Inducing Constraints in Paraphrase Generation and Consistency in Paraphrase Detection

Ashutosh Kumar
Indian Institute of Science, Bangalore

Definition - Paraphrases

Paraphrases are sentences or phrases that convey the same meaning using different words. Although the logical definition of paraphrases requires strict semantic equivalence, ***linguistics accepts a broader, approximate equivalence—thereby allowing far more examples of "quasi-paraphrases."****

***For a list of possible definitions of paraphrases, please refer to [1]**

Definition - Paraphrases

Paraphrases are sentences or phrases that convey the same meaning using different words. Although the logical definition of paraphrases requires strict semantic equivalence, ***linguistics accepts a broader, approximate equivalence—thereby allowing far more examples of "quasi-paraphrases."****

The finalists will play in Giants stadium. ⇔ Giants stadium will be the playground for the finalists.

***For a list of possible definitions of paraphrases, please refer to [1]**

Definition - Paraphrases

Paraphrases are sentences or phrases that convey the same meaning using different words. Although the logical definition of paraphrases requires strict semantic equivalence, ***linguistics accepts a broader, approximate equivalence—thereby allowing far more examples of "quasi-paraphrases."****

The finalists will play in Giants stadium. ⇔ Giants stadium will be the playground for the finalists.
The pilot took off despite the stormy weather. ⇔ The plane took off despite the stormy weather.

***For a list of possible definitions of paraphrases, please refer to [1]**

Definition - Paraphrases

Paraphrases are sentences or phrases that convey the same meaning using different words. Although the logical definition of paraphrases requires strict semantic equivalence, ***linguistics accepts a broader, approximate equivalence—thereby allowing far more examples of "quasi-paraphrases."****

The finalists will play in Giants stadium. ⇔ Giants stadium will be the playground for the finalists.
The pilot took off despite the stormy weather. ⇔ The plane took off despite the stormy weather.
The pilot took off despite the stormy weather. ⇔ The pilot is checking for stormy weather.

***For a list of possible definitions of paraphrases, please refer to [1]**

Subproblems of Paraphrasing

- Paraphrase **Generation**: Reword a sentence into a text such that the text preserves most of the semantics of the sentence.

- Paraphrase **Detection**: Given two sentences, predict if they are paraphrases of each other.

Scope of the Thesis

Problems	Examples
Diversity in paraphrase generation	**Input (X):** - how do i increase body height ? **Output (Y):** - how do i increase my height ? - how do i increase my body height ? - how do i increase the height ? - how would i increase my body height ?
Syntacticality in paraphrase generation	**Input** **(X):** What are pure substances ? What are some examples ? **Exemplar sentence** **(Z)**: What are the characteristics of the Elizabeth theatre ? **Output** **(Y):** Which is the sample of a substance ?
Consistency in paraphrase detection	**Input** **X:** a provision government or a revolutionary government has been declared several times by insurgent groups in philippines . **Y:** a provision government or a revolutionary government has been declared several times in philippines by insurgent groups . **Output** **For (X, Y) as input:** 1 (98.6) **For (Y, X) as input:** 0 (92.2)

Scope of the Thesis

Problems	Examples
Diversity in paraphrase generation	**Input (X):** - how do i increase body height ? **Output (Y):** - how do i increase my height ? - how do i increase my body height ? - how do i increase the height ? - how would i increase my body height ?
Syntacticality in paraphrase generation	**Input** **(X):** What are pure substances ? What are some examples ? **Exemplar sentence** **(Z)**: What are the characteristics of the Elizabeth theatre ? **Output** **(Y):** Which is the sample of a substance ?
Consistency in paraphrase detection	**Input** **X:** a provision government or a revolutionary government has been declared several times by insurgent groups in philippines . **Y:** a provision government or a revolutionary government has been declared several times in philippines by insurgent groups . **Output** **For (X, Y) as input:** 1 (98.6) **For (Y, X) as input:** 0 (92.2)

Scope of the Thesis

Problems	Examples
Diversity in paraphrase generation	**Input (X):** - how do i increase body height ? **Output (Y):** - how do i increase my height ? - how do i increase my body height ? - how do i increase the height ? - how would i increase my body height ?
Syntacticality in paraphrase generation	**Input** **(X):** What are pure substances ? What are some examples ? **Exemplar sentence** **(Z)**: What are the characteristics of the Elizabeth theatre ? **Output** **(Y):** Which is the sample of a substance ?
Consistency in paraphrase detection	**Input** **X:** a provision government or a revolutionary government has been declared several times by insurgent groups in philippines . **Y:** a provision government or a revolutionary government has been declared several times in philippines by insurgent groups . **Output** **For (X, Y) as input:** 1 (98.6) **For (Y, X) as input:** 0 (92.2)

Scope of the Thesis

Problems	Examples
Diversity in paraphrase generation	**Input (X):** - how do i increase body height ? **Output (Y):** - how do i increase my height ? - how do i increase my body height ? - how do i increase the height ? - how would i increase my body height ?
Syntacticality in paraphrase generation	**Input** **(X):** What are pure substances ? What are some examples ? **Exemplar sentence** **(Z)**: What are the characteristics of the Elizabeth theatre ? **Output** **(Y):** Which is the sample of a substance ?
Consistency in paraphrase detection	**Input** **X:** a provision government or a revolutionary government has been declared several times by insurgent groups in philippines . **Y:** a provision government or a revolutionary government has been declared several times in philippines by insurgent groups . **Output** **For (X, Y) as input:** 1 (98.6) **For (Y, X) as input:** 0 (92.2)

Scope of the Thesis

Problems	Examples
Diversity in paraphrase generation	**Input (X):** - how do i increase body height ? **Output (Y):** - how do i increase my height ? - how do i increase my body height ? - how do i increase the height ? - how would i increase my body height ?
Syntacticality in paraphrase generation	**Input** **(X):** What are pure substances ? What are some examples ? **Exemplar sentence** **(Z)**: What are the characteristics of the Elizabeth theatre ? **Output** **(Y):** Which is the sample of a substance ?
Consistency in paraphrase detection	**Input** **X:** a provision government or a revolutionary government has been declared several times by insurgent groups in philippines . **Y:** a provision government or a revolutionary government has been declared several times in philippines by insurgent groups . **Output** **For (X, Y) as input:** 1 (98.6) **For (Y, X) as input:** 0 (92.2)

Scope of the Thesis

Problems	Examples
Diversity in paraphrase generation	**Input (X):** - how do i increase body height ? **Output (Y):** - how do i increase my height ? - how do i increase my body height ? - how do i increase the height ? - how would i increase my body height ?
Syntacticality in paraphrase generation	**Input** **(X):** What are pure substances ? What are some examples ? **Exemplar sentence** **(Z)**: What are the characteristics of the Elizabeth theatre ? **Output** **(Y):** Which is the sample of a substance ?
Consistency in paraphrase detection	**Input** **X:** a provision government or a revolutionary government has been declared several times by insurgent groups in philippines . **Y:** a provision government or a revolutionary government has been declared several times in philippines by insurgent groups . **Output** **For (X, Y) as input:** 1 (98.6) **For (Y, X) as input:** 0 (92.2)

Scope of the Thesis

Problems	Examples
Diversity in paraphrase generation	**Input (X):** - how do i increase body height ? **Output (Y):** - how do i increase my height ? - how do i increase my body height ? - how do i increase the height ? - how would i increase my body height ?
Syntacticality in paraphrase generation	**Input** **(X):** What are pure substances ? What are some examples ? **Exemplar sentence** **(Z)**: What are the characteristics of the Elizabeth theatre ? **Output** **(Y):** Which is the sample of a substance ?
Consistency in paraphrase detection	**Input** **X:** a provision government or a revolutionary government has been declared several times by insurgent groups in philippines . **Y:** a provision government or a revolutionary government has been declared several times in philippines by insurgent groups . **Output** **For (X, Y) as input:** 1 (98.6) **For (Y, X) as input:** 0 (92.2)

Scope of the Thesis

Problems	Examples
Diversity in paraphrase generation	**Input (X):** - how do i increase body height ? **Output (Y):** - how do i increase my height ? - how do i increase my body height ? - how do i increase the height ? - how would i increase my body height ?
Syntacticality in paraphrase generation	**Input** **(X):** What are pure substances ? What are some examples ? **Exemplar sentence** **(Z)**: What are the characteristics of the Elizabeth theatre ? **Output** **(Y):** Which is the sample of a substance ?
Consistency in paraphrase detection	**Input** **X:** a provision government or a revolutionary government has been declared several times by insurgent groups in philippines . **Y:** a provision government or a revolutionary government has been declared several times in philippines by insurgent groups . **Output** **For (X, Y) as input:** 1 (98.6) **For (Y, X) as input:** 0 (92.2)

Scope of the Thesis

Problems	Examples
Diversity in paraphrase generation	**Input (X):** - how do i increase body height ? **Output (Y):** - how do i increase my height ? - how do i increase my body height ? - how do i increase the height ? - how would i increase my body height ?
Syntacticality in paraphrase generation	**Input** **(X):** What are pure substances ? What are some examples ? **Exemplar sentence** **(Z)**: What are the characteristics of the Elizabeth theatre ? **Output** **(Y):** Which is the sample of a substance ?
Consistency in paraphrase detection	**Input** **X:** a provision government or a revolutionary government has been declared several times by insurgent groups in philippines . **Y:** a provision government or a revolutionary government has been declared several times in philippines by insurgent groups . **Output** **For (X, Y) as input:** 1 (98.6) **For (Y, X) as input:** 0 (92.2)

Scope of the Thesis

Problems	Examples
Diversity in paraphrase generation	**Input (X):** - how do i increase body height ? **Output (Y):** - how do i increase my height ? - how do i increase my body height ? - how do i increase the height ? - how would i increase my body height ?
Syntacticality in paraphrase generation	**Input** **(X):** What are pure substances ? What are some examples ? **Exemplar sentence** **(Z)**: What are the characteristics of the Elizabeth theatre ? **Output** **(Y):** Which is the sample of a substance ?
Consistency in paraphrase detection	**Input** **X:** a provision government or a revolutionary government has been declared several times by insurgent groups in philippines . **Y:** a provision government or a revolutionary government has been declared several times in philippines by insurgent groups . **Output** **For (X, Y) as input:** 1 (98.6) **For (Y, X) as input:** 0 (92.2)

Scope of the Thesis

Problems	Examples
Diversity in paraphrase generation	**Input (X):** - how do i increase body height ? **Output (Y):** - how do i increase my height ? - how do i increase my body height ? - how do i increase the height ? - how would i increase my body height ?
Syntacticality in paraphrase generation	**Input** **(X):** What are pure substances ? What are some examples ? **Exemplar sentence** **(Z)**: What are the characteristics of the Elizabeth theatre ? **Output** **(Y):** Which is the sample of a substance ?
Consistency in paraphrase detection	**Input** **X:** a provision government or a revolutionary government has been declared several times by insurgent groups in philippines . **Y:** a provision government or a revolutionary government has been declared several times in philippines by insurgent groups . **Output** **For (X, Y) as input:** 1 (98.6) **For (Y, X) as input:** 0 (92.2)

DISCLAIMER All experiments conducted on English language datasets

Outline

Outline

- **[Part 1] Inducing Constraints In Paraphrase Generation**
  - Motivation for inducing diversity in paraphrase generation
  - Baselines - Strengths and limitations
  - Introducing DiPS (our method) and providing component details **[Conf.: NAACL 2019]**
  - Introducing SGCP (our method) and providing component details **[Journal:TACL 2020]**
  - Results on standard datasets

Outline

- **[Part 1] Inducing Constraints In Paraphrase Generation**
  - Motivation for inducing diversity in paraphrase generation
  - Baselines - Strengths and limitations
  - Introducing DiPS (our method) and providing component details **[Conf.: NAACL 2019]**
  - Introducing SGCP (our method) and providing component details **[Journal:TACL 2020]**
  - Results on standard datasets

- **[Part 2] Inducing Consistency In Paraphrase Detection**
  - Motivation for inducing consistency in paraphrase detection
  - Defining the consistency objective **[Conf.: Findings ACL 2022]**
  - Baselines - Limitations
  - Results on standard datasets

Part 1: Constraints in Paraphrase Generation

Part 1: Constraints in Paraphrase Generation

Addresses the gap in diversity and syntacticality in paraphrase generation research

Part 1: Constraints in Paraphrase Generation

Addresses the gap in diversity and syntacticality in paraphrase generation research

