香港中文大學
The Chinese University of Hong Kong

Tencent AI Lab

Interconnected Question Generation with Coreference Alignment and Conversation Flow Modeling

Yifan Gao[1], Piji Li[2], Irwin King[1], Michael R. Lyu[1]

[1]The Chinese University of Hong Kong [2]Tencent AI Lab

ACL 2019, Florence

Question Generation: Background

Question Generation: Related Work

- Dialogue
  - Seeking Information in Task-oriented Chatbot
  - Asking Clarification Questions (Rao and Daume, 2018)
  - Interactiveness and Persistance (Wang et al, 2018)

Question Generation: Related Work

- Dialogue
  - Seeking Information in Task-oriented Chatbot
  - Asking Clarification Questions (Rao and Daume, 2018)
  - Interactiveness and Persistance (Wang et al, 2018)
- Question Answering
  - Reading Comprehension Question Generation (Du et al., 2017)
  - Harvesting Question Answer Pairs (Du et al., 2018)
  - Improving QA model (Yang et at., 2017)

Question Generation: Related Work

- Dialogue
  - Seeking Information in Task-oriented Chatbot
  - Asking Clarification Questions (Rao and Daume, 2018)
  - Interactiveness and Persistance (Wang et al, 2018)
- Question Answering
  - Reading Comprehension Question Generation (Du et al., 2017)
  - Harvesting Question Answer Pairs (Du et al., 2018)
  - Improving QA model (Yang et at., 2017)

Our Focus

Question Generation: Previous Setting

Incumbent Democratic President Bill Clinton was ineligible to serve a third term due to term limitations in the 22nd Amendment of the Constitution, and Vice President Gore was able to secure the Democratic nomination with relative ease.

Question Generation: Previous Setting

Incumbent Democratic President Bill Clinton was ineligible to serve a third term due to term limitations in the 22nd Amendment of the Constitution, and Vice President Gore was able to secure the Democratic nomination with relative ease.

Q: What political party is Clinton a member of?
A: Democratic
Q: What was Clinton ineligible to serve?
A: third term
Q: Why was he ineligible to serve a third term?
A: term limitations

Question Generation: Previous Setting

Incumbent Democratic President Bill Clinton was ineligible to serve a third term due to term limitations in the 22nd Amendment of the Constitution, and Vice President Gore was able to secure the Democratic nomination with relative ease.

Q: What political party is Clinton a member of?
A: Democratic
Q: What was Clinton ineligible to serve?
A: third term
Q: Why was he ineligible to serve a third term?
A: term limitations

Standalone Interaction ☹

Question Generation: Previous Setting

Incumbent Democratic President Bill Clinton was ineligible to serve

Human test knowledge through conversations involving interconnected questions

Democratic nomination with relative ease.

Q: What political party is Clinton a member of?
A: Democratic
Q: What was Clinton ineligible to serve?
A: third term
Q: Why was he ineligible to serve a third term?
A: term limitations

Standalone Interaction ☹

Question Generation: Previous Setting

Incumbent Democratic President Bill Clinton was ineligible to serve

Human test knowledge through conversations involving interconnected questions

Democratic nomination with relative ease.

Machines?

Q: What political party is Clinton a member of?
A: Democratic
Q: What was Clinton ineligible to serve?
A: third term
Q: Why was he ineligible to serve a third term?
A: term limitations

Standalone Interaction ☹

Question Generation + Conversation

Conversation is the ultimate way for human-machine interactions

Standalone Interaction

Q: What political party is Clinton a member of?

A: Democratic

Q: What was Clinton ineligible to serve?

A: third term

Q: Why was he ineligible to serve a third term?

A: term limitations

Question Generation + Conversation

Conversation is the ultimate way for human-machine interactions

Standalone Interaction	**Conversational Questions**
Q: What political party is Clinton a member of?	Q1: What political party is Clinton a member of?
A: Democratic	A1: Democratic
Q: What was Clinton ineligible to serve? →	Q2: What was **he** ineligible to serve?
A: third term	A2: third term
Q: Why was he ineligible to serve a third term? →	Q3: **Why**?
A: term limitations	A3: term limitations

Conversational Question Generation

- Our Goal
  - A system needs to ask a series of **interconnected** questions grounded in a passage through a question-answering style conversation
  - Every question after the first turn might be dependent on the conversation history.

Incumbent Democratic President Bill Clinton was ineligible to serve a third term due to term limitations in the 22nd Amendment of the Constitution, and Vice President Gore was able to secure the Democratic nomination with relative ease.

Q1: What political party is Clinton a member of?
A1: Democratic
Q2: What was **he** ineligible to serve?
A2: third term
Q3: **Why**?
A3: term limitations

Challenges

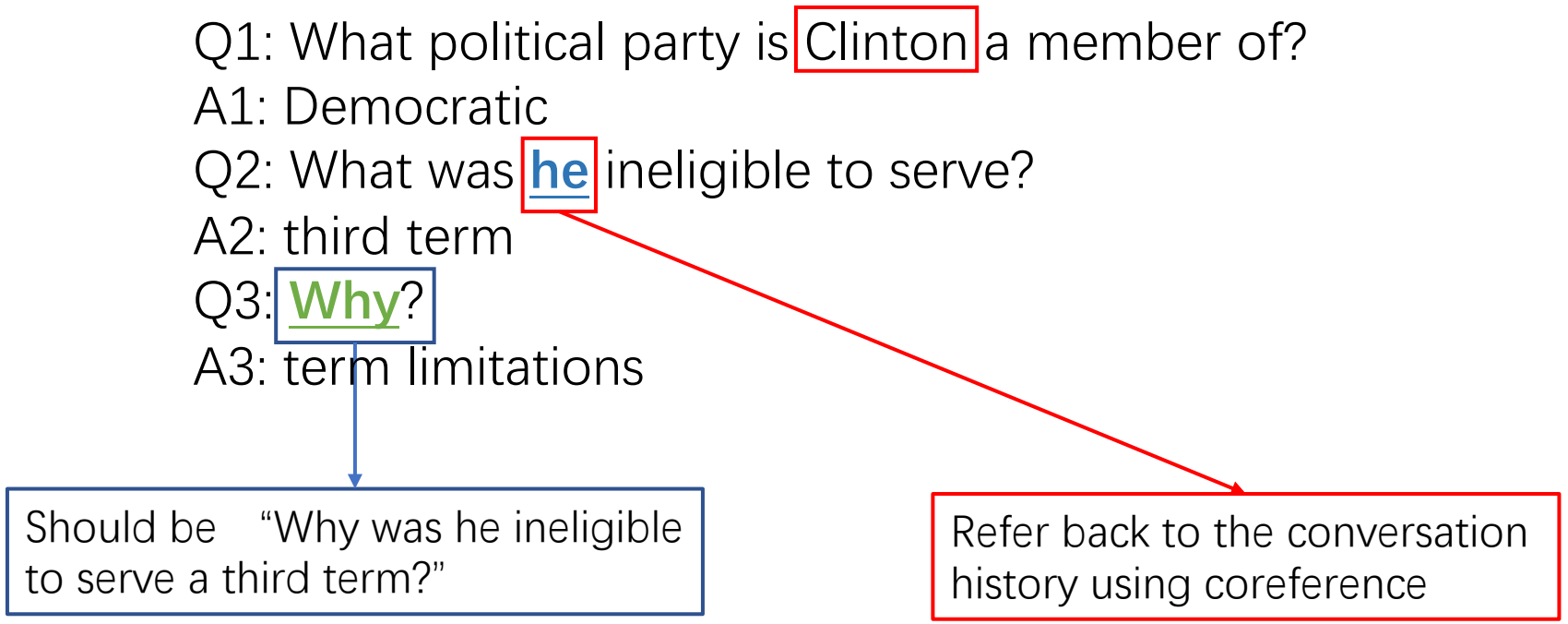
1. Generate **conversational interconnected** questions depending on the conversation so far

Q1: What political party is Clinton a member of?
A1: Democratic
Q2: What was **he** ineligible to serve?
A2: third term
Q3: Why?
A3: term limitations

Refer back to the conversation history using coreference

Challenges

1. Generate **conversational interconnected** questions depending on the conversation so far



Challenges

1. Generate **conversational interconnected** questions depending on the conversation so far
2. A coherent conversation must have **smooth transitions** between turns
   - We expect the narrative structure of passages can influence the conversation flow of our interconnected questions

Incumbent Democratic President Bill Clinton was ineligible to serve a third term due to term limitations in the 22nd Amendment of the Constitution, and Vice President Gore was able to secure the Democratic nomination with relative ease. Bush was seen as the early favorite for the Republican nomination and, despite a contentious primary battle with Senator John McCain and other candidates, secured the nomination by Super Tuesday.

