Improving **Long Distance Slot Carryover** in Spoken Dialogue Systems

Tongfei Chen*, Chetan Naik, Hua He, Puspendre Rastogi, Lambert Mathias

Alexa AI, Amazon.com

** Johns Hopkins University*

Task: Contextual carryover

- U: Find me a Mexican restaurant in Portland
  PLACETYPE CITY
- A: A few Mexican restaurants in Portland are A, B and C
  PLACETYPE CITY
- U: What movies are in theaters there
  MEDIATYPE PLACETYPE ANAPHOR

Problem definition

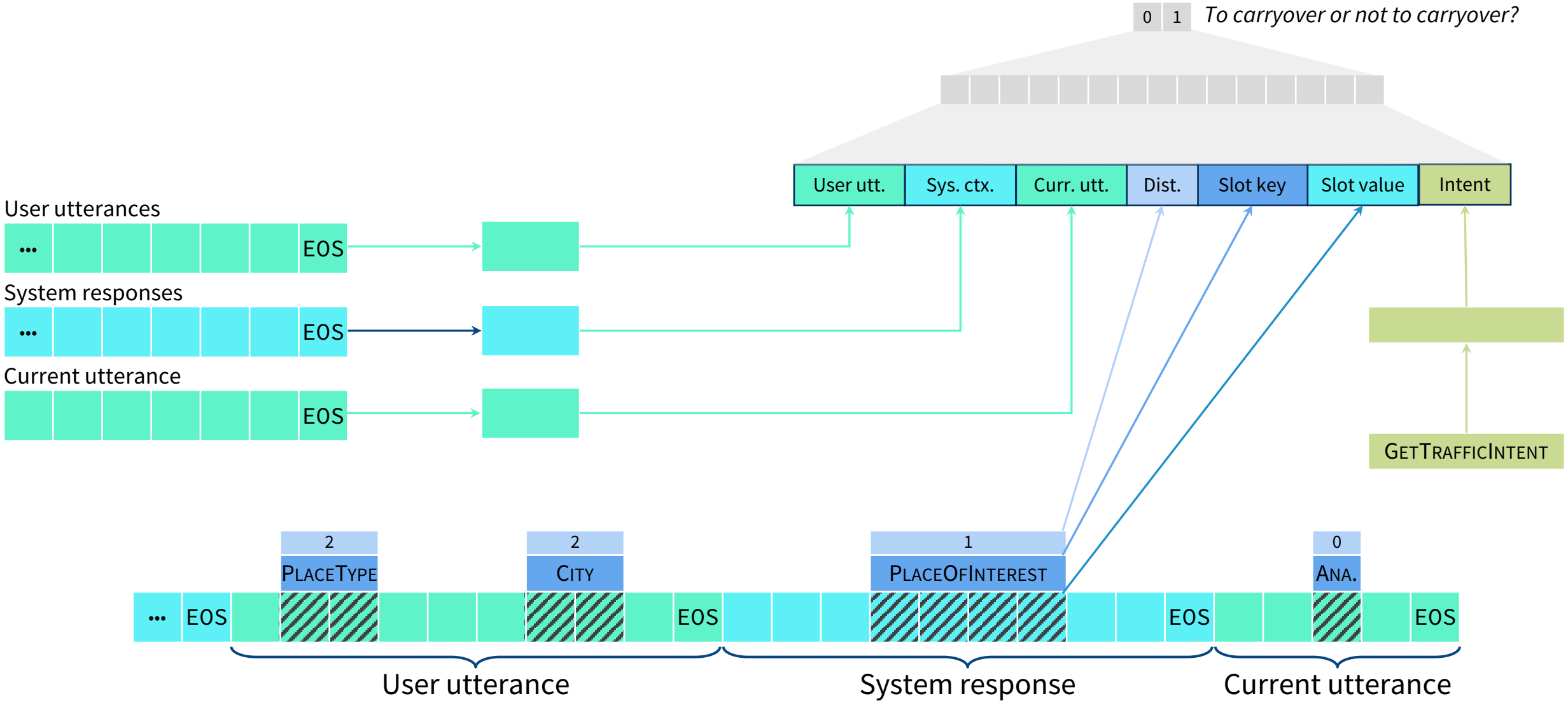
- A dialog session between a user and an agent can be represented as

$$\{(u_w, a_w), (u_{w-1}, a_{w-1}), \cdots, (u_1, a_1), u_0\}$$

- where $u_i = (t_i^u, s_i^u, I_i^u)$
  - *t*: ASR transcription of an utterance: represented as a sequence of tokens
  - *s*: Slots detected
    - Might be the results from an upstream candidate generation system
  - *I*: Intent inferred from the NLU system