NAACL 2019, TACL 2020

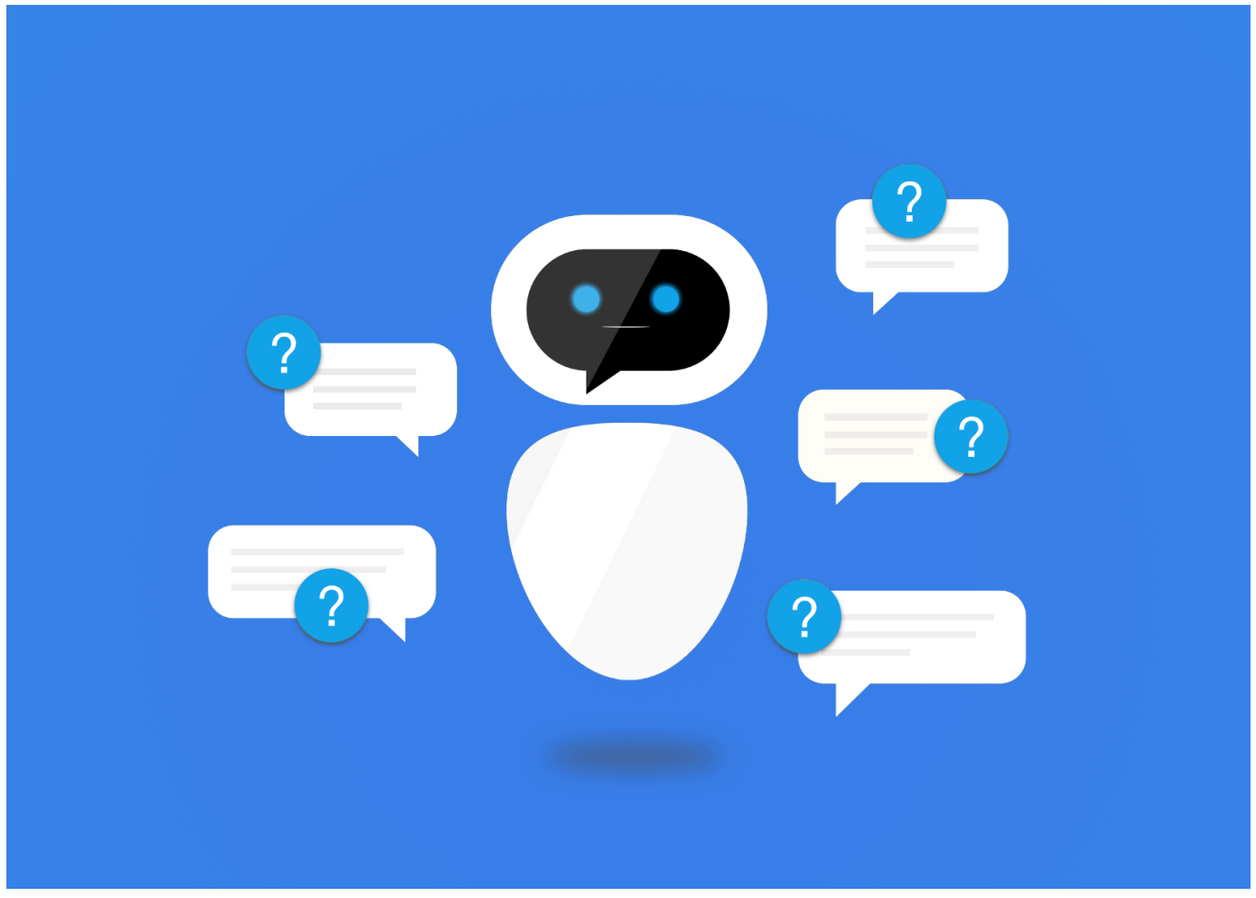
NAACL 2019: Submodular optimization-based diverse paraphrasing and its effectiveness in data augmentation.
Kumar et. al. 2019

TACL 2020: Syntax-Guided Controlled Generation of Paraphrases.
Kumar et. al. 2020

