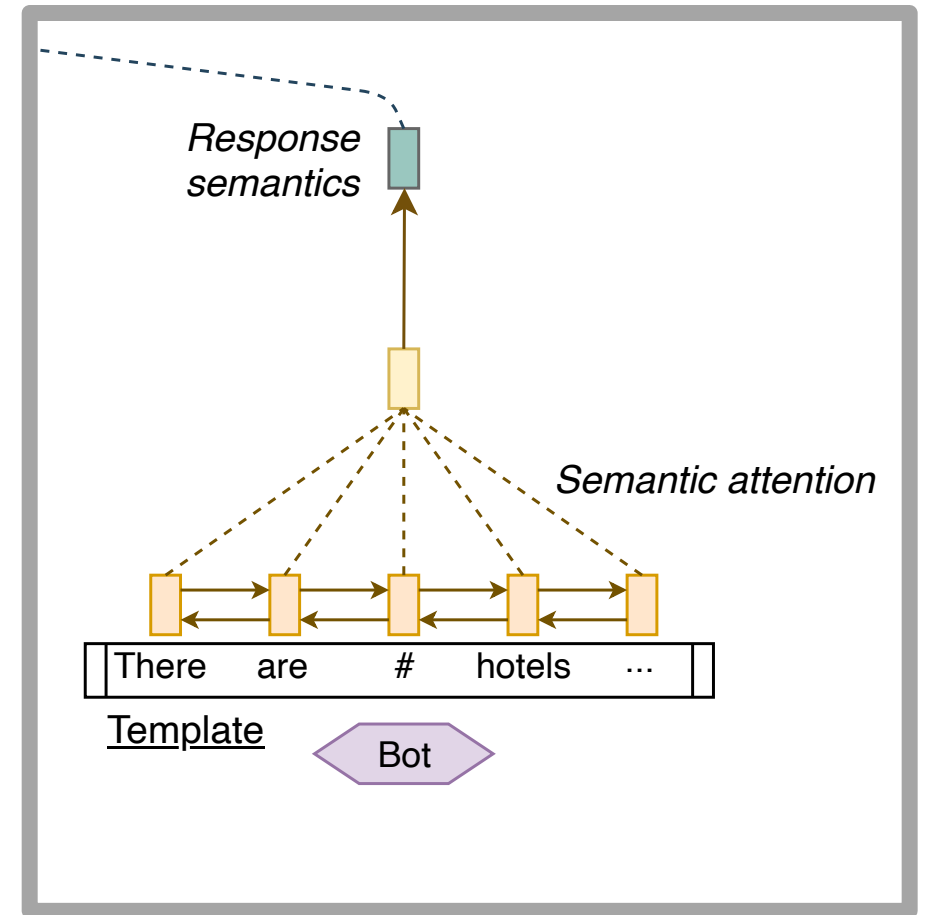# Model architecture
Overview

Model architecture
P2-Net

Simultaneously Improving Utility and User Experience in Task-oriented Dialogue Systems