Q1

Conversation Flow

Challenges

1. Generate **conversational interconnected** questions depending on the conversation so far
2. A coherent conversation must have **smooth transitions** between turns
   - We expect the narrative structure of passages can influence the conversation flow of our interconnected questions

Incumbent Democratic President Bill Clinton was ineligible to serve a third term due to term limitations in the 22nd Amendment of the Constitution, and Vice President Gore was able to secure the Democratic nomination with relative ease. Bush was seen as the early favorite for the Republican nomination and, despite a contentious primary battle with Senator John McCain and other candidates, secured the nomination by Super Tuesday.

Q1
Q2
Q3

Conversation Flow

Challenges

1. Generate **conversational interconnected** questions depending on the conversation so far
2. A coherent conversation must have **smooth transitions** between turns
   - We expect the narrative structure of passages can influence the conversation flow of our interconnected questions

Incumbent Democratic President Bill Clinton was ineligible to serve a third term due to term limitations in the 22nd Amendment of the Constitution, and Vice President Gore was able to secure the Democratic nomination with relative ease. Bush was seen as the early favorite for the Republican nomination and, despite a contentious primary battle with Senator John McCain and other candidates, secured the nomination by Super Tuesday.

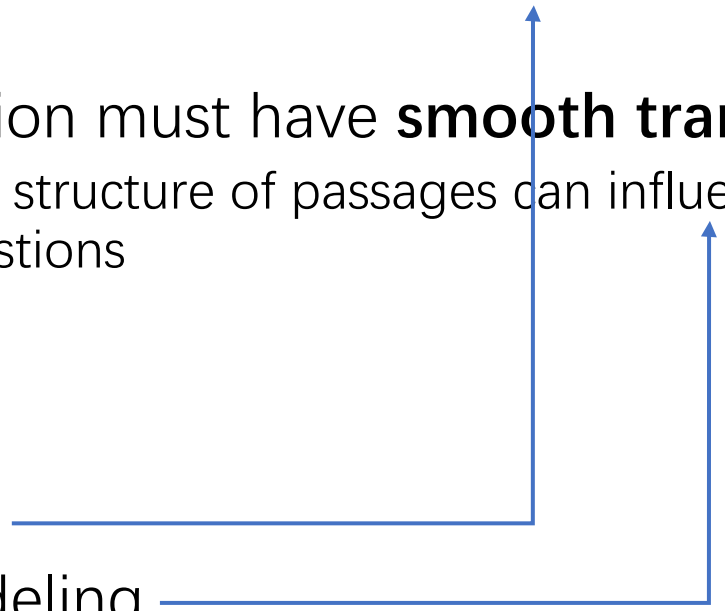
Q1
Q2
Q3
Q4
…
Qn

Conversation Flow

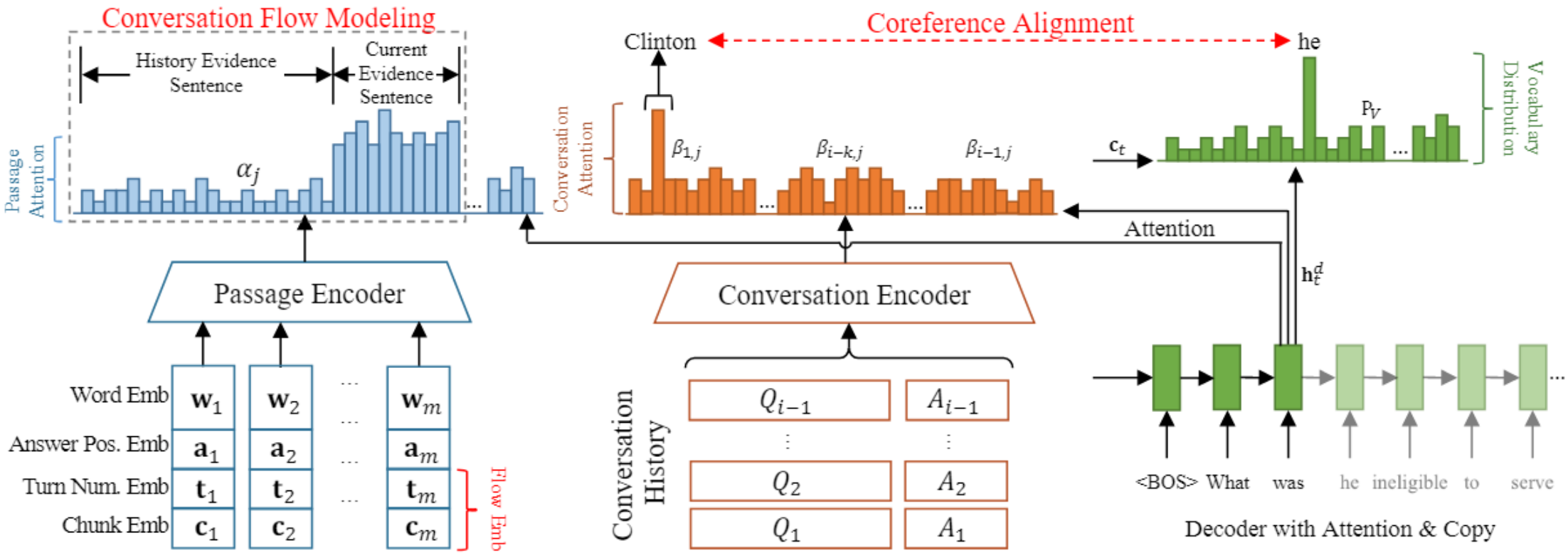
Conversational Question Generation

- Challenges
  1. Generate **conversational interconnected** questions depending on the conversation so far
  2. A coherent conversation must have **smooth transitions** between turns
     - We expect the narrative structure of passages can influence the conversation flow of our interconnected questions