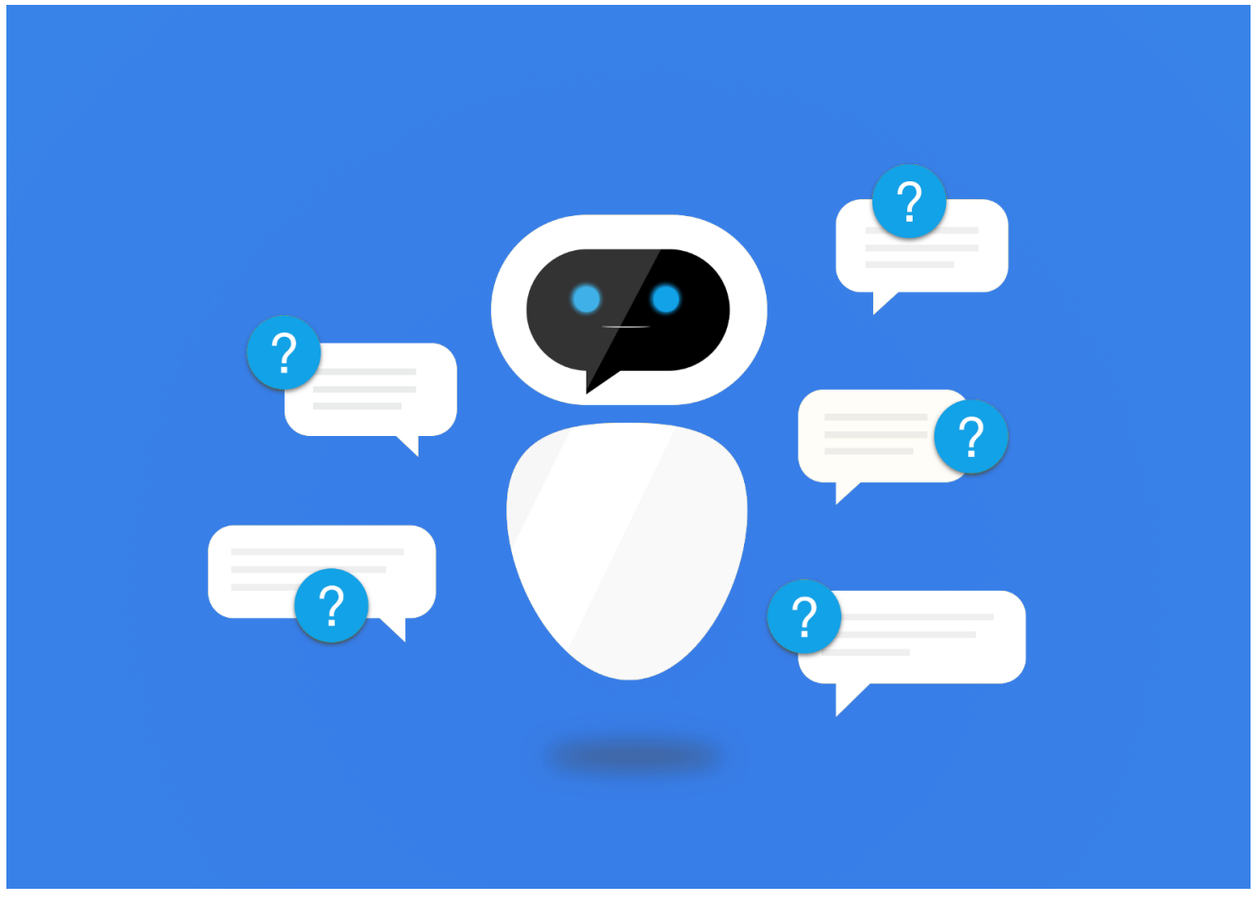
Diversity In Paraphrase Generation

Conversational Agents

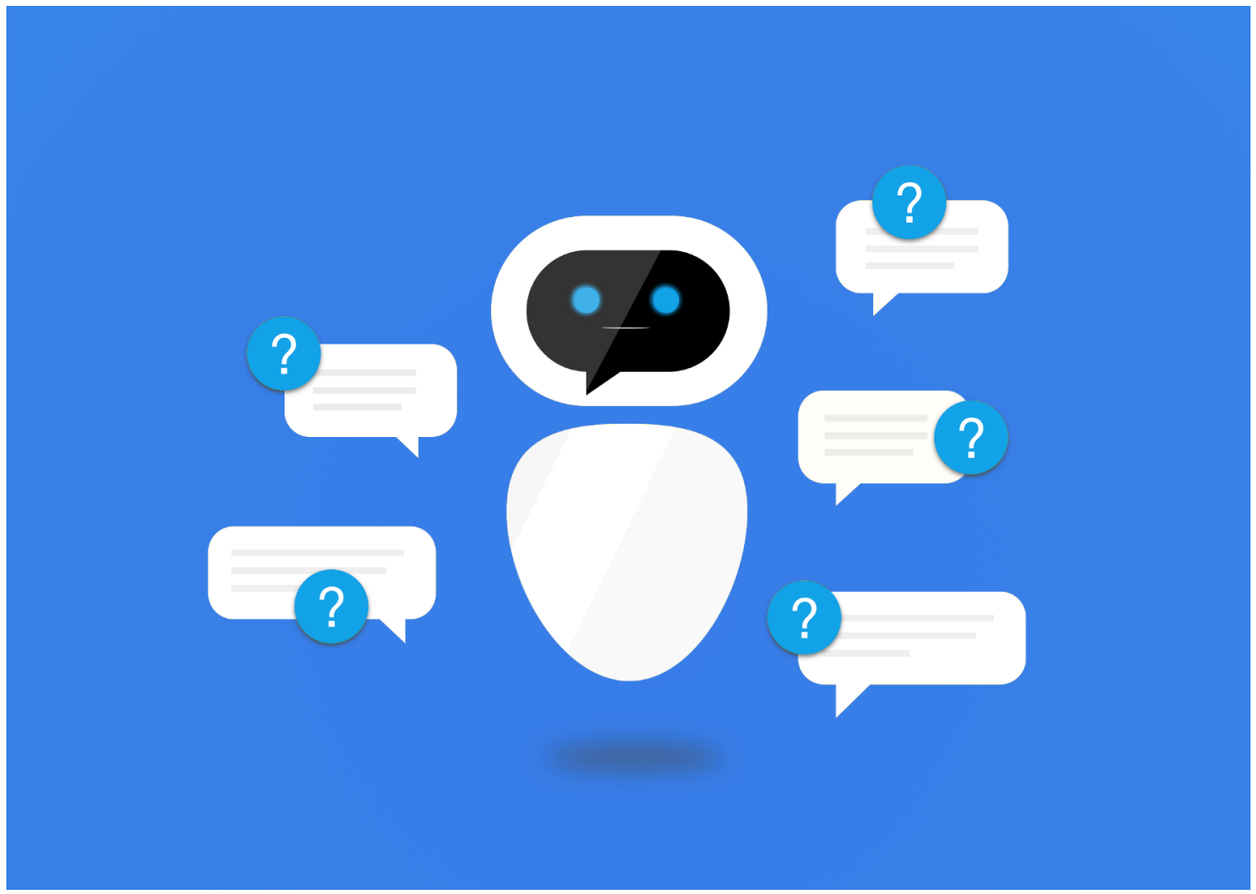
Conversational Agents



Conversational Agents

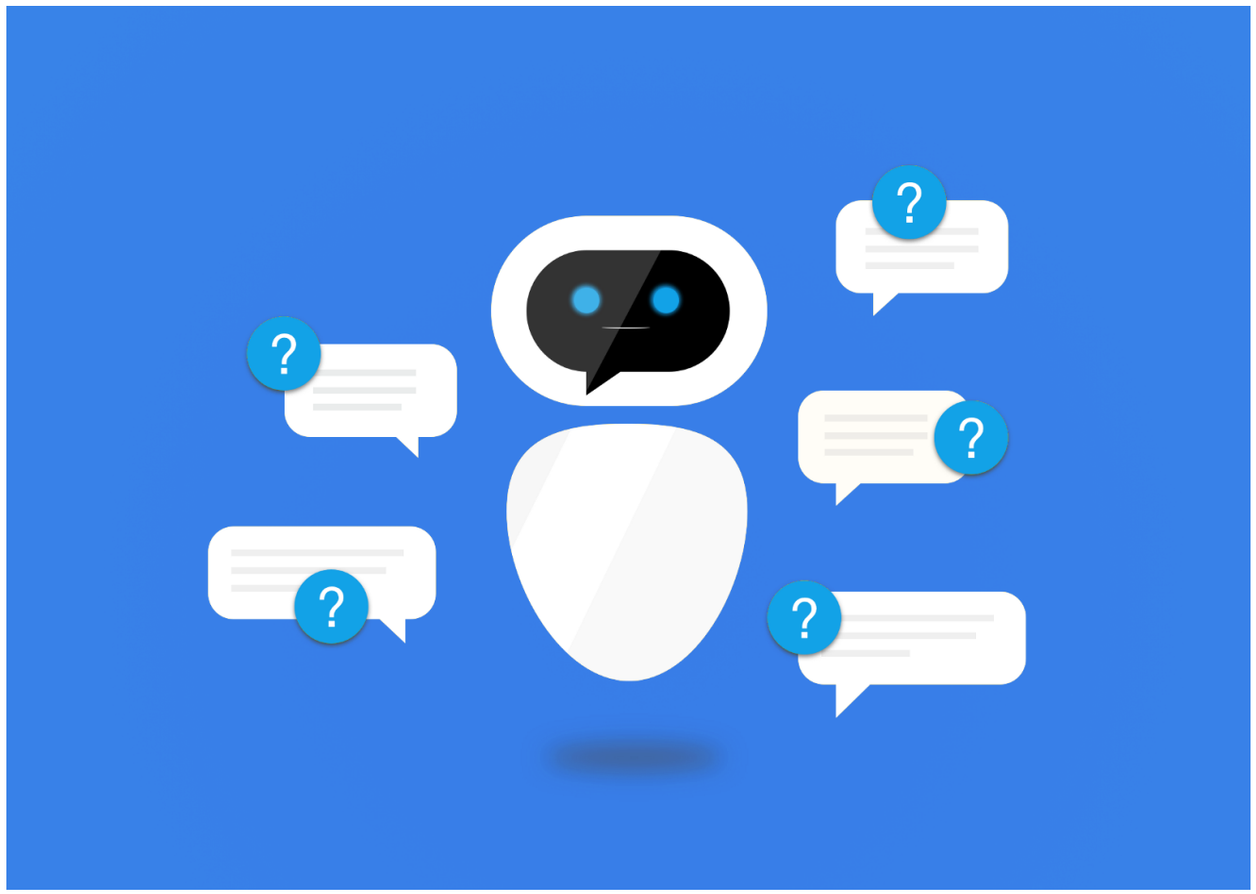


Conversational Agents



User: I want to book a flight from Minneapolis to New York

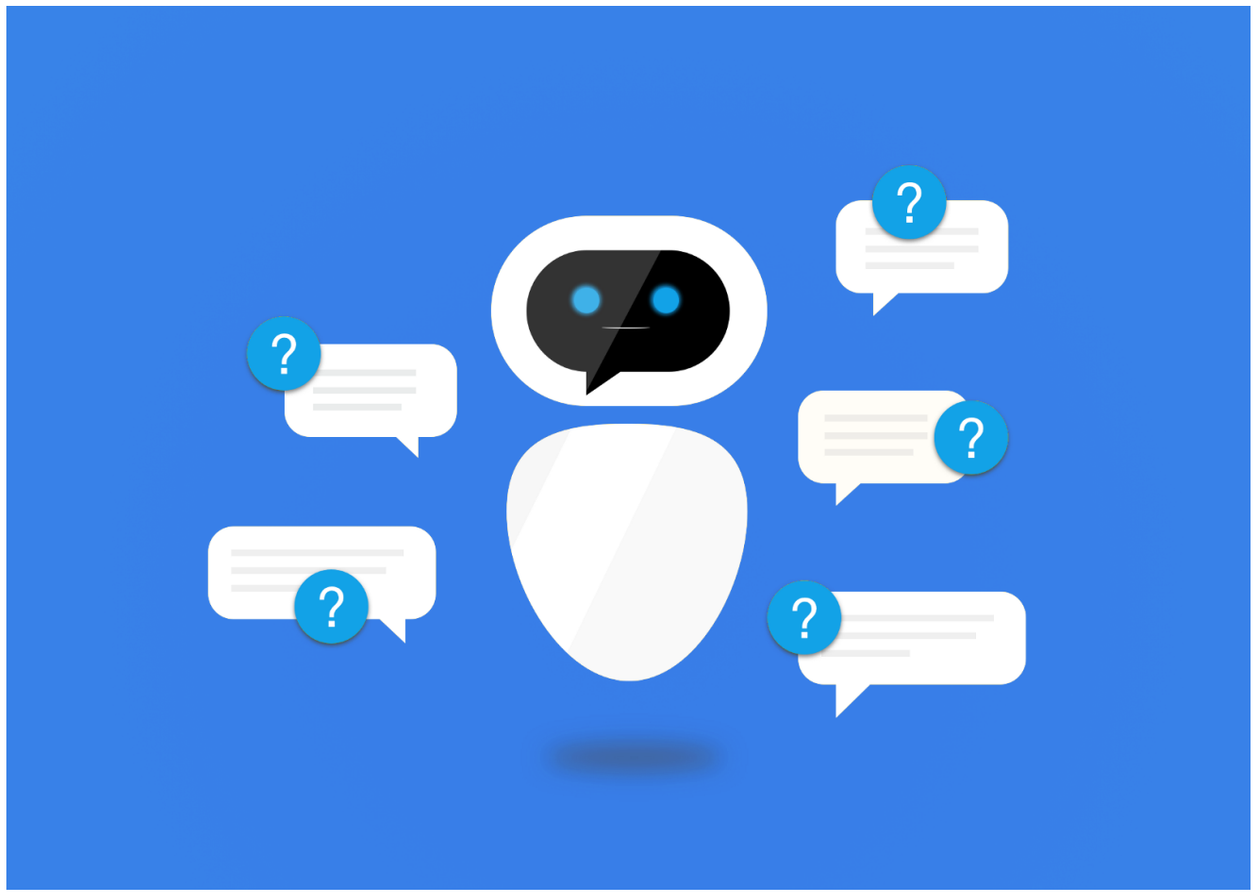
Conversational Agents



User: I want to book a flight from Minneapolis to New York

Bot: Sure. When are you planning to travel ?

Conversational Agents



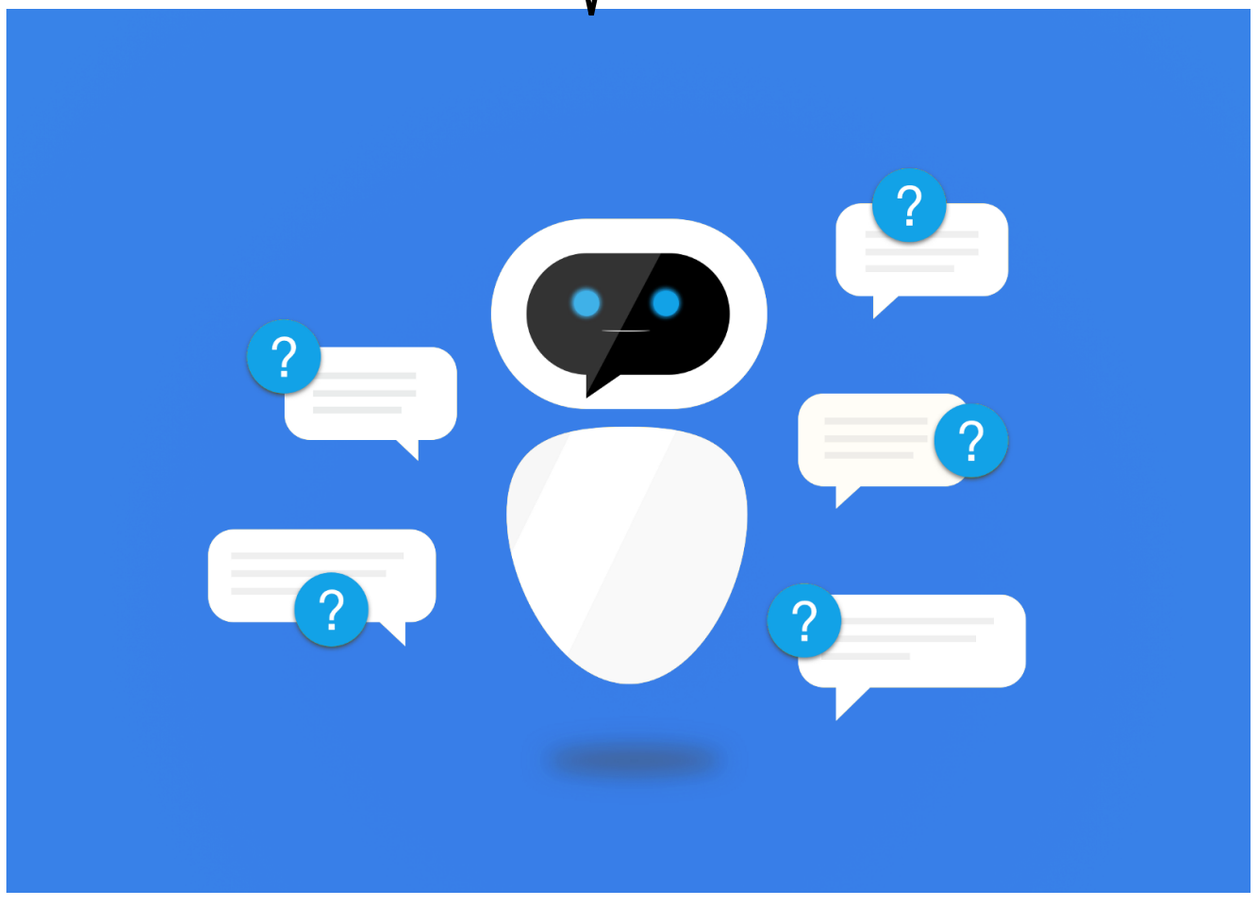
User: I want to book a flight from Minneapolis to New York

Bot: Sure. When are you planning to travel ?

User: Can you book plane to New York from Minneapolis

Conversational Agents

Sorry, I don’t understand what you’re saying



User: I want to book a flight from Minneapolis to New York

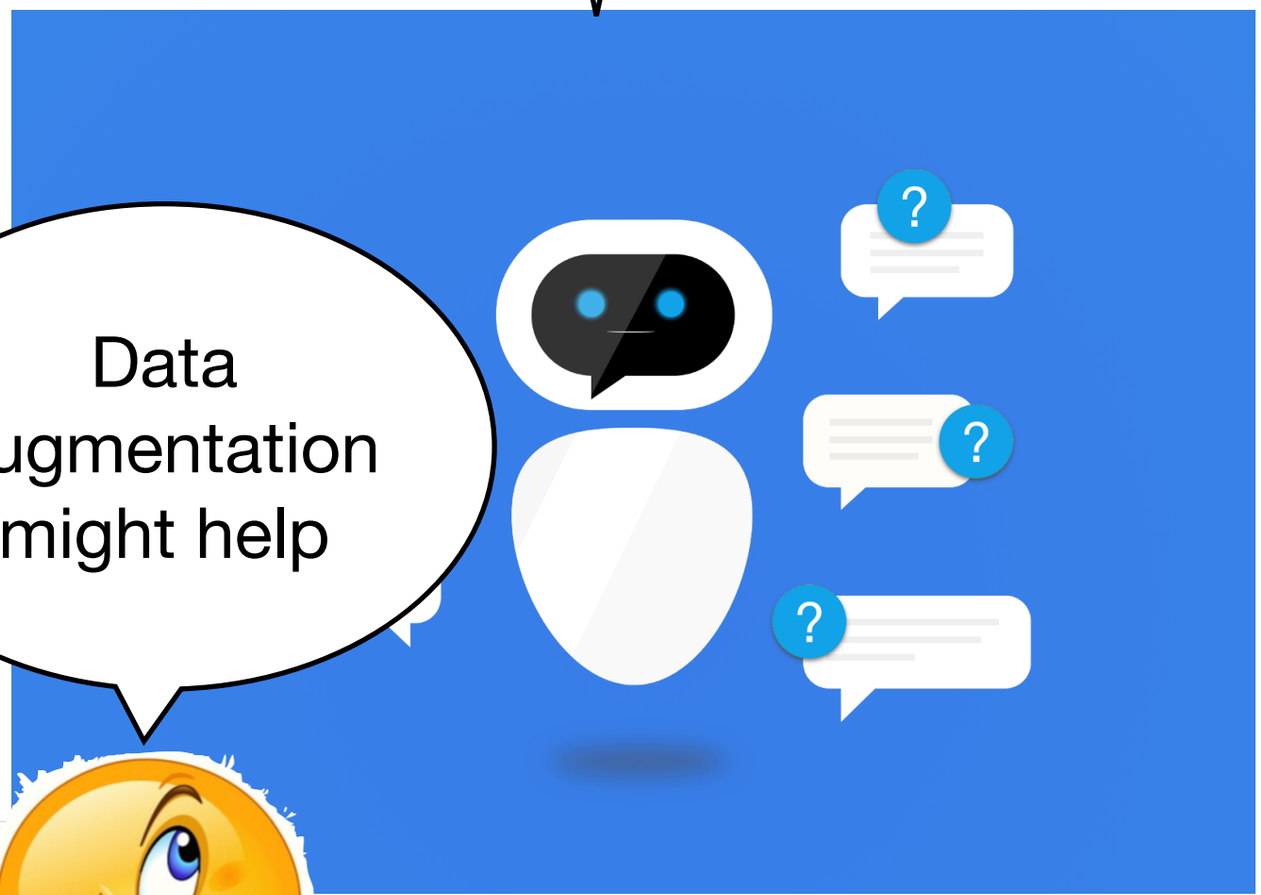
Bot: Sure. When are you planning to travel ?