Previous work

(Naik et al., 2018)



Motivation

- U: What's the thermostat set to
  APPLIANCE
- A: The thermostat is set to 76 degrees
- U: Turn it down to 72 degrees
  SETTINGVALUE

Motivation

- U: What's adele 's latest album?
  ARTISTNAME
- A: It is 25
  ALBUMNAME
- U: play it

Motivation

- Modeling slot interdependence instead of isolated decisions

- Instead of doing independent decisions

$$f : \texttt{Slot} \rightarrow \{0, 1\}$$

- Predict slots to be carried over in one round

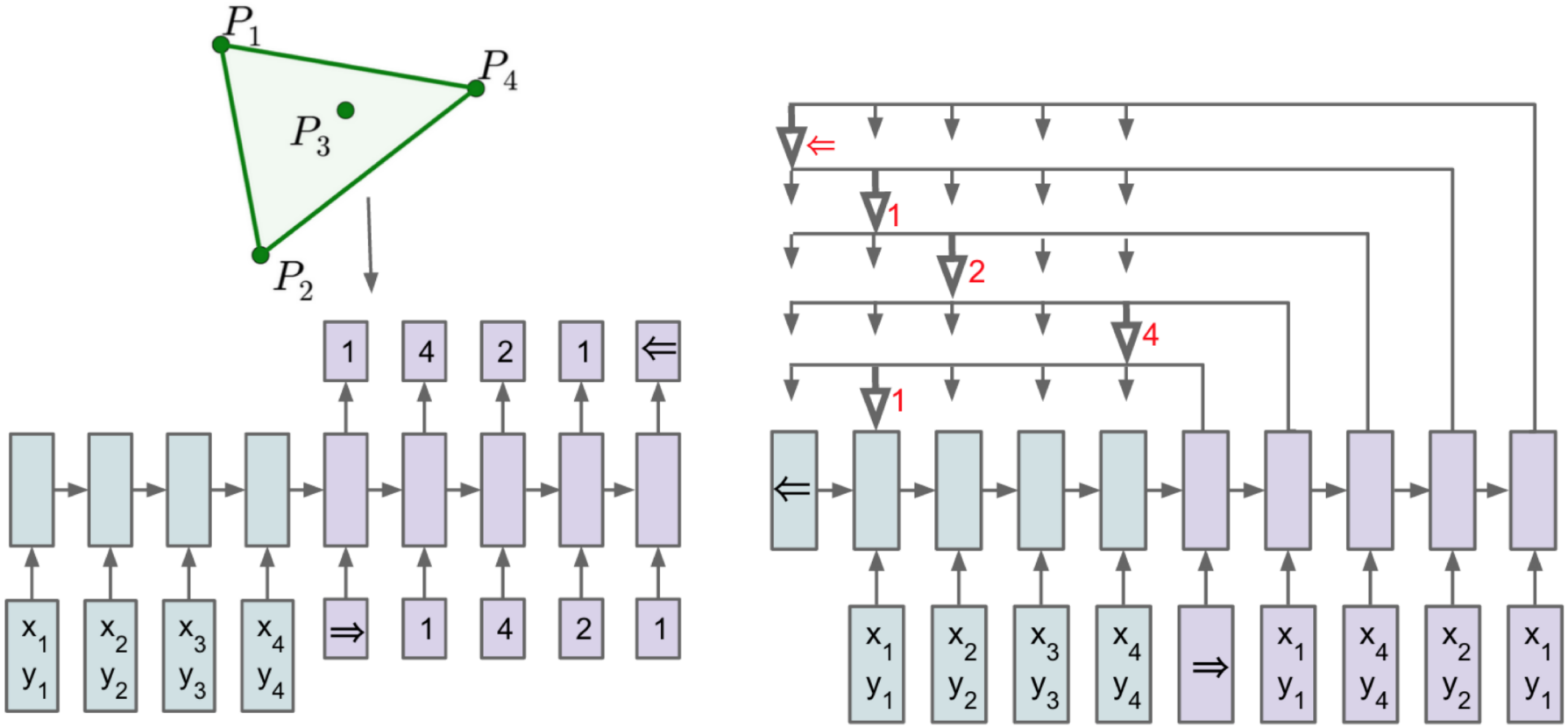
$$f : \texttt{Set}\langle\texttt{Slot}\rangle \rightarrow \texttt{Set}\langle\texttt{Slot}\rangle$$