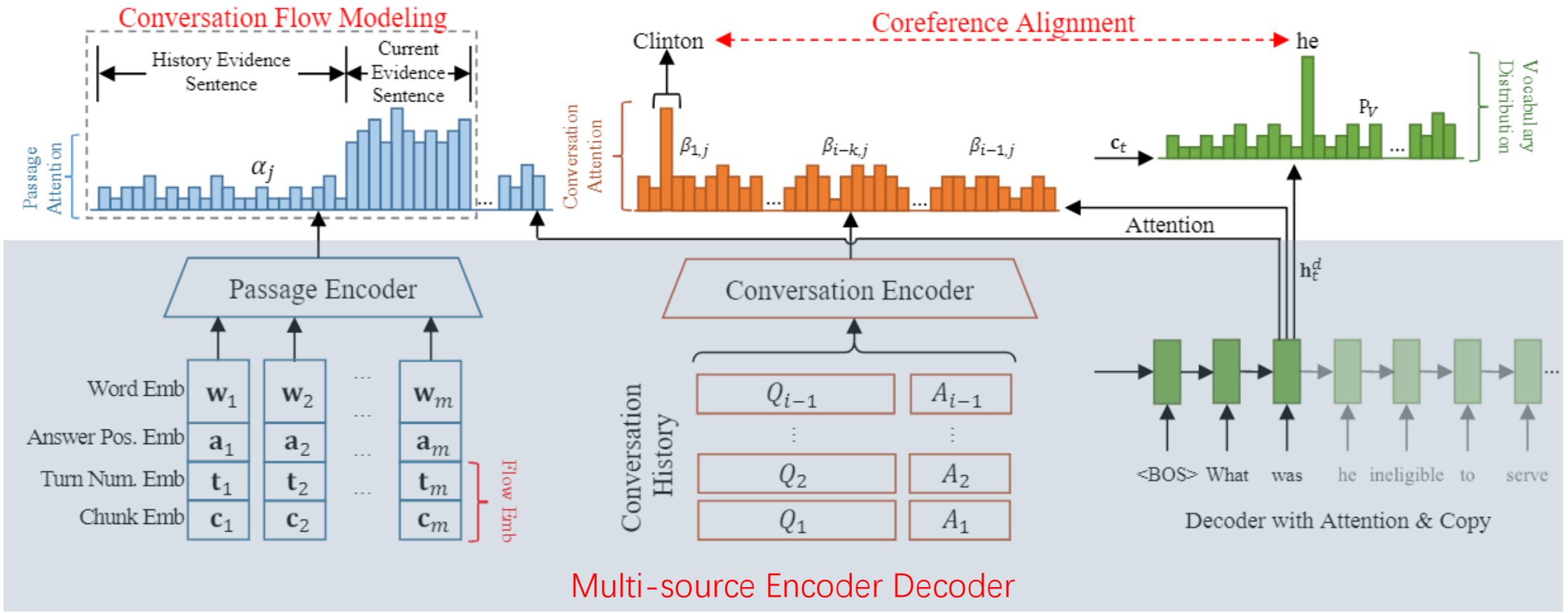
- Solutions
  - Coreference Alignment
  - Conversation Flow Modeling