User: Can you book plane to New York from Minneapolis

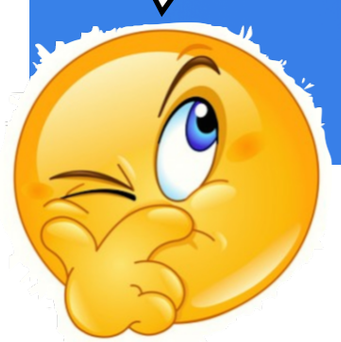
Bot: Sorry, I don’t understand what you’re saying

Conversational Agents

Sorry, I don’t understand what you’re saying



Data augmentation might help



User: I want to book a flight from Minneapolis to New York

Bot: Sure. When are you planning to travel ?

User: Can you book plane to New York from Minneapolis

Bot: Sorry, I don’t understand what you’re saying

Paraphrase Generation

Rephrasing a given text in multiple ways

Paraphrase Generation

Rephrasing a given text in multiple ways

Source	how do i increase body height ?

Paraphrase Generation

Rephrasing a given text in multiple ways

Source	how do i increase body height ?
Paraphrases	• how could i increase my height ? • what should i do to increase body height ? • what are the ways to increase height ? • are there some ways to increase body height ?

Paraphrase Generation

Rephrasing a given text in multiple ways

Source	how do i increase body height ?
Paraphrases	• how could i increase my height ? • what should i do to increase body height ? • what are the ways to increase height ? • are there some ways to increase body height ?

Fidelity
(Meaning preserving)

Paraphrase Generation

Rephrasing a given text in multiple ways

Source	how do i increase body height ?
Paraphrases	• how could i increase my height ? • what should i do to increase body height ? • what are the ways to increase height ? • are there some ways to increase body height ?

Fidelity
(Meaning preserving)

Diversity
(Lexical & syntactical variety)

Some baselines miss Diversity

Some baselines miss Diversity

Synonym or phrase replacement

Some baselines miss Diversity

Synonym or phrase replacement	
Source	how do i increase body height ?

Some baselines miss Diversity

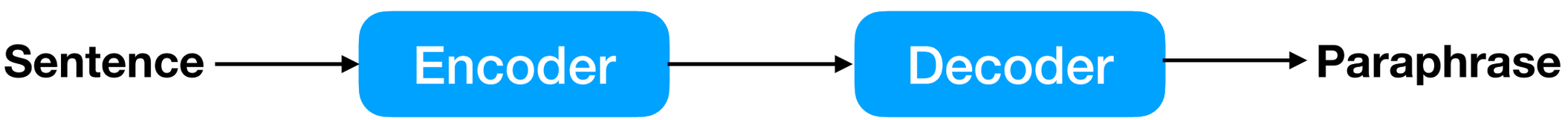
Synonym or phrase replacement	
Source	how do i increase body height ?
Synonym	how do i grow body height ?

Some baselines miss Diversity

Synonym or phrase replacement	
Source	how do i increase body height ?
Synonym	how do i grow body height ?
Phrase	how do i increase the body measurement vertically?

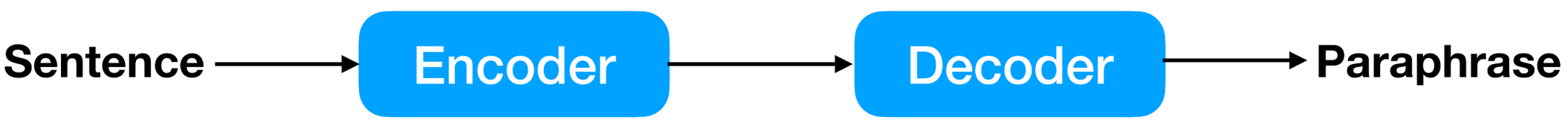
Some baselines miss Diversity

Synonym or phrase replacement	
Source	how do i increase body height ?
Synonym	how do i grow body height ?
Phrase	how do i increase the body measurement vertically?



Some baselines miss Diversity

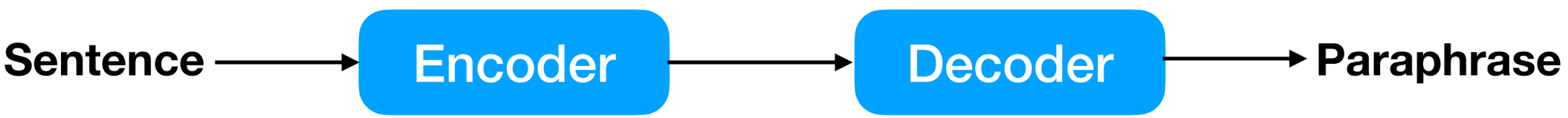
Synonym or phrase replacement	
Source	how do i increase body height ?
Synonym	how do i grow body height ?
Phrase	how do i increase the body measurement vertically?



Subsequence Selection - Beam Search (Top-k)

Some baselines miss Diversity

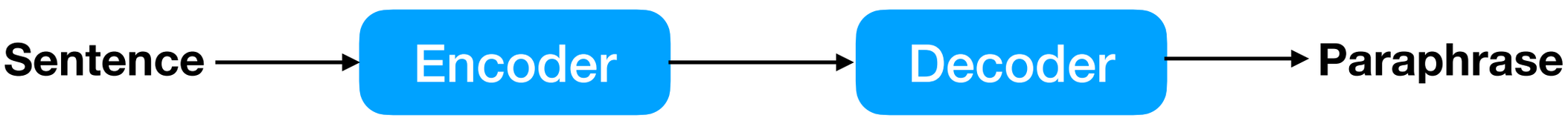
Synonym or phrase replacement	
Source	how do i increase body height ?
Synonym	how do i grow body height ?
Phrase	how do i increase the body measurement vertically?



Subsequence Selection - Beam Search (Top-k)	
Source	how do i increase body height ?

Some baselines miss Diversity

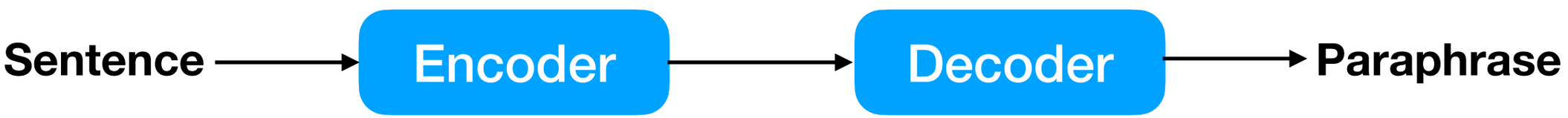
Synonym or phrase replacement	
Source	how do i increase body height ?
Synonym	how do i grow body height ?
Phrase	how do i increase the body measurement vertically?



Subsequence Selection - Beam Search (Top-k)	
Source	how do i increase body height ?
Beam	• how do i increase my height ? • how do i increase my body height ? • how do i increase the height ? • how would i increase my body height ?

Some baselines miss Diversity

Synonym or phrase replacement	
Source	how do i increase body height ?
Synonym	how do i grow body height ?
Phrase	how do i increase the body measurement vertically?

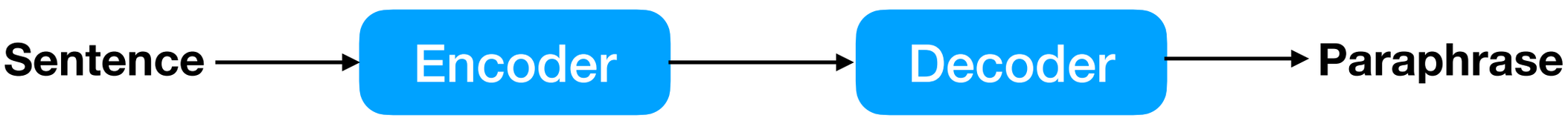


Subsequence Selection - Beam Search (Top-k)	
Source	how do i increase body height ?
Beam	• how do i increase my height ? • how do i increase my body height ? • how do i increase the height ? • how would i increase my body height ?

Fidelity ✓

Some baselines miss Diversity

Synonym or phrase replacement	
Source	how do i increase body height ?
Synonym	how do i grow body height ?
Phrase	how do i increase the body measurement vertically?



Subsequence Selection - Beam Search (Top-k)	
Source	how do i increase body height ?
Beam	• how do i increase my height ? • how do i increase my body height ? • how do i increase the height ? • how would i increase my body height ?

Fidelity ✓ **Diversity** ✗

Some baselines miss Fidelity

Some baselines miss Fidelity

Subsequence Selection - Beam Search (Diverse selection)

Some baselines miss Fidelity

Subsequence Selection - Beam Search (Diverse selection)	
Source	how do i increase body height ?

Some baselines miss Fidelity

Subsequence Selection - Beam Search (Diverse selection)	
Source	how do i increase body height ?
Beam	• how do i increase my height ? • how can i decrease my body weight ? • what do i do to increase the height ? • i am 17, what steps to take to decrease weight ?

Some baselines miss Fidelity

Subsequence Selection - Beam Search (Diverse selection)	
Source	how do i increase body height ?
Beam	• how do i increase my height ? • how can i decrease my body weight ? • what do i do to increase the height ? • i am 17, what steps to take to decrease weight ?

Diversity ✓

Some baselines miss Fidelity

Subsequence Selection - Beam Search (Diverse selection)	
Source	how do i increase body height ?
Beam	• how do i increase my height ? • how can i decrease my body weight ? • what do i do to increase the height ? • i am 17, what steps to take to decrease weight ?

Fidelity ✗ **Diversity ✓**

What we need

What we need

Fidelity ✓

Diversity ✓

What we need

Fidelity ✓ **Diversity** ✓

DiPS

Find *k* diverse paraphrases with high fidelity

Method based on subset
selection of candidate (sub)sequences

Subset Selection

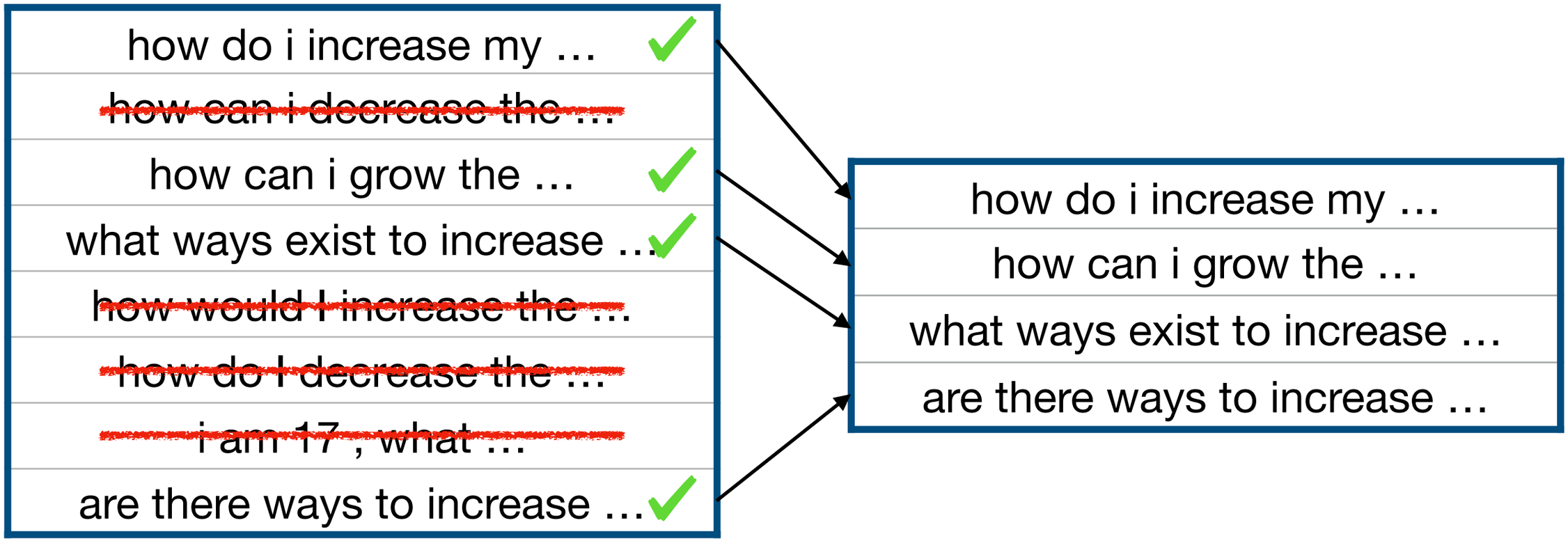
Induce Diversity while not compromising on Fidelity