- where a subset is selected from the input

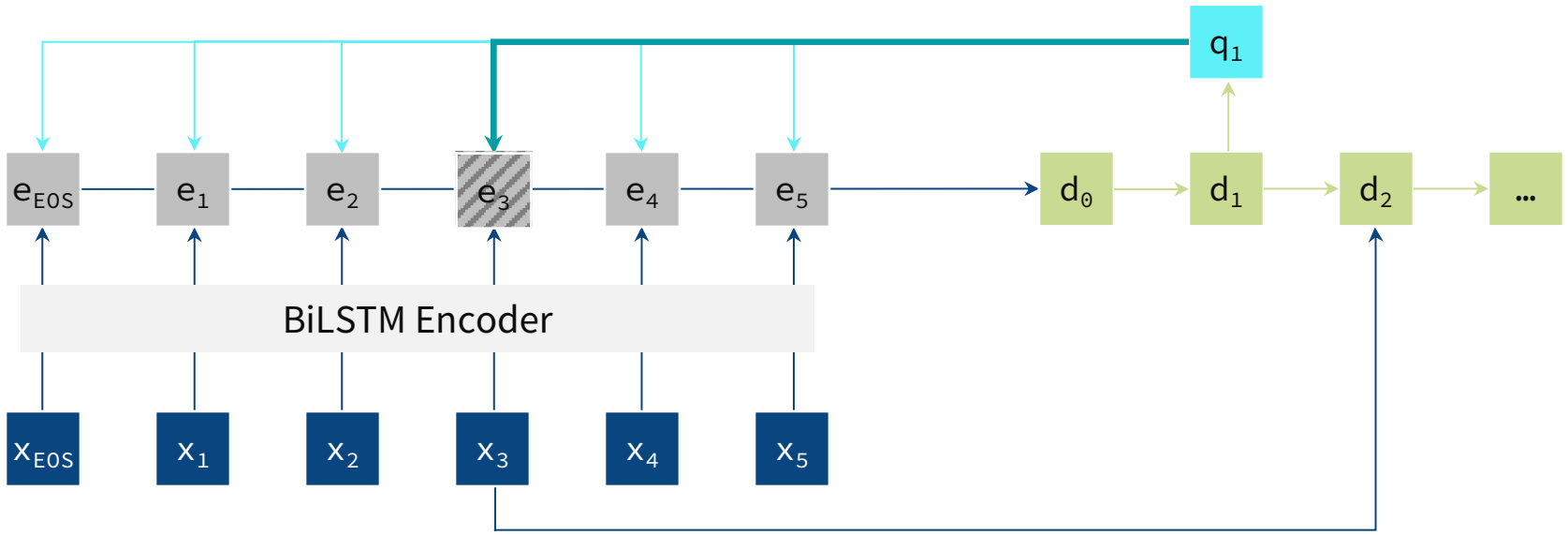
$$y = f(x); y \subseteq x$$

Pointer networks

- Vinyals et al. (2015)
- Modified Seq2Seq that generates a subset of the original sequence



Pointer networks for subset selection



Pointer networks for contextual carryover

- Select a subset of candidate slots detected in the dialog

- Input: candidate slot set $X = \{s_1, \cdots, s_n\}$
- Output: selected slots to carryover $Y \subseteq X$

- With additional external information:
  - Current utterance
  - Past history
  - Intent