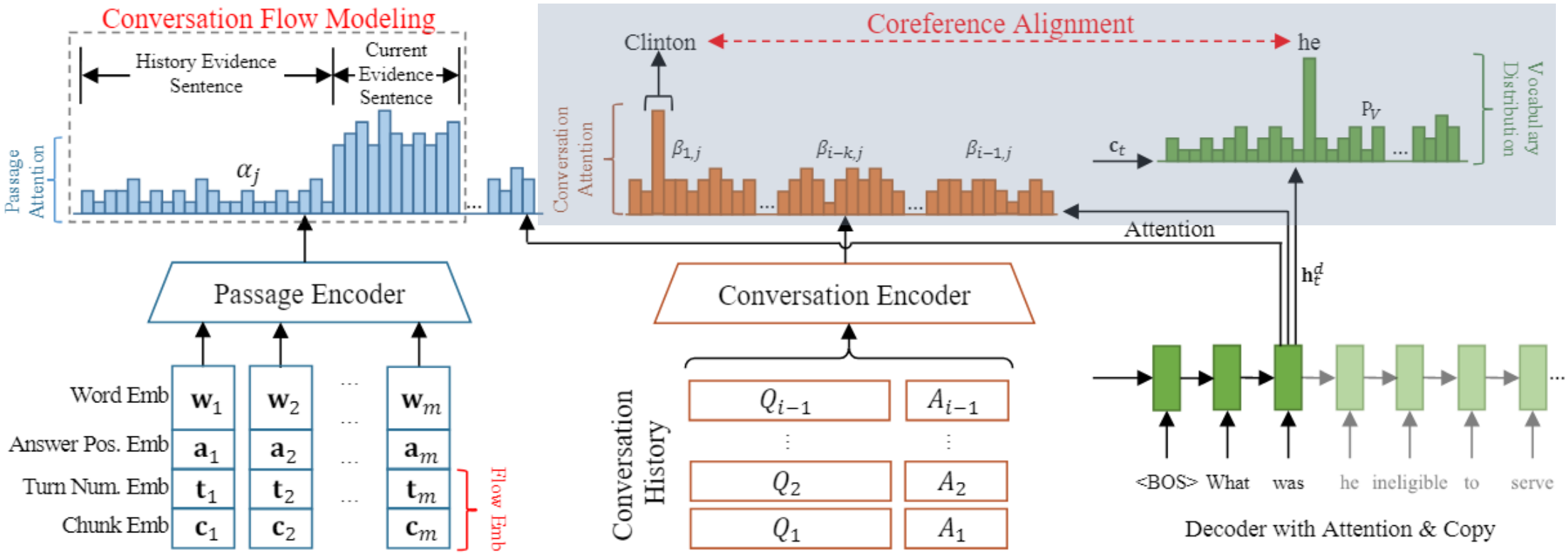
Proposed Framework



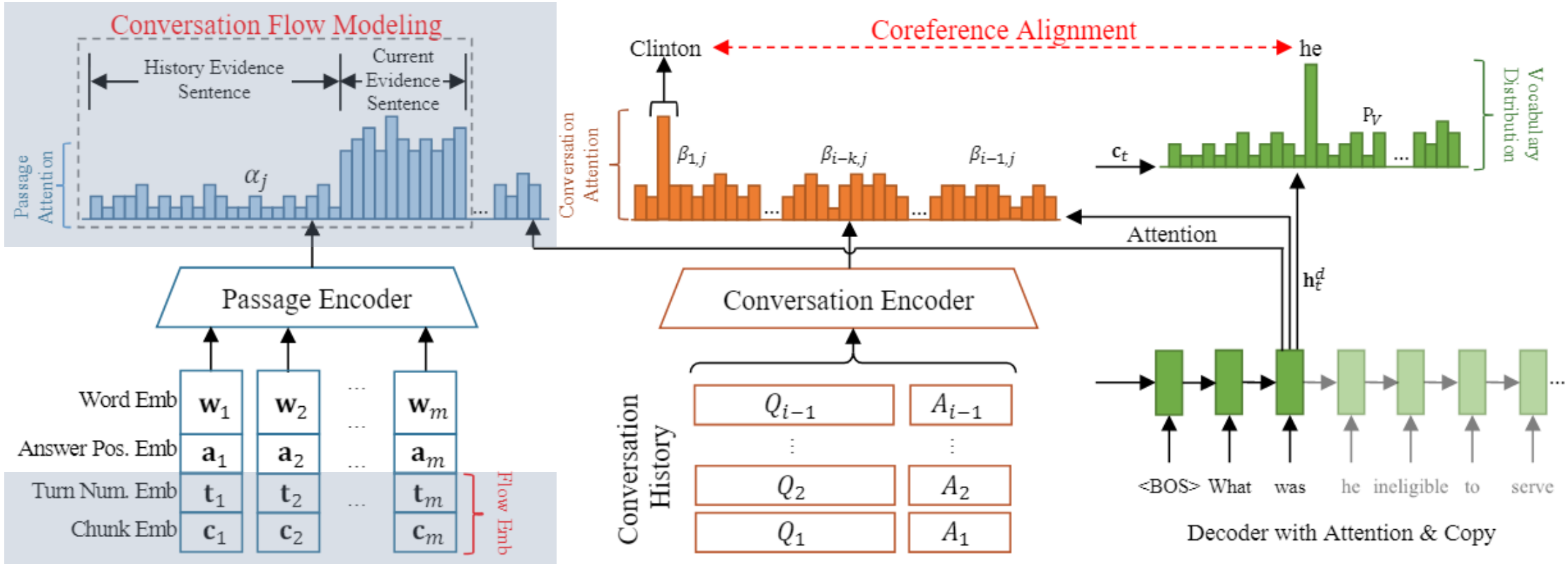
Proposed Framework



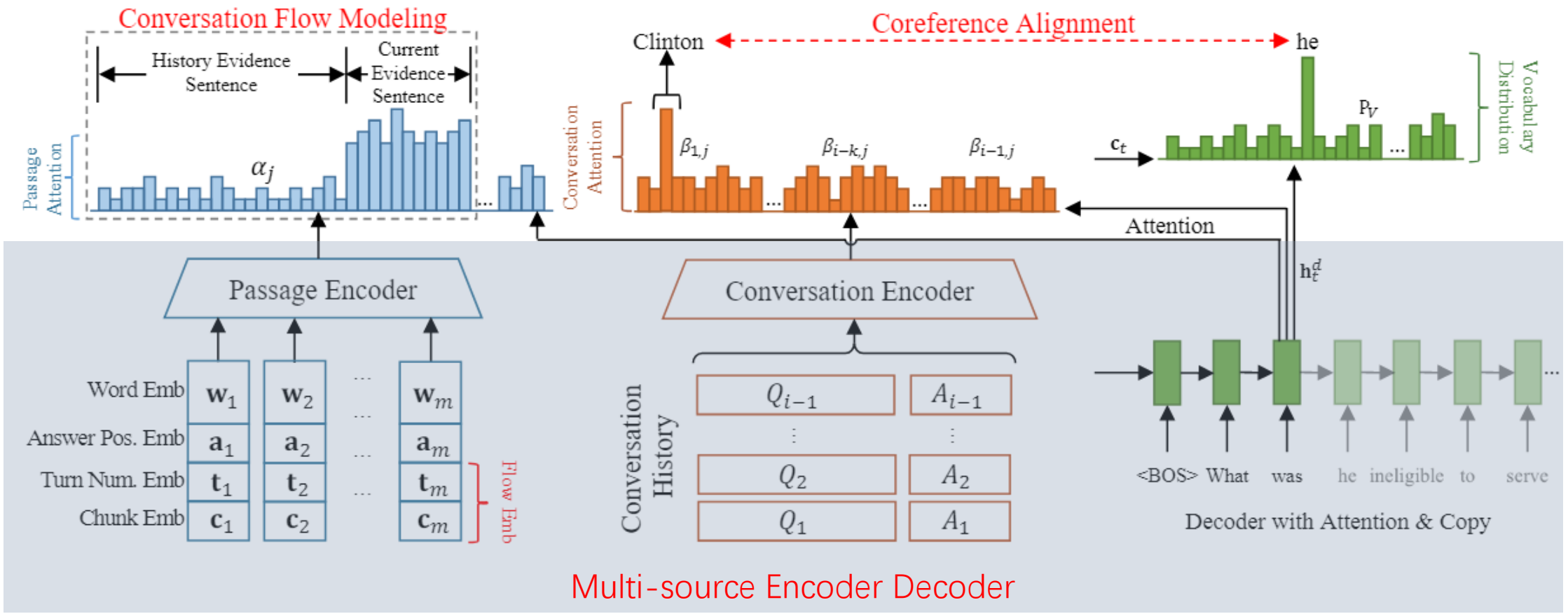
Proposed Framework



Proposed Framework

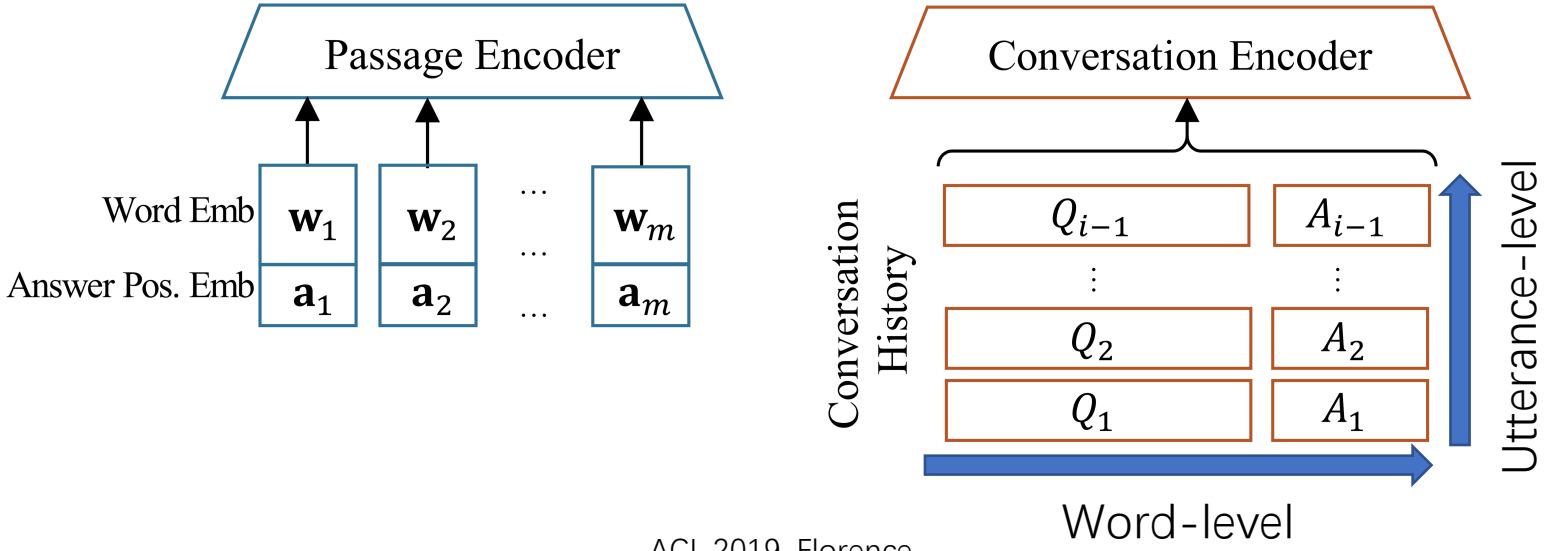


Multi-source Encoder Decoder

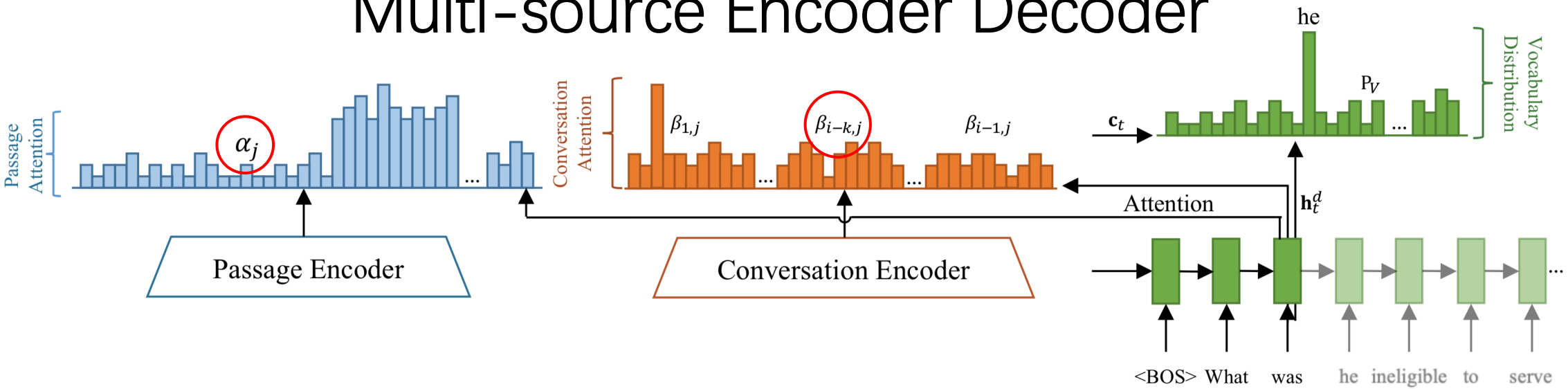


Multi-source Encoder Decoder

- Multi-source encoder jointly encodes the passage and conversation
  - Passage Representation: $(\mathbf{h}_1^p, \mathbf{h}_2^p, \ldots, \mathbf{h}_m^p)$
  - Conversation History Representation:
    - Word-level representation: $\left(\mathbf{h}_{i-k,1}^w, \ldots, \mathbf{h}_{i-k,m}^w\right)$, where $\mathrm{i} - k \in [1, i)$ is the turn number.
    - Utterance-level representation: $(\mathbf{h}_1^c, \ldots, \mathbf{h}_{i-1}^c)$



Multi-source Encoder Decoder

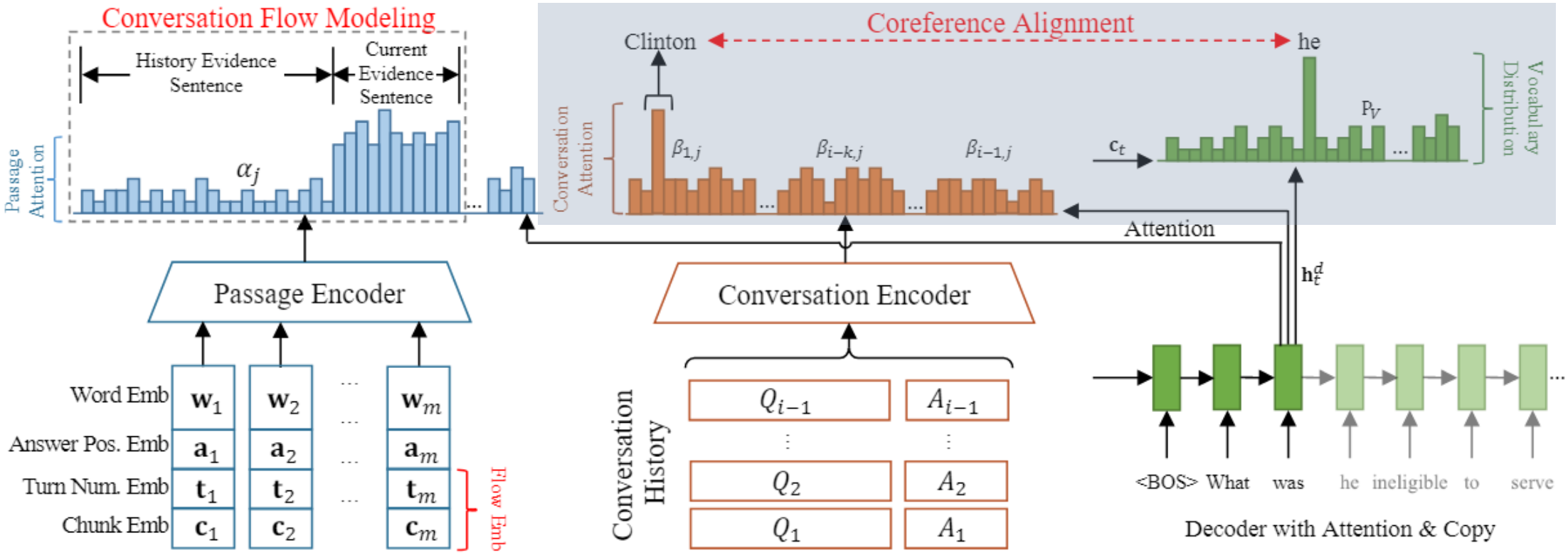


- The decoder itself decides to focus more on passage or conversation history:
  - Passage: $\mathbf{h}_j^p$, Conversation: $\mathbf{h}_{i-k,j}^w$, Decoder: $\mathbf{h}_t^d$