how do i increase my …
how can i decrease the …
how can i grow the …
what ways exist to increase …
how would I increase the …
how do I decrease the …
i am 17 , what …
are there ways to increase …

$$V^t$$

Subset Selection

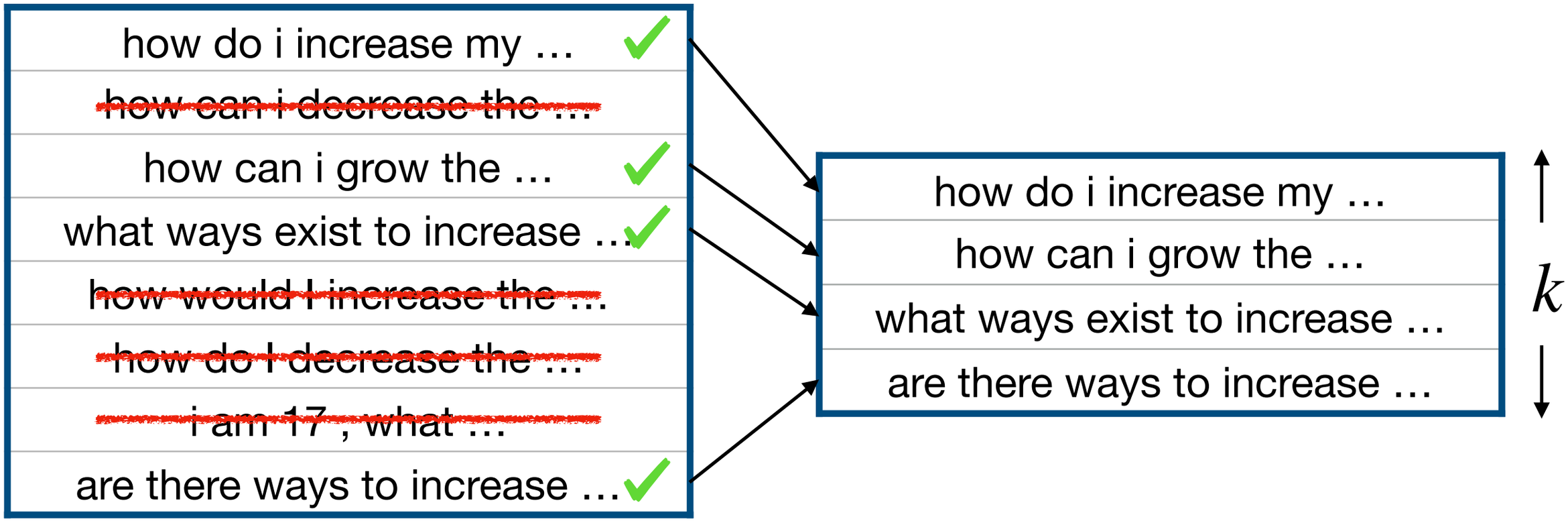
Induce Diversity while not compromising on Fidelity



$$V^t \longrightarrow X$$

Subset Selection

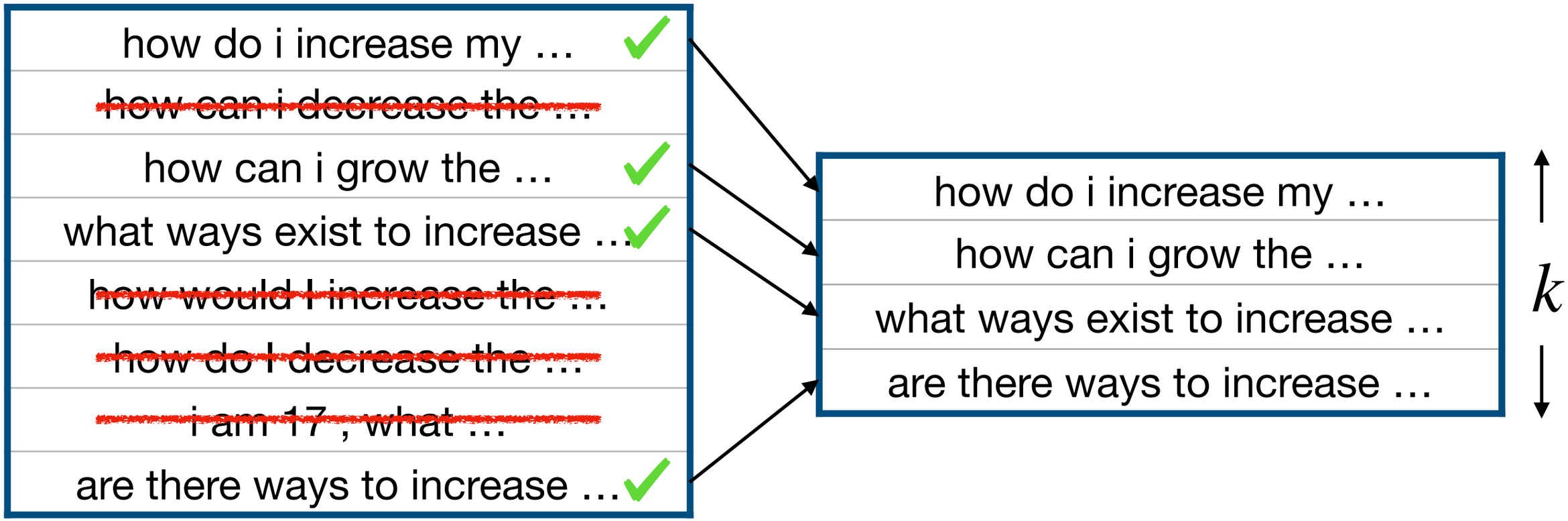
Induce Diversity while not compromising on Fidelity



$$V^t \xrightarrow{\textbf{argmax}_{X \subseteq V^t,\, |X|=k} F(X)} X$$

Subset Selection

Induce Diversity while not compromising on Fidelity



$$V^t \xrightarrow{\textbf{argmax}_{X \subseteq V^t,\, |X|=k} F(X)} X$$

If F is sub modular + monotone = Greedy algo. with good bounds exists

Sub-modularity

Sub-modularity

$$F : 2^V \to \mathbb{R}$$

$$X \subset Y \subset V, s \in V \backslash Y, \qquad F(X \cup \{s\}) - F(X) \geq F(Y \cup \{s\}) - F(Y)$$

Sub-modularity

$$F : 2^V \to \mathbb{R}$$

$$X \subset Y \subset V, s \in V \backslash Y, \qquad F(X \cup \{s\}) - F(X) \geq F(Y \cup \{s\}) - F(Y)$$

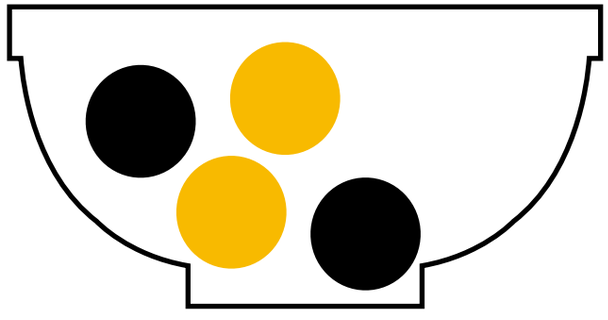
F = # Unique Coloured items

Sub-modularity

$$F : 2^V \rightarrow \mathbb{R}$$

$$X \subset Y \subset V, s \in V \backslash Y, \qquad F(X \cup \{s\}) - F(X) \geq F(Y \cup \{s\}) - F(Y)$$

F = # Unique Coloured items



 # Items = 4

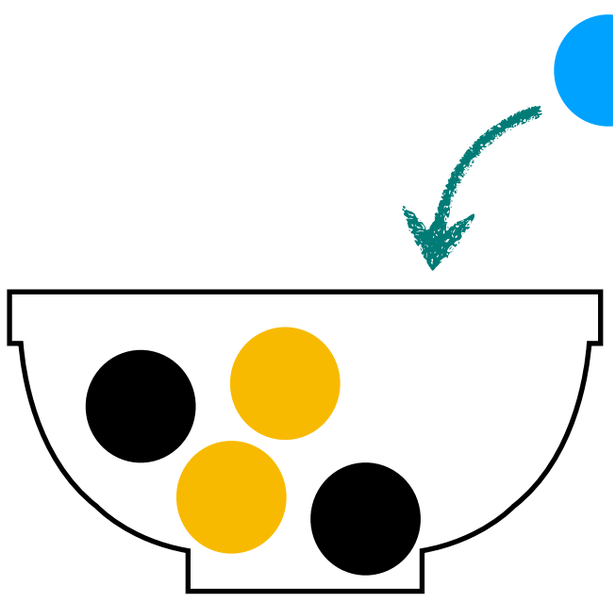
F = 2

Sub-modularity

$$F : 2^V \rightarrow \mathbb{R}$$

$$X \subset Y \subset V, s \in V \backslash Y, \qquad F(X \cup \{s\}) - F(X) \geq F(Y \cup \{s\}) - F(Y)$$

F = # Unique Coloured items



 # Items = 4

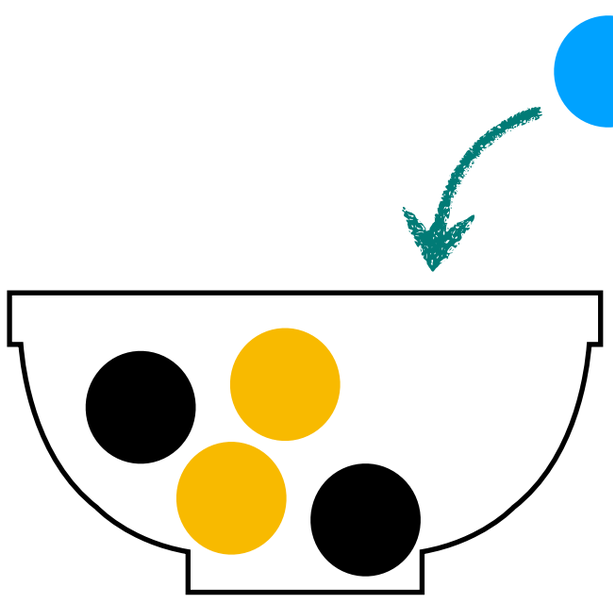
F = 2

Sub-modularity

$$F : 2^V \to \mathbb{R}$$

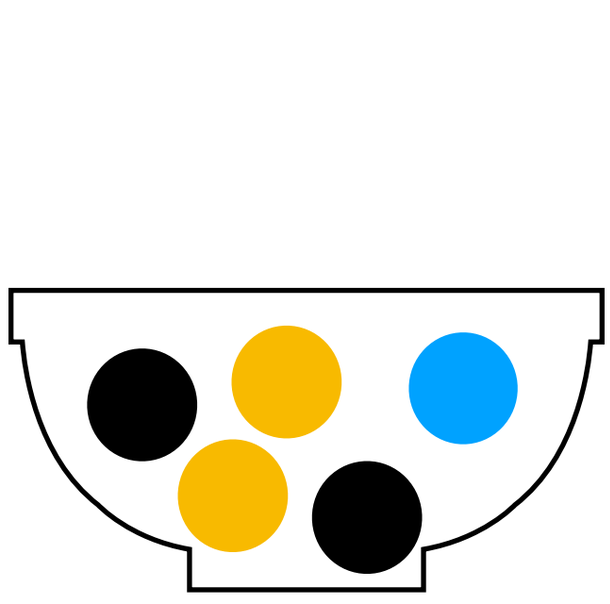
$$X \subset Y \subset V, s \in V \backslash Y, \qquad F(X \cup \{s\}) - F(X) \geq F(Y \cup \{s\}) - F(Y)$$