Model architecture
Response semantic

- Bi-LSTM encoder with global attention

- Single, fixed-sized feature vector representation

Model architecture
Response Style

- Bi-LSTM encoder with global attention

- Prototype layer
  - Predict weighted sum of prototype styles

- Sample style during inference

Response Style

Response Style Prototypes

Style attention

| I | found | # | hotels | ... |

Ground Truth

Bot

Model architecture
Context Style

- Hierarchical RNN

- Prototype layer for noise reduction
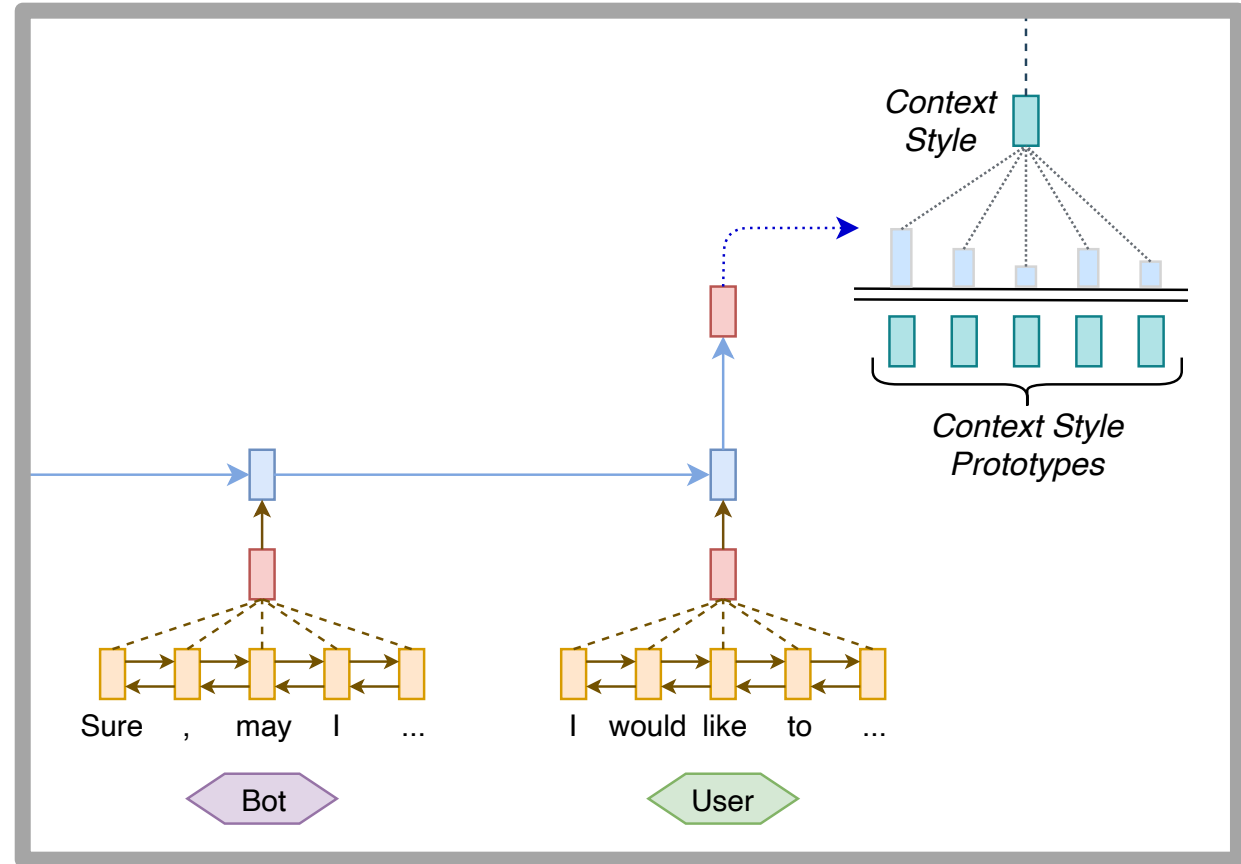
Model architecture
Response Generation