  - Use attention to calculate the importance score for each token in the passage and the conversation history as $\boldsymbol{\alpha_j}$ and $\boldsymbol{\beta_{i-k,j}}$ respectively;
  - Derive the context vector $\mathbf{c}_t$ and final vocabulary distribution P_V:

$$c_t = \Sigma_j \alpha_j \mathbf{h}_j^p + \Sigma_{k,j} \beta_{i-k,j} \mathbf{h}_{i-k,j}^w, \qquad P_V = \text{softmax}(\mathbf{W}_v \tanh(\mathbf{W}_a[\mathbf{h}_t^d; \mathbf{c}_t]))$$

Coreference Alignment

Generate conversational interconnected questions depending on the conversation so far



Coreference Alignment

Generate conversational interconnected questions depending on the conversation so far

- Explicitly **align** coreferent mentions in conversation history with corresponding pronominal references in generated questions
  - Preprocessing Stage

What political party is Clinton a member of?

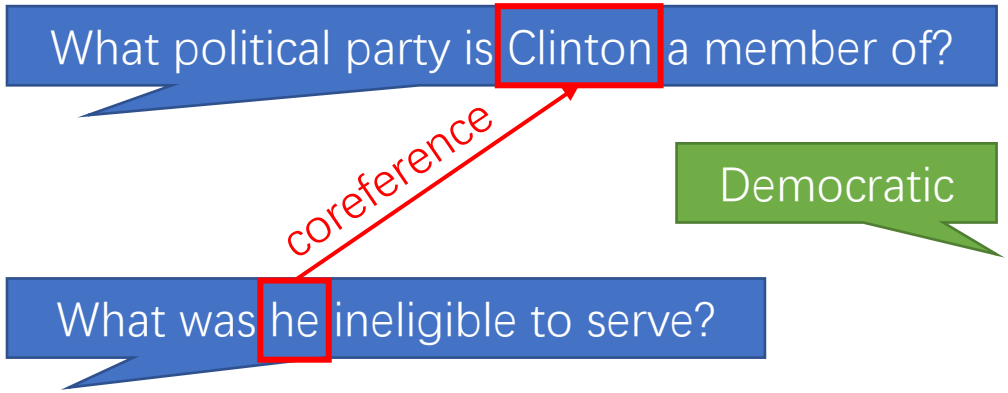
Democratic

What was he ineligible to serve?