F = # Unique Coloured items



 # Items = 4

F = 2



 # Items = 4 + 1

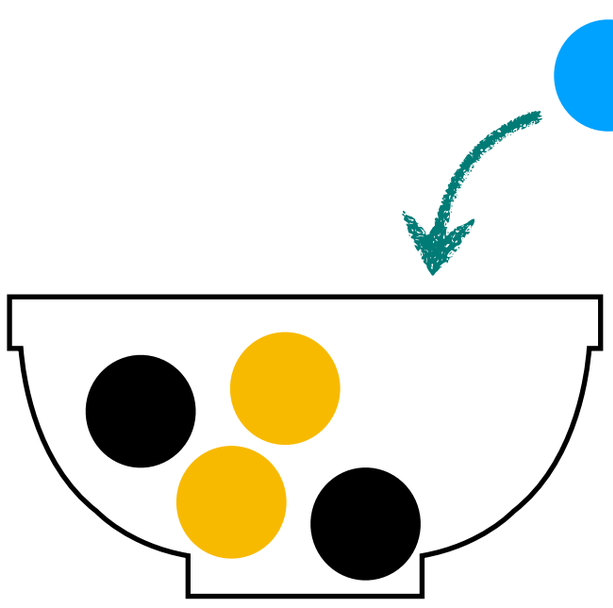
F = 2 + 1

Sub-modularity

$$F : 2^V \to \mathbb{R}$$

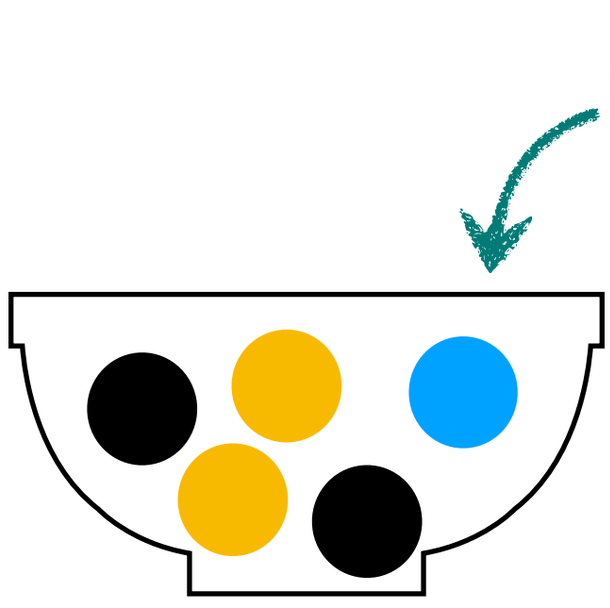
$$X \subset Y \subset V, s \in V \backslash Y, \qquad F(X \cup \{s\}) - F(X) \geq F(Y \cup \{s\}) - F(Y)$$

F = # Unique Coloured items



 # Items = 4

F = 2



 # Items = 4 + 1

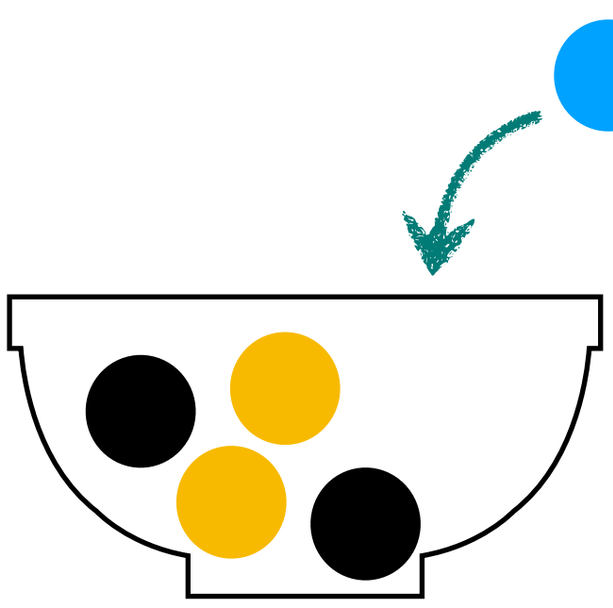
F = 2 + 1

Sub-modularity

$$F : 2^V \to \mathbb{R}$$

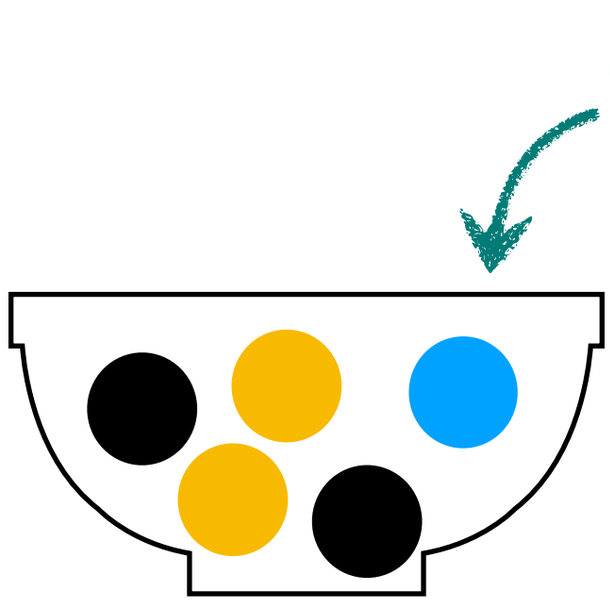
$$X \subset Y \subset V, s \in V \backslash Y, \qquad F(X \cup \{s\}) - F(X) \geq F(Y \cup \{s\}) - F(Y)$$

F = # Unique Coloured items



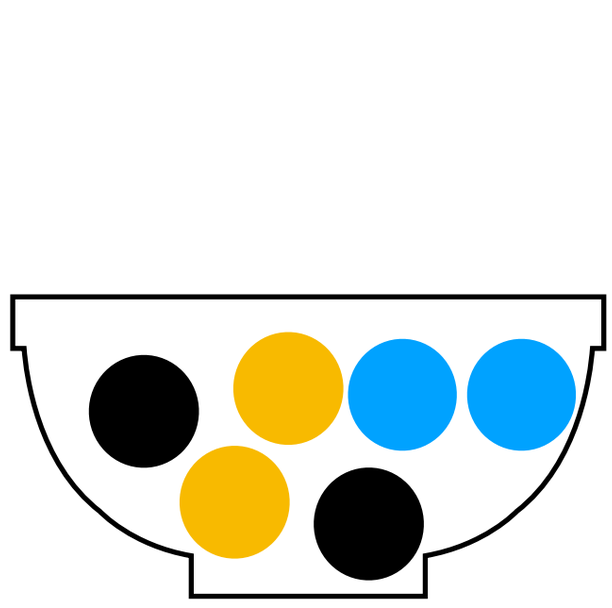
 # Items = 4

F = 2



 # Items = 4 + 1

F = 2 + 1



 # Items = 5 + 1

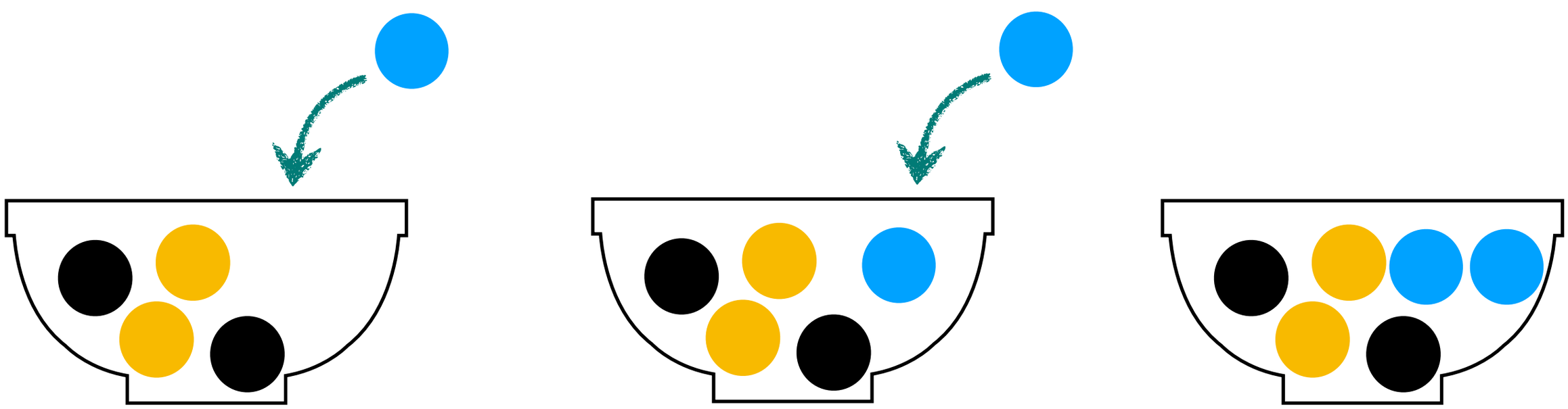
F = 3 + 0

Sub-modularity

$$F : 2^V \rightarrow \mathbb{R}$$

$$X \subset Y \subset V, s \in V \backslash Y, \qquad F(X \cup \{s\}) - F(X) \geq F(Y \cup \{s\}) - F(Y)$$