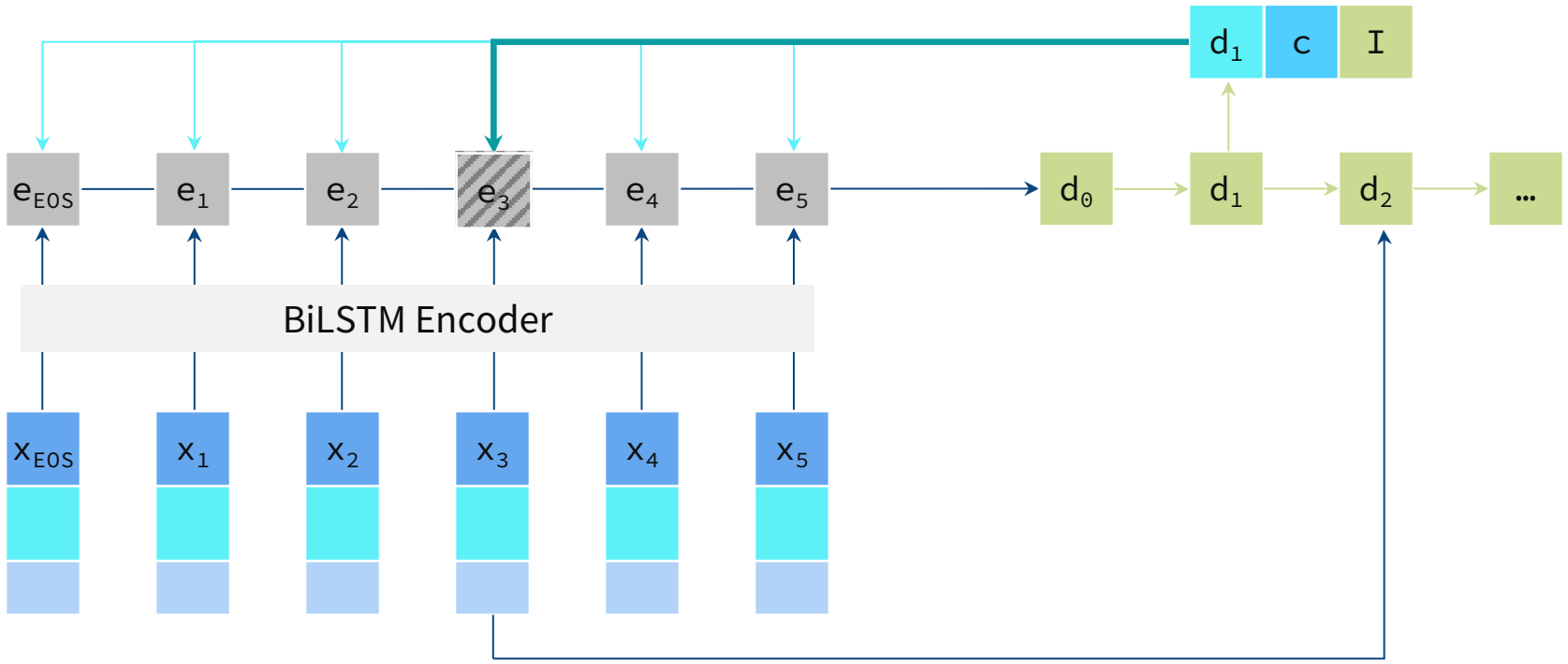
Enforcing order on candidate slots

- Pointer networks are adapted from Seq2Seq
- Input and output are all ORDERED

- We can enforce a *temporal order* on the slots

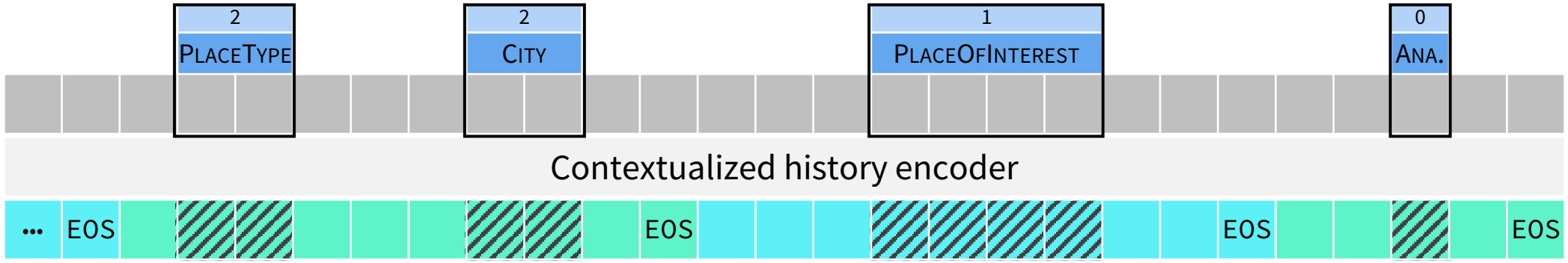
- Order inputs reversely and outputs in normal order (LIFO property)