- Autoregressive language model

- Pointer network architecture for integrating slots
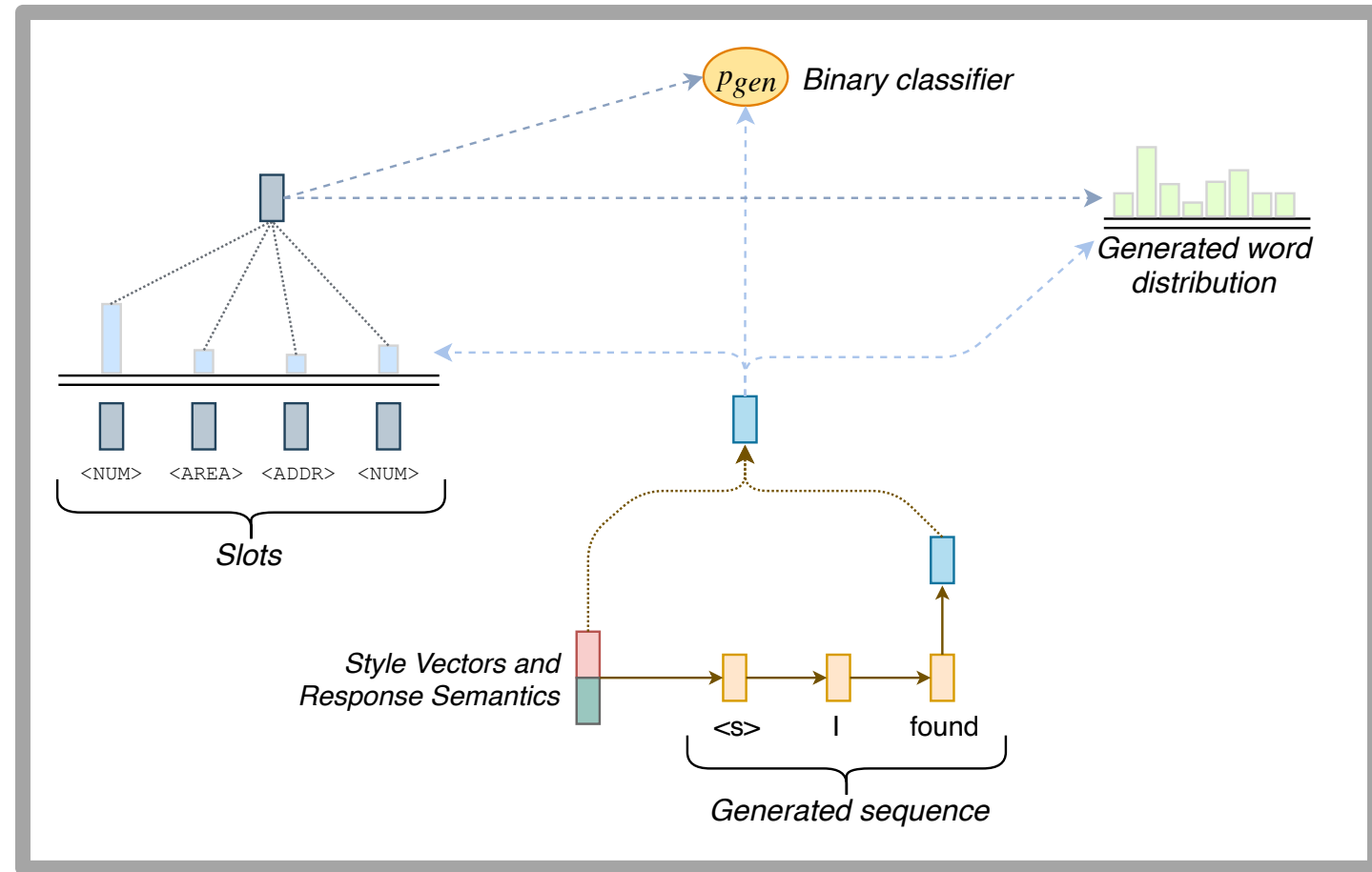
Model architecture
P2-Net

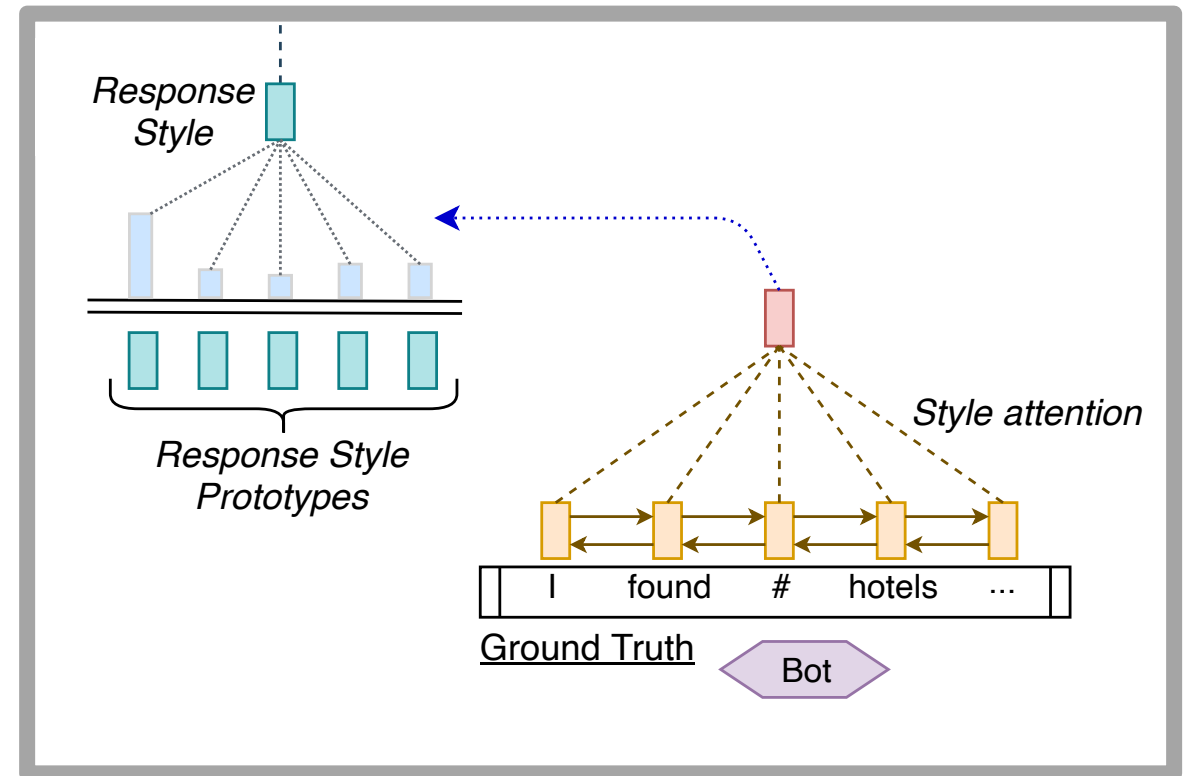Simultaneously Improving Utility and User Experience in Task-oriented Dialogue Systems