F = # Unique Coloured items



 # Items = 4

F = 2

 # Items = 4 + 1

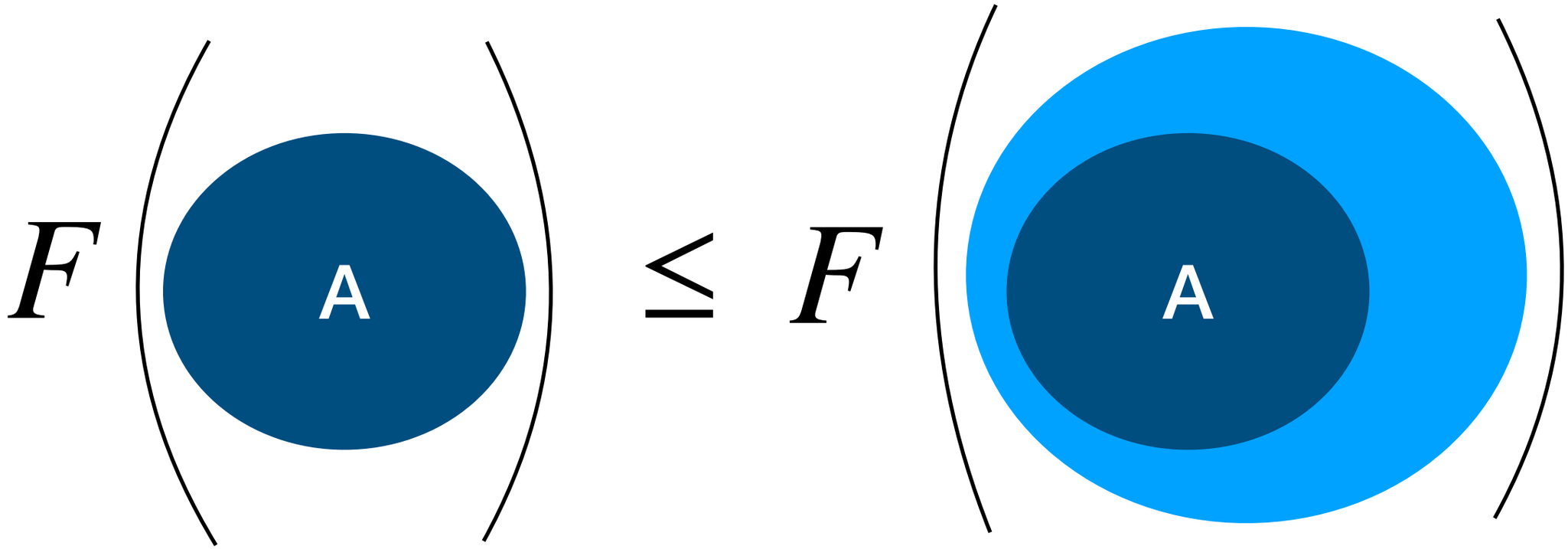
F = 2 + 1

 # Items = 5 + 1

F = 3 + 0

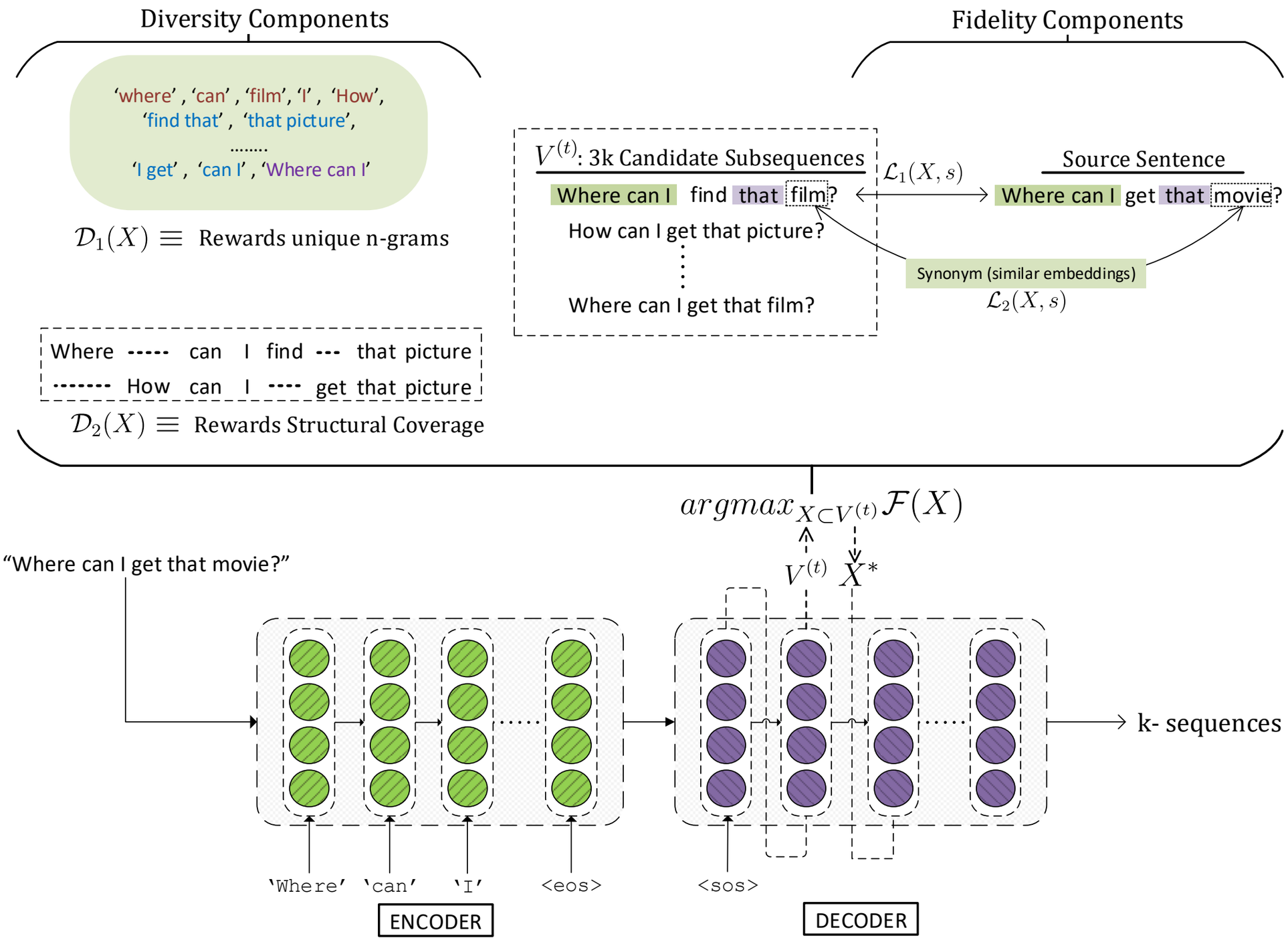
Diminishing Returns

Monotonicity



DiPS

Induce Diversity while not compromising on Fidelity



DiPS

Induce Diversity while not compromising on Fidelity

Diversity Components

'where' , 'can' , 'film', 'I' , 'How',
'find that' , 'that picture',
.....
'I get' , 'can I' , 'Where can I'

$\mathcal{D}_1(X) \equiv$ Rewards unique n-grams

Where ----- can I find --- that picture
------- How can I ---- get that picture

$\mathcal{D}_2(X) \equiv$ Rewards Structural Coverage

Fidelity Components

$V^{(t)}$: 3k Candidate Subsequences

Where can I find that film?

How can I get that picture?

⋮

Where can I get that film?

$\mathcal{L}_1(X, s)$

Source Sentence

Where can I get that movie?

Synonym (similar embeddings)

$\mathcal{L}_2(X, s)$

$argmax_{X \subset V^{(t)}} \mathcal{F}(X)$

$V^{(t)}$ X^*

"Where can I get that movie?"

Encoder

'Where' 'can' 'I' <eos>

ENCODER

<sos>

DECODER

k- sequences

DiPS

Induce Diversity while not compromising on Fidelity

Diversity Components

'where' , 'can' , 'film', 'I' , 'How',
'find that' , 'that picture',
.....
'I get' , 'can I' , 'Where can I'

$\mathcal{D}_1(X) \equiv$ Rewards unique n-grams

Where ----- can I find --- that picture
------- How can I ---- get that picture

$\mathcal{D}_2(X) \equiv$ Rewards Structural Coverage

Fidelity Components

$V^{(t)}$: 3k Candidate Subsequences

Where can I find that film?

How can I get that picture?

Where can I get that film?

$\mathcal{L}_1(X, s)$

Source Sentence

Where can I get that movie?

Synonym (similar embeddings)

$\mathcal{L}_2(X, s)$

$argmax_{X \subset V^{(t)}} \mathcal{F}(X)$

"Where can I get that movie?"

'Where' 'can' 'I' <eos>

ENCODER

Encoder

$V^{(t)}$ X^*

<sos>

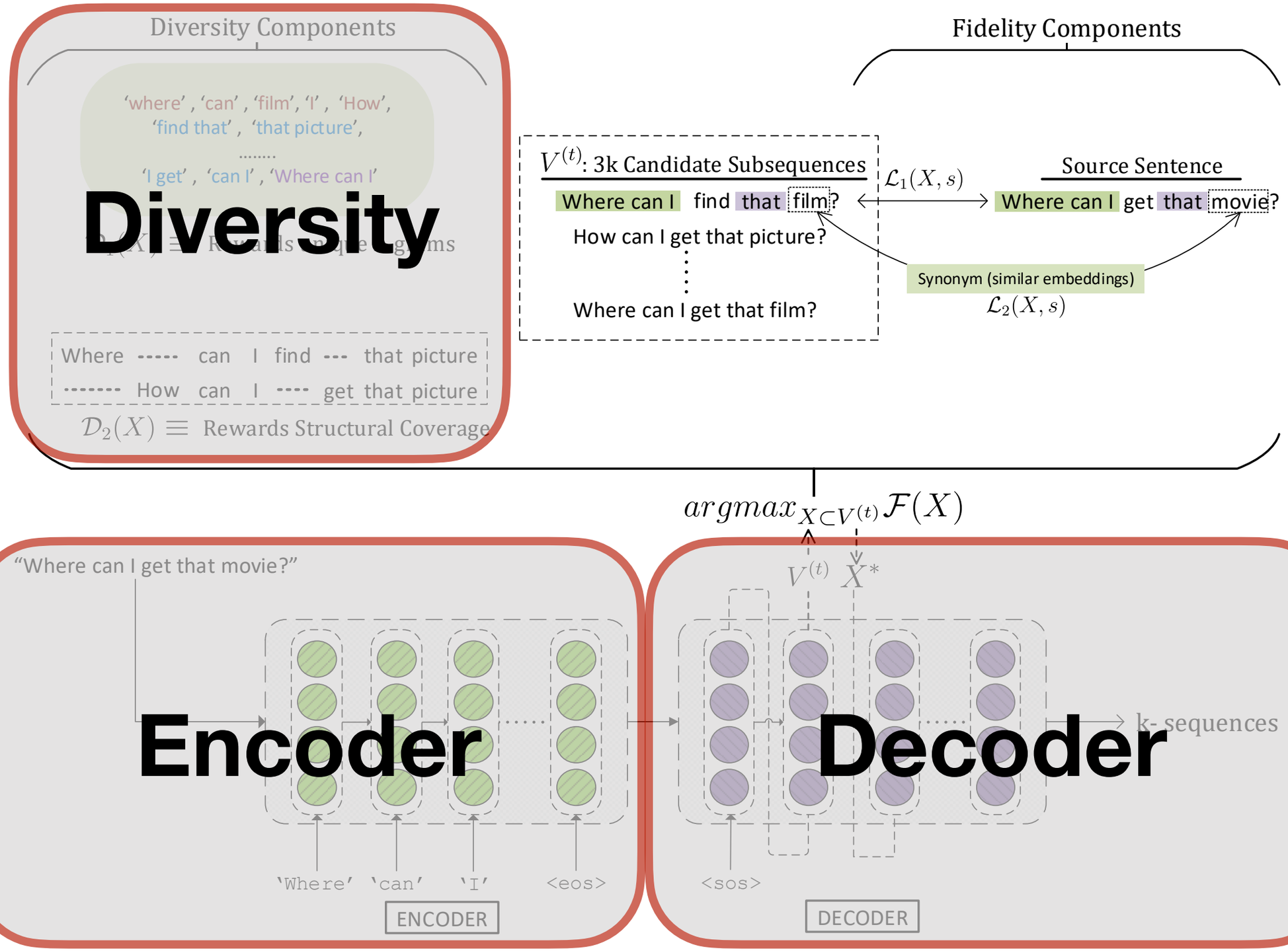
DECODER

Decoder

k- sequences

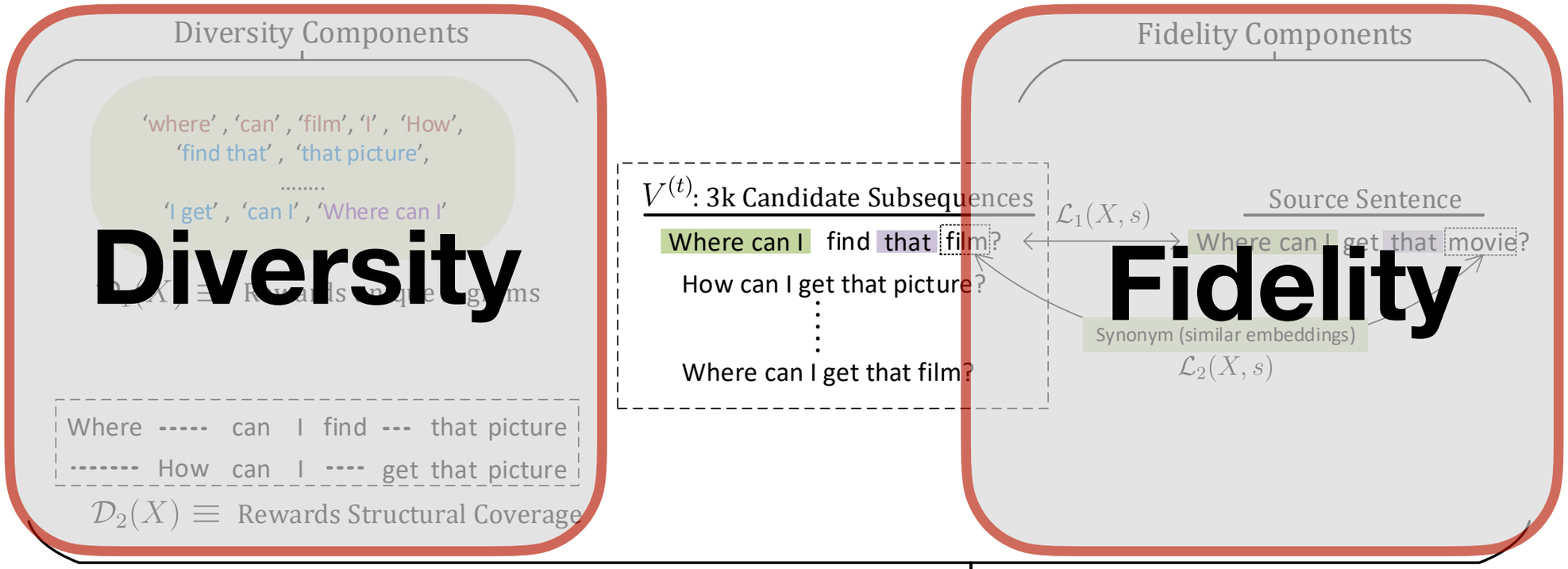
DiPS

Induce Diversity while not compromising on Fidelity



DiPS

Induce Diversity while not compromising on Fidelity



DiPS

Induce Diversity while not compromising on Fidelity

Diversity Components

'where' , 'can' , 'film', 'I' , 'How',
'find that' , 'that picture',
.....
'I get' , 'can I' , 'Where can I'

$\mathcal{D}_1(X) \equiv$ Rewards unique n-grams

Where ----- can I find --- that picture
------- How can I ---- get that picture

$\mathcal{D}_2(X) \equiv$ Rewards Structural Coverage

Fidelity Components

$V^{(t)}$: 3k Candidate Subsequences

Where can I find that film?

How can I get that picture?

Where can I get that film?

$\mathcal{L}_1(X, s)$

Source Sentence

Where can I get that movie?

Synonym (similar embeddings)

$\mathcal{L}_2(X, s)$

$argmax_{X \subset V^{(t)}} \mathcal{F}(X)$

$V^{(t)}$ X^*

"Where can I get that movie?"