Pointer networks for contextual carryover



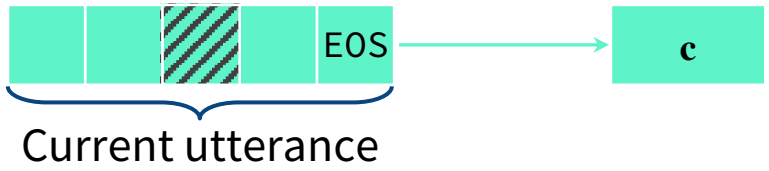
Contextualized slot embeddings

- Slot value encoding: averaged from the words after a contextualized encoder

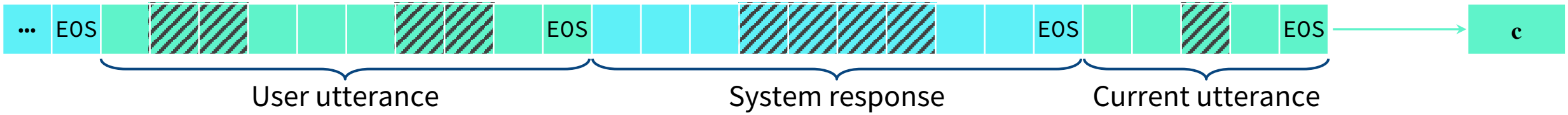


Context vector in queries