Coreference Alignment

Generate conversational interconnected questions depending on the conversation so far

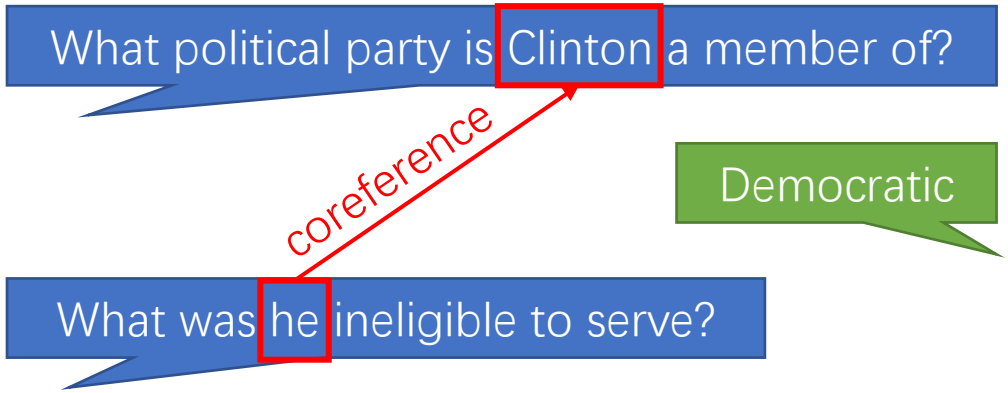
- Explicitly **align** coreferent mentions in conversation history with corresponding pronominal references in generated questions
  - Preprocessing Stage



Coreference Alignment

Generate conversational interconnected questions depending on the conversation so far

- Explicitly **align** coreferent mentions in conversation history with corresponding pronominal references in generated questions
  - Preprocessing Stage

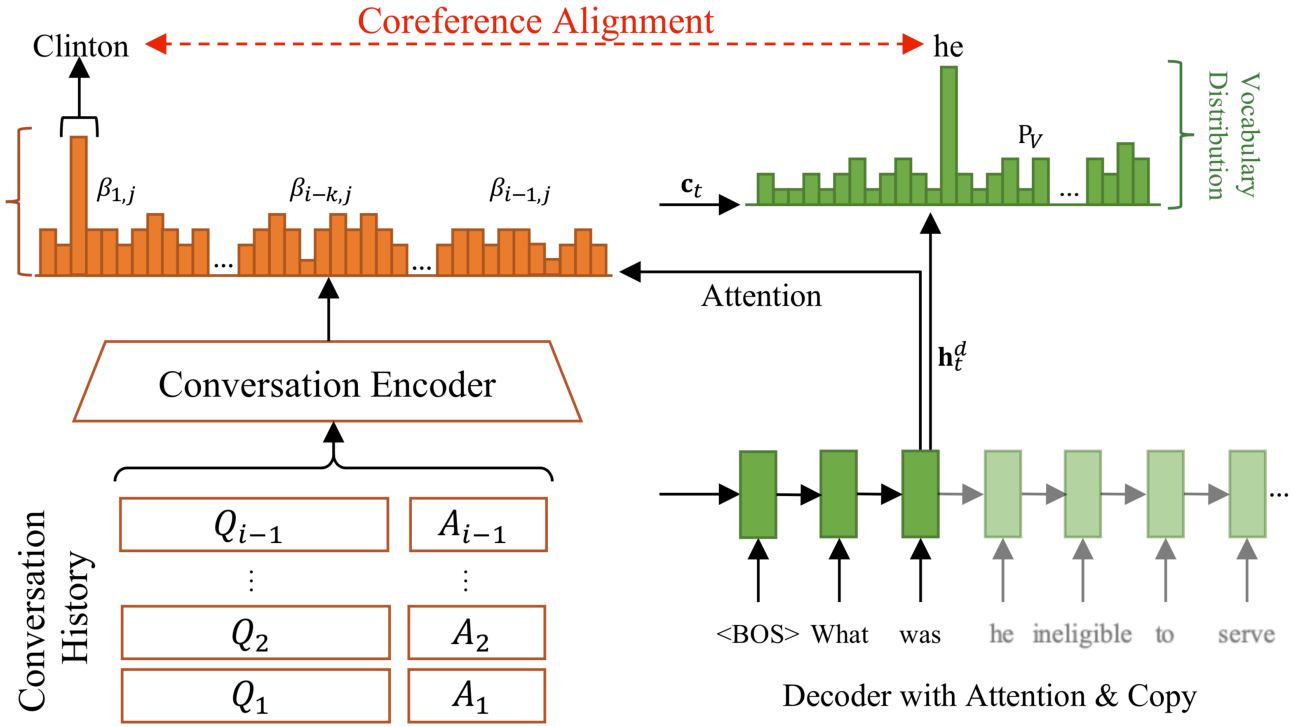


  - Training Stage

e.g., Clinton

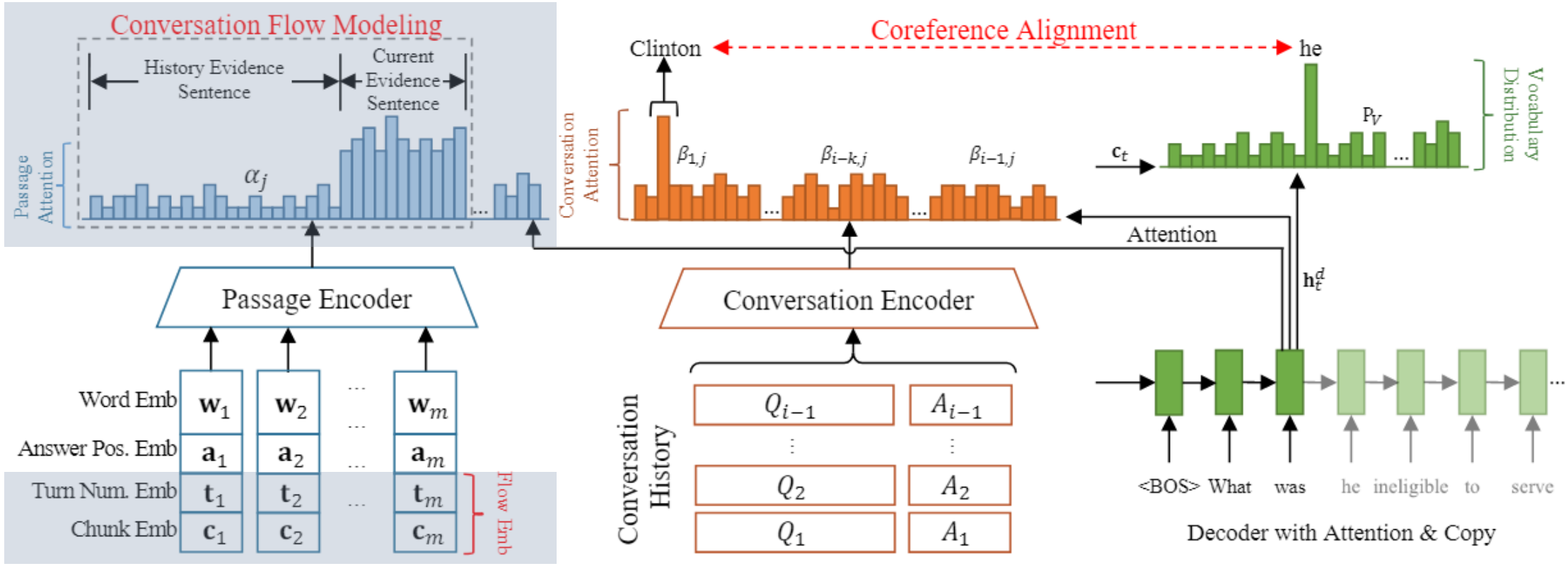
e.g., he

$$\mathcal{L}_{\text{coref}} = -(\lambda_1 \log \frac{\sum_j \beta_j^c}{\sum_{k,j} \beta_{i-k,j}} + \lambda_2 \log p_{\text{coref}}) * s_c$$