Experimental results
Research Questions

(RQ1) Can P2-Net generate more diverse responses than post-processing methods?

(RQ2) Can P2-Net paraphrase the template without changing its semantic?

Experimental results
RQ1 - Qualitative results

Dialogue action: Offer two choices for booking a table at a restaurant.
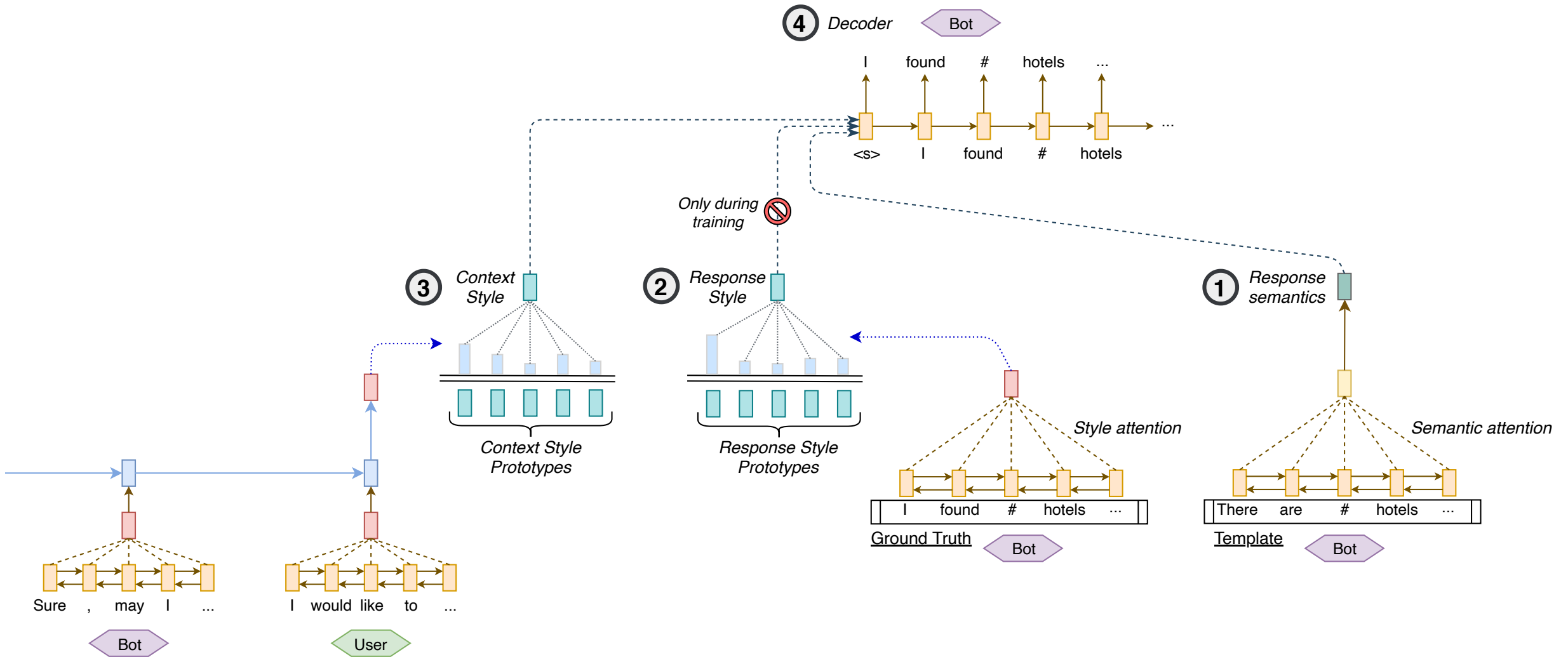
Template: I could try the <name="charlie chan"> , or <name="the golden house"> for you, if you wish .
Context: Can you book a table for seven people on Thursday at 15:00 ?
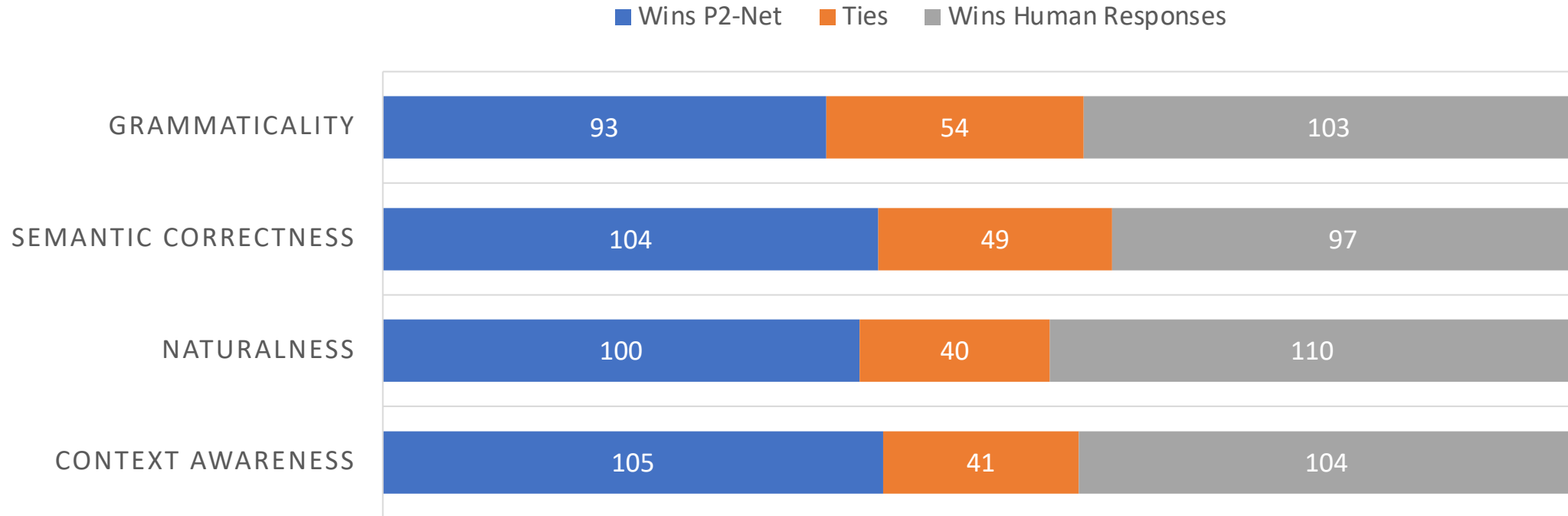**Slots:** <name="La Mimosa">, <name="Shiraz">