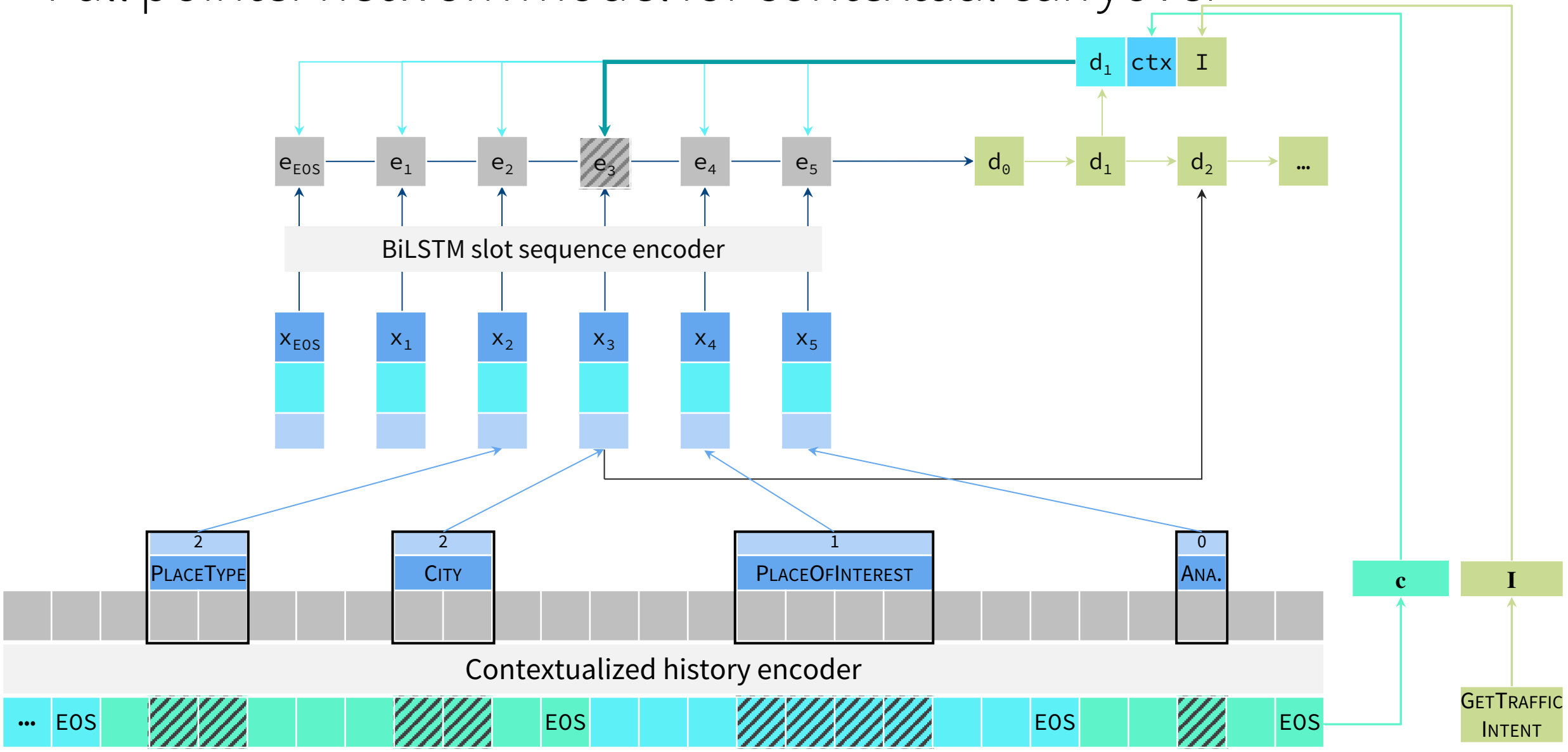
- Could be either the current utterance



- Or the whole history



Full pointer network model for contextual carryover

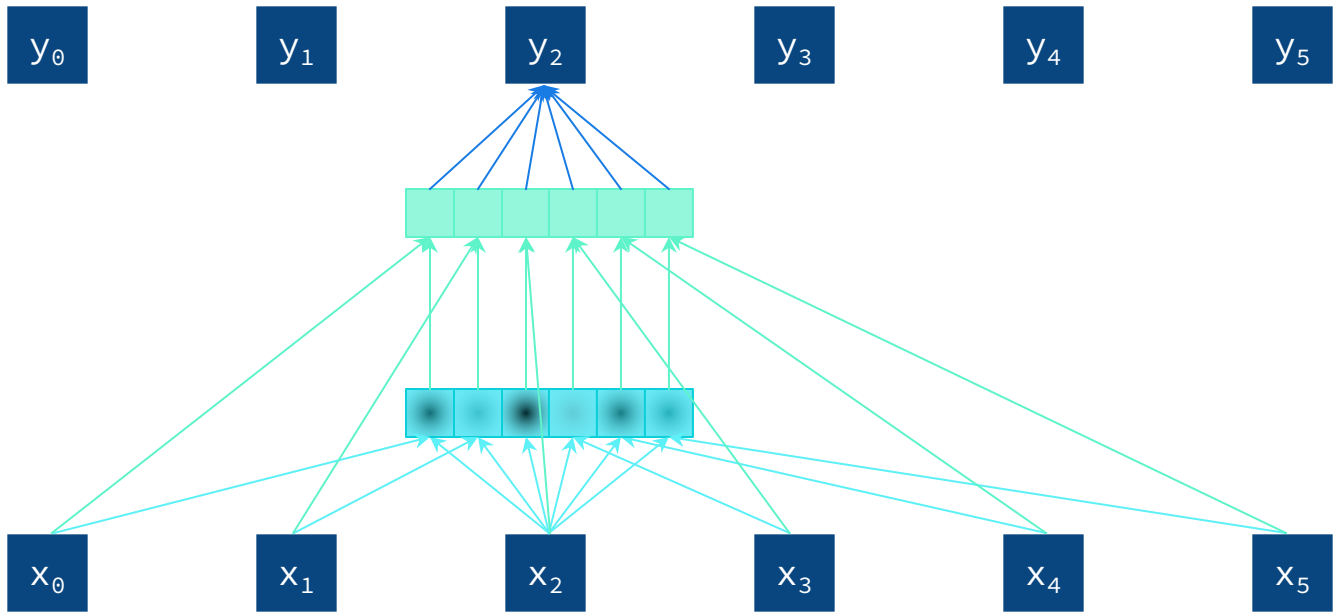


Self-attention model

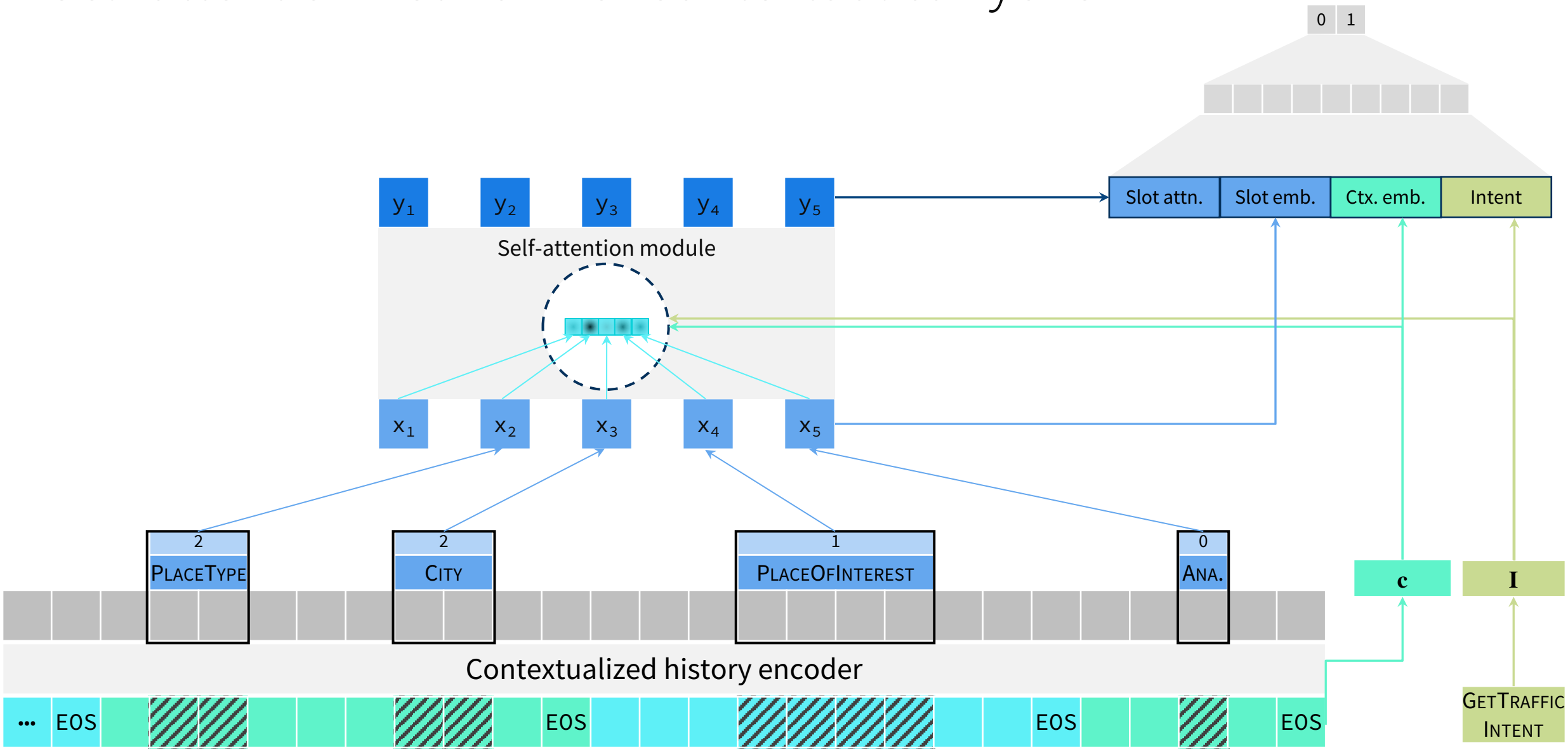
- An order is enforced on the slots