Conversation Flow Modeling

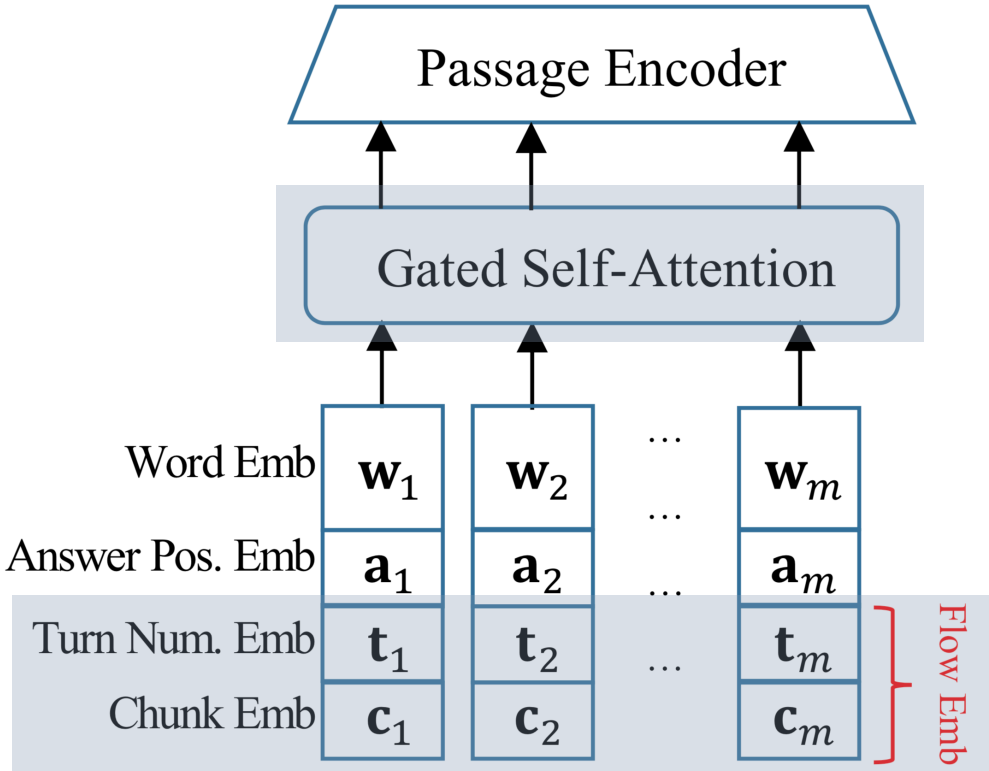
Model the conversation flow to transit focus inside the passage smoothly across turns



Conversation Flow Modeling: Flow Embedding

Convey the correlations between number of turns and narrative structure of passages

- Turn number embedding
  - Map the turn number into its feature embedding space
- Chunk embedding
  - Split the passage into L uniform chunks, and create an embedding vector for each chunk
- A gated self-attention mechanism over different embeddings
  - Learn the latent alignment between the turn number embedding and the chunk embedding



Conversation Flow Modeling: Flow Loss

1. Focus on sentences contain key information to generate the current turn question
2. Ignore sentences questioned several turns ago

Conversation Flow Modeling: Flow Loss

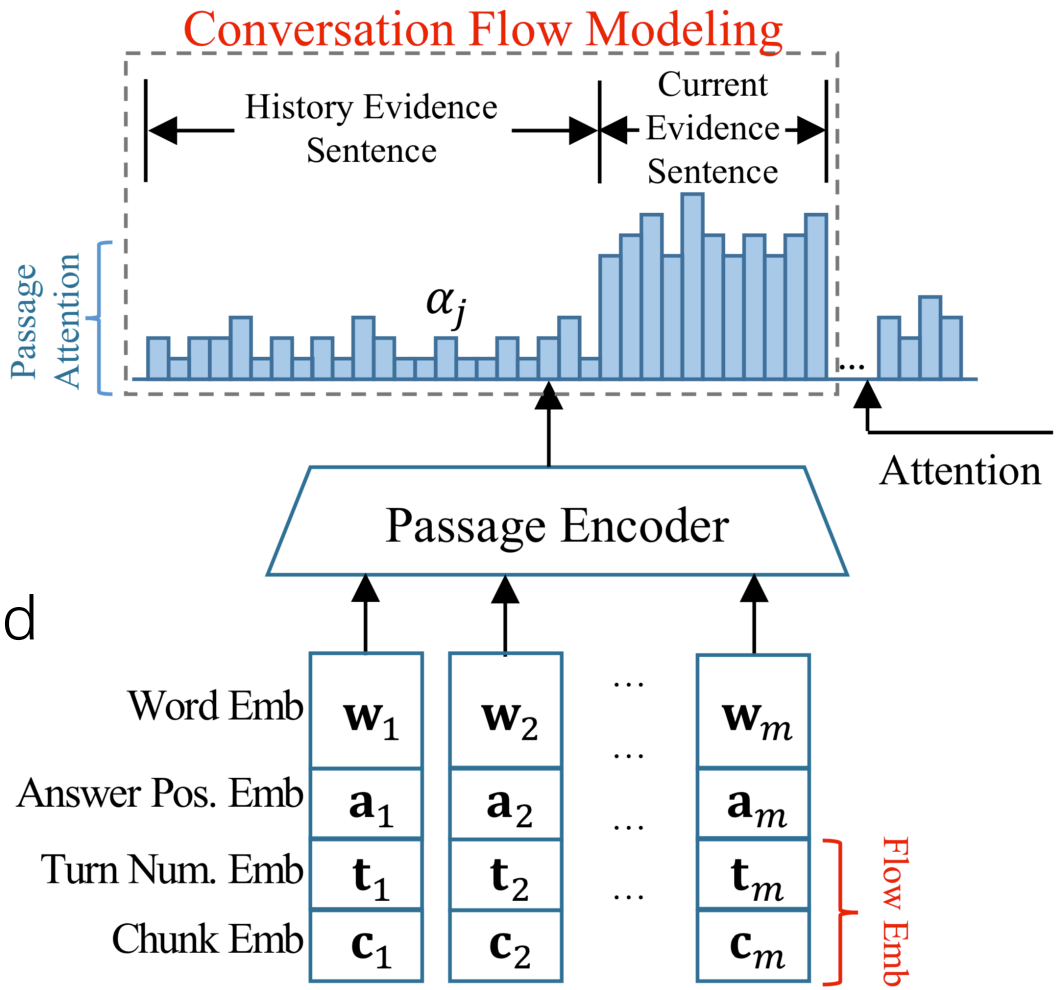
1. Focus on sentences contain key information to generate the current turn question Current Evidence Sentence
2. Ignore sentences questioned several turns ago History Evidence Sentence

Conversation Flow Modeling: Flow Loss

1. Focus on sentences contain key information to generate the current turn question — Current Evidence Sentence
2. Ignore sentences questioned several turns ago — History Evidence Sentence

- Focus on Current Evidence Sentence (CES): and ignore History Evidence Sentence (HES): via a flow loss:

$$\mathcal{L}_{\text{flow}} = -\lambda_3 \log \frac{\sum_{j:w_j \in \text{CES}} \alpha_j}{\sum_j \alpha_j} + \lambda_4 \frac{\sum_{j:w_j \in \text{HES}} \alpha_j}{\sum_j \alpha_j}$$



Dataset & Evaluation Metrics

- CoQA Dataset (Reddy et al., 2019)
  - A large-scale conversational question answering dataset
  - 8k conversations, 127k QA pairs
  - Short question length: 5.5 tokens (SQuAD: 10.1 tokens)
- Evaluation
  - Automatic Evaluation: BLEU, ROUGE
  - Human Evaluation

Main Results

- Baselines:
  - PGNet: Pointer-Generator Network
  - NQG: [Du and Cardie, 2018]
- Ablations:
  - MSNet: Multi-source EncDec
  - CorefNet: Coreference Alignment
  - FlowNet: Conversation Flow Modeling
  - CFNet: Our Full Model