Diverse generations from P2-Net	Diverse generations from stochastic beam search
(1) Would you like to try La Mimosa or Shiraz?	(1) You can choose from La Mimosa , Shiraz .
(2) The La Mimosa and Shiraz are both available	(2) You can choose from La Mimosa or Shiraz.
(3) What about La Mimosa or Shiraz?	(3) you can choose from La Mimosa, or Shiraz.
(4) I can recommend the La Mimosa or Shiraz.	(4) Sure, we have the La Mimosa or Shiraz.
(5) Okay . Would you like to try La Mimosa or Shiraz?	(5) You can choose from La Mimosa, or Shiraz.
(6) How about the La Mimosa or Shiraz?	(6) Sure, we have the La Mimosa and the Shiraz.
(7) I have the La Mimosa, Shiraz.	(7) Sure, we have the La Mimosa and the Shiraz. Do you want to book a of them?
(8) Okay, I have two options for you. La Mimosa and Shiraz.	(8) Sure, we have La Mimosa and Shiraz. Do you want to go?

Experimental results
RQ2 - Human Evaluation

Legend: Wins P2-Net | Ties | Wins Human Responses

Category	Wins P2-Net	Ties	Wins Human Responses
GRAMMATICALITY	93	54	103
SEMANTIC CORRECTNESS	104	49	97
NATURALNESS	100	40	110
CONTEXT AWARENESS	105	41	104

⟹ P2-Net can achieve human-like user experience

Conclusion
Summary

- Combine template- and corpus-based dialogue systems for diverse, natural conversations

- Unsupervised learning of decomposing a response into three independent parts: semantics, context style and paraphrasing noise

- Sampling a response style shows a significantly higher diversity than post-processing methods

- Simplifying the task for the corpus-based model supports accurate language generation
  $\implies$ Less grammar mistakes