- What if we completely forgo order?
- Remove autoregressive encoders/decoders
- Parallel – may leads to faster performance

- Self-attention (Vaswani et al., 2017)

Self-attention module



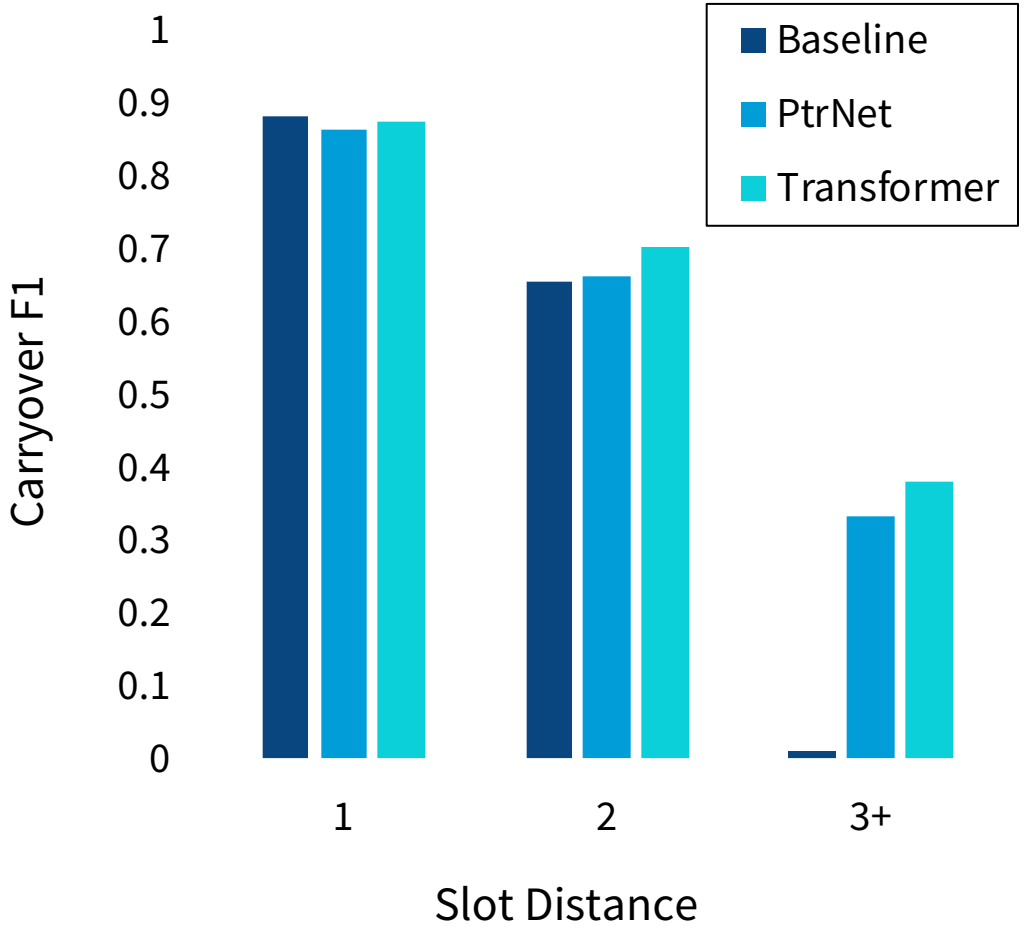
Self-attention network for contextual carryover