	B1	B2	B3	R-L
PGNet	28.84*	13.74*	8.16*	39.18*
NQG	35.56*	21.14*	14.84*	45.58*
MSNet	36.27*	21.92*	15.51*	46.01*
CorefNet	36.89	22.28	15.77	46.53
FlowNet	36.87	22.49	15.98	46.64
CFNet	**37.38**	**22.81**	**16.25**	**46.90**

underline: p-value<0.05, *: p-value<0.01

Main Results

- Baselines:
  - PGNet: Pointer-Generator Network
  - NQG: [Du and Cardie, 2018]
- Ablations:
  - MSNet: Multi-source EncDec
  - CorefNet: Coreference Alignment
  - FlowNet: Conversation Flow Modeling
  - CFNet: Our Full Model

	B1	B2	B3	R-L
PGNet	28.84*	13.74*	8.16*	39.18*
NQG	35.56*	21.14*	14.84*	45.58*
MSNet	36.27*	21.92*	15.51*	46.01*
CorefNet	36.89	22.28	15.77	46.53
FlowNet	36.87	22.49	15.98	46.64
CFNet	**37.38**	**22.81**	**16.25**	**46.90**

underline: p-value<0.05, *: p-value<0.01

Main Results

- Baselines:
  - PGNet: Pointer-Generator Network
  - NQG: [Du and Cardie, 2018]
- Ablations:
  - MSNet: Multi-source EncDec
  - CorefNet: Coreference Alignment
  - FlowNet: Conversation Flow Modeling
  - CFNet: Our Full Model

	B1	B2	B3	R-L
PGNet	28.84*	13.74*	8.16*	39.18*
NQG	35.56*	21.14*	14.84*	45.58*
MSNet	36.27*	21.92*	15.51*	46.01*
CorefNet	<u>36.89</u>	<u>22.28</u>	<u>15.77</u>	<u>46.53</u>
FlowNet	36.87	22.49	15.98	46.64
CFNet	**37.38**	**22.81**	**16.25**	**46.90**

<u>underline</u>: p-value<0.05, *: p-value<0.01

Main Results

- Baselines:
  - PGNet: Pointer-Generator Network
  - NQG: [Du and Cardie, 2018]
- Ablations:
  - MSNet: Multi-source EncDec
  - CorefNet: Coreference Alignment
  - FlowNet: Conversation Flow Modeling
  - CFNet: Our Full Model

	B1	B2	B3	R-L
PGNet	28.84*	13.74*	8.16*	39.18*
NQG	35.56*	21.14*	14.84*	45.58*
MSNet	36.27*	21.92*	15.51*	46.01*
CorefNet	36.89	22.28	15.77	46.53
FlowNet	36.87	22.49	15.98	46.64
CFNet	**37.38**	**22.81**	**16.25**	**46.90**

underline: p-value<0.05, *: p-value<0.01

Coreference Alignment Analysis

- Coreference Set
  - Each sample in the coreference set requires a pronoun resolution

Precision, Recall, F-score of pronouns in generated questions

	B1	B2	B3	R-L	P	R	F
PGNet	27.66*	13.82*	8.96*	38.40*	26.87*	25.17*	25.68*
NQG	34.75*	21.52*	15.96*	45.04*	34.46*	32.97*	33.25*
MSNet	36.31*	22.92	17.07	45.97*	35.34*	33.80*	34.07*
CorefNet	**37.51**	**24.14**	**18.44**	**47.45**	**42.09**	**40.35**	**40.64**

A large margin!

underline: p-value<0.05, *: p-value<0.01

Coreference Alignment Analysis

The attention probability when the model predicts a pronoun

Passage: … however , mccain has a very different life story . he grew up in a navy family and was a pilot during the vietnam war in the 1960s …

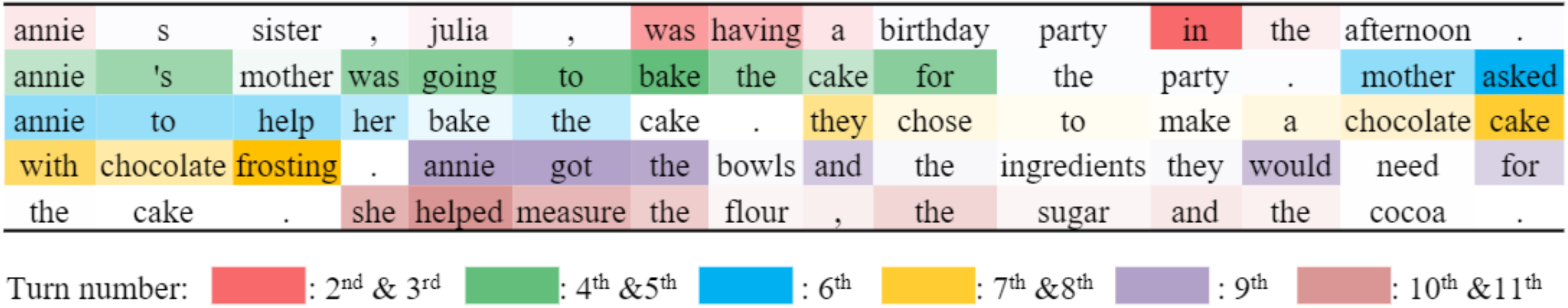
Conversation History:

<q>	what	war	was	mccain	in	?
0.0000	0.0001	0.0049	0.0138	0.7710	0.0055	0.0069
<a>	vietnam	war				
0.0000	0.0140	0.0095				
<q>	was	he	in	the	army	?
0.0000	0.0045	0.1303	0.0005	0.0139	0.0001	0.0250
<a>	no					
0.0000	0.0000					

Question (Human): what was his job ?

Question (Our Model): what was his job ?

Conversation Flow Analysis



The transition of passage attention distribution across turns

Human Evaluation

- We hire 5 annotators to rate 93 questions
- Rating criteria (1-3 scale, 3 for the best):
  - Grammaticality: the grammatical correctness and fluency
  - Answerability: whether the generated question can be answered by the current answer
  - **Interconnectedness**: whether the generated questions are *conversational* or not

	Grammaticality	Answerability	Interconnectedness
PGNet	2.74	1.39	1.59
MSNet	2.85	2.39	1.74
CFNet	2.89	**2.74***	**2.67***

*: p-value<0.01

Conclusion

- A new setting: Conversational Question Generation
- Coreference Alignment
- Conversation Flow Modeling

- Limitations and Future Work
  - Incorporate answer span identification into the current system
  - The answerer may also want to ask clarification questions
  - Domain adaptation (7 domains in CoQA dataset)
  - Gender Bias (his/he appears more frequently than her/she)

Reference

1. Sudha Rao and Hal Daumé III. Learning to Ask Good Questions: Ranking Clarification Questions using Neural Expected Value of Perfect Information. In ACL 2018.
2. Yansen Wang, Chenyi Liu, Minlie Huang, Liqiang Nie. Learning to Ask Questions in Open-domain Conversational Systems with Typed Decoders. In ACL 2018.
3. Xinya Du and Claire Cardie. Harvesting Paragraph-Level Question-Answer Pairs from Wikipedia. In ACL 2018.
4. Xinya Du, Junru Shao and Claire Cardie. Learning to Ask: Neural Question Generation for Reading Comprehension. In ACL 2017.
5. Zhilin Yang, Junjie Hu, Ruslan Salakhutdinov, William W. Cohen. Semi-Supervised QA with Generative Domain-Adaptive Nets. In ACL 2017
6. Siva Reddy, Danqi Chen, and Christopher D Manning. CoQA: A conversational question answering challenge. In TACL 2018.

Thanks

Code:

scan this

or https://github.com/Evan-Gao/conversational-QG