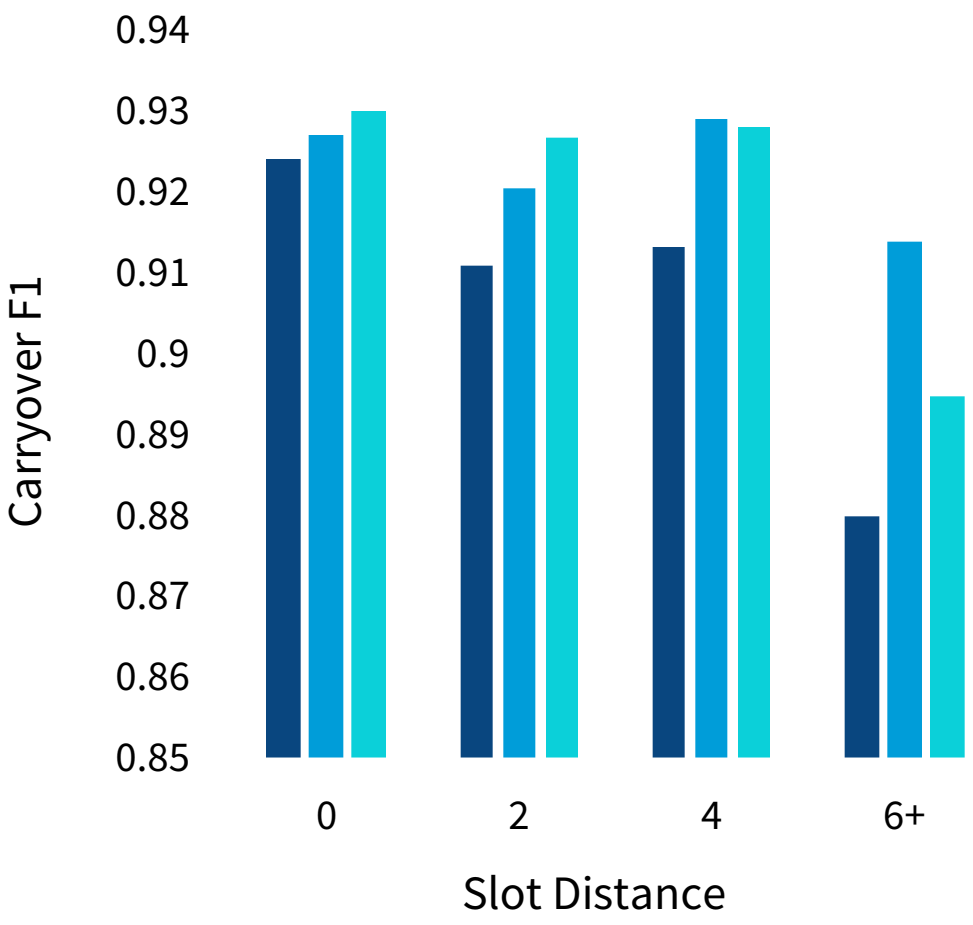
Datasets

- Alexa Internal
  - 156k dialogues from various domains
  - Music, Q&A, Video, Weather, Local Businesses, Home Automation
- DSTC2
  - Top ASR hypothesis as the user utterance
  - All slots from SLU with score > 0.1 as candidate slots

Results on Alexa Internal Dataset



Results on DSTC2



Summary

- Jointly models contextual slots
- Subset decoding:
  - Via pointer networks
  - Via transformers

- Leads to improved long distance slot carryover