'Where' 'can' 'I' <eos>

<sos>

k- sequences

ENCODER

DECODER

DiPS

Induce Diversity while not compromising on Fidelity

Diversity Components

'where' , 'can' , 'film', 'I' , 'How',
'find that' , 'that picture',
.....
'I get' , 'can I' , 'Where can I'

$\mathcal{D}_1(X) \equiv$ Rewards unique n-grams

Where ----- can I find --- that picture
------- How can I ---- get that picture

$\mathcal{D}_2(X) \equiv$ Rewards Structural Coverage

Fidelity Components

$V^{(t)}$: 3k Candidate Subsequences

Where can I find that film?

How can I get that picture?

Where can I get that film?

$\mathcal{L}_1(X, s)$

Source Sentence

Where can I get that movie?

Synonym (similar embeddings)

$\mathcal{L}_2(X, s)$

$argmax_{X \subset V^{(t)}} \mathcal{F}(X)$

"Where can I get that movie?"

$V^{(t)}$ X^*

'Where' 'can' 'I' <eos>

<sos>

ENCODER

DECODER

k- sequences

DiPS

Induce Diversity while not compromising on Fidelity

Diversity Components

'where' , 'can' , 'film', 'I' , 'How',
'find that' , 'that picture',
.....
'I get' , 'can I' , 'Where can I'

$\mathcal{D}_1(X) \equiv$ Rewards unique n-grams

Where ----- can I find --- that picture
------- How can I ---- get that picture

$\mathcal{D}_2(X) \equiv$ Rewards Structural Coverage

Fidelity Components

$V^{(t)}$: 3k Candidate Subsequences

Where can I find that film?

How can I get that picture?

Where can I get that film?

$\mathcal{L}_1(X, s)$

Source Sentence

Where can I get that movie?

Synonym (similar embeddings)

$\mathcal{L}_2(X, s)$

$argmax_{X \subset V^{(t)}} \mathcal{F}(X)$

$V^{(t)}$ X^*

"Where can I get that movie?"

'Where' 'can' 'I' <eos>

<sos>

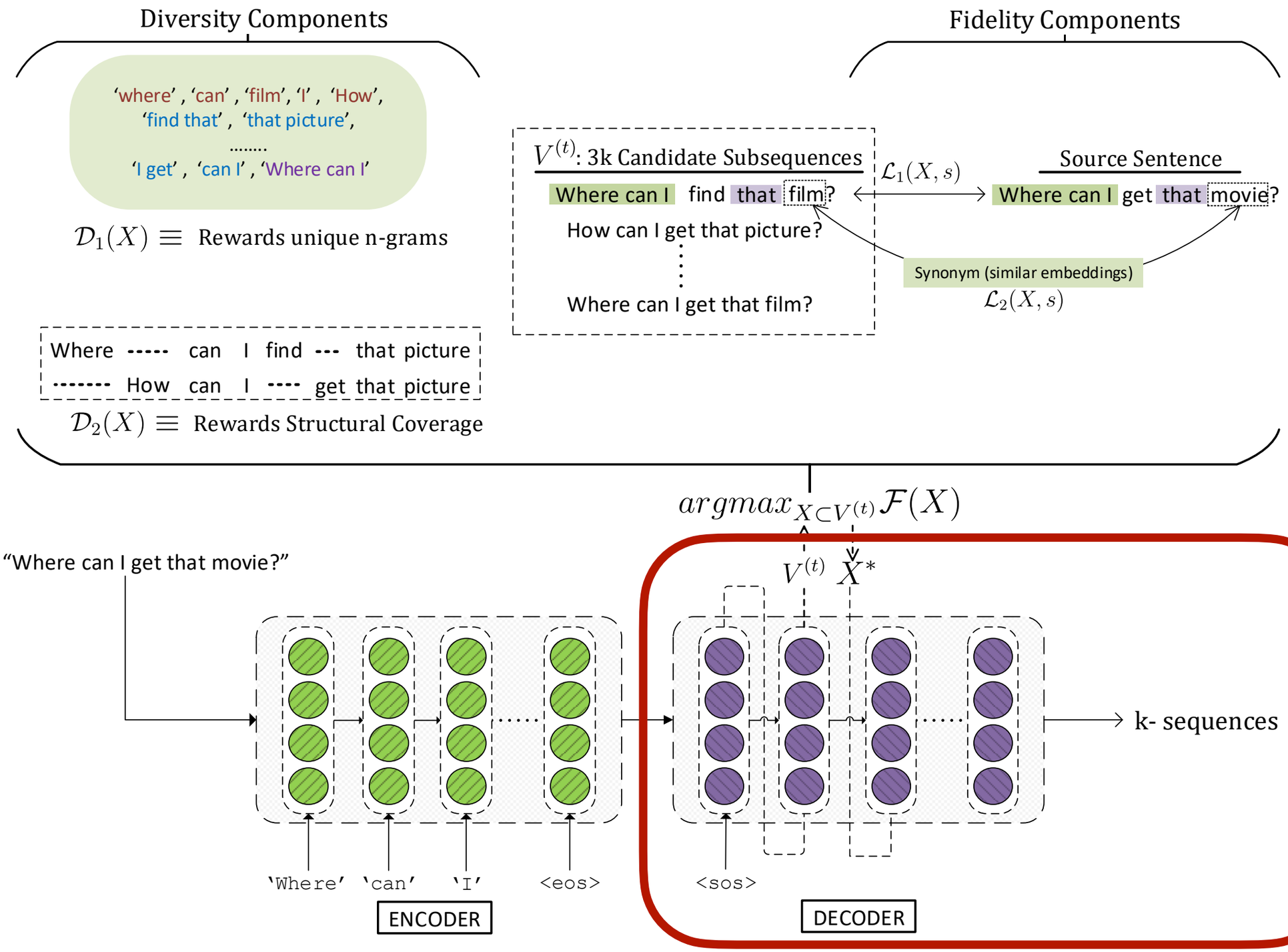
k- sequences

ENCODER

DECODER

DiPS

Induce Diversity while not compromising on Fidelity



DiPS

Induce Diversity while not compromising on Fidelity

Diversity Components

'where' , 'can' , 'film', 'I' , 'How',
'find that' , 'that picture',
.....
'I get' , 'can I' , 'Where can I'

$\mathcal{D}_1(X) \equiv$ Rewards unique n-grams

Where ----- can I find --- that picture
------- How can I ---- get that picture

$\mathcal{D}_2(X) \equiv$ Rewards Structural Coverage

Fidelity Components

$V^{(t)}$: 3k Candidate Subsequences

Where can I find that film?

How can I get that picture?

Where can I get that film?

$\mathcal{L}_1(X, s)$

Source Sentence

Where can I get that movie?

Synonym (similar embeddings)

$\mathcal{L}_2(X, s)$

$argmax_{X \subset V^{(t)}} \mathcal{F}(X)$

$V^{(t)}$ X^*

"Where can I get that movie?"

'Where' 'can' 'I' <eos>

<sos>

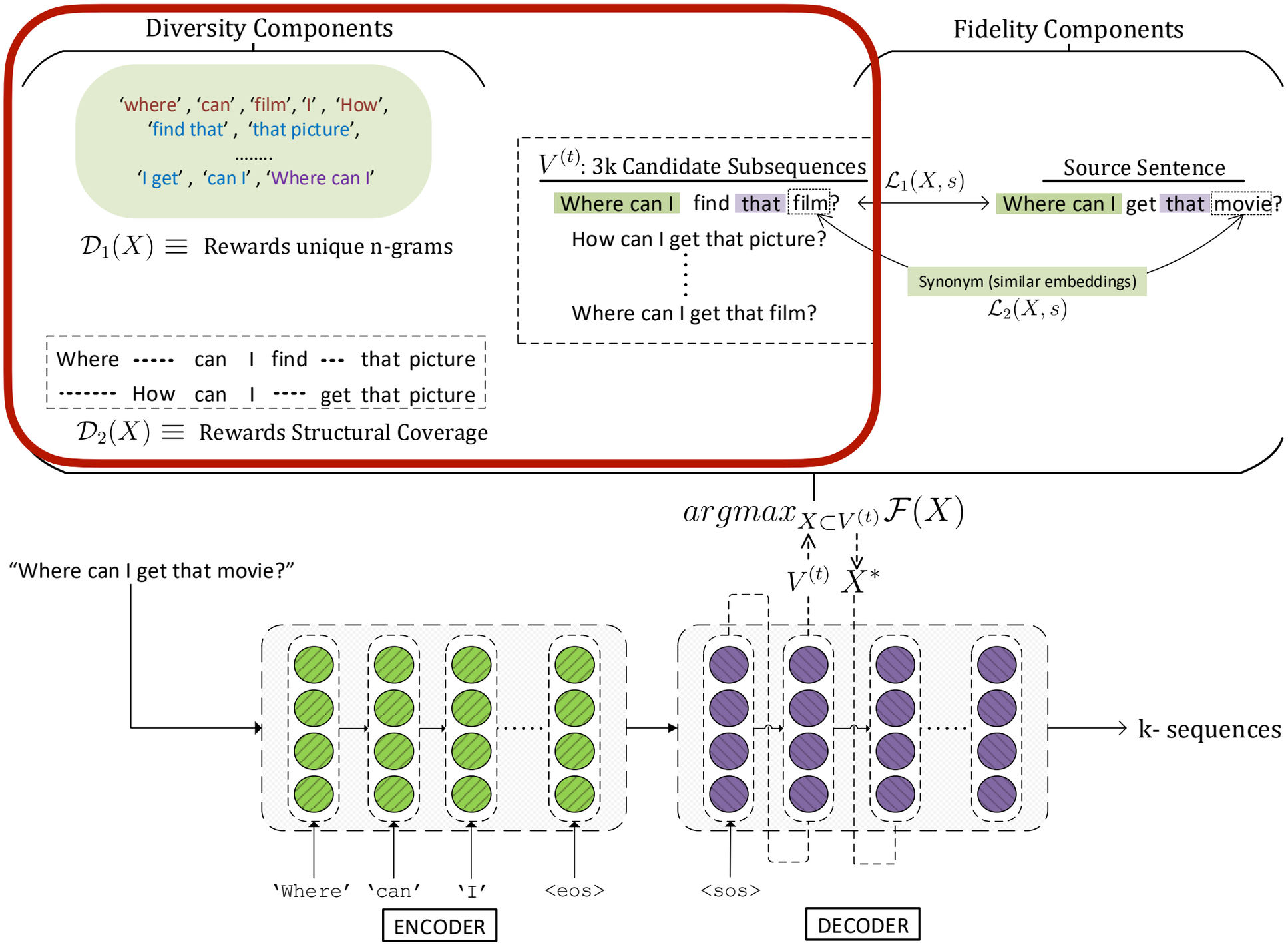
k- sequences

ENCODER

DECODER

DiPS

Induce Diversity while not compromising on Fidelity



Diversity Components

$V^{(t)}$: 3k Candidate Subsequences

Where can I find that film?

How can I get that picture?

⋮

Where can I get that film?

Diversity Components

'where' , 'can' , 'film', 'I' , 'How',
'find that' , 'that picture',
.....
'I get' , 'can I' , 'Where can I'

$\mathcal{D}_1(X) \equiv$ Rewards unique n-grams

Where ----- can I find --- that picture

------- How can I ---- get that picture

$\mathcal{D}_2(X) \equiv$ Rewards Structural Coverage

Diversity Components

$V^{(t)}$: 3k Candidate Subsequences

Where can I find that film?

How can I get that picture?

⋮

Where can I get that film?

Diversity Components

'where' , 'can' , 'film', 'I' , 'How',
'find that' , 'that picture',
.....
'I get' , 'can I' , 'Where can I'

$\mathcal{D}_1(X) \equiv$ Rewards unique n-grams

N-gram uniqueness

$$\sum_{n=1}^{N} \beta^n \left| \bigcup_{x \in X} x_{n-gram} \right|$$

Where ----- can I find --- that picture

------- How can I ---- get that picture

$\mathcal{D}_2(X) \equiv$ Rewards Structural Coverage

Diversity Components

$V^{(t)}$: 3k Candidate Subsequences

Where can I find that film?

How can I get that picture?

⋮

Where can I get that film?

Diversity Components

'where' , 'can' , 'film', 'I' , 'How',
'find that' , 'that picture',
.....
'I get' , 'can I' , 'Where can I'

$\mathcal{D}_1(X) \equiv$ Rewards unique n-grams

N-gram uniqueness

$$\sum_{n=1}^{N} \beta^n \left| \bigcup_{x \in X} x_{n-gram} \right|$$

Where ----- can I find --- that picture
------- How can I ---- get that picture

$\mathcal{D}_2(X) \equiv$ Rewards Structural Coverage

Structural Coverage

$$\sum_{x_i \in V^{(t)}} \sum_{x_j \in X} \mathcal{R}(x_i, x_j)$$

$$\mathcal{R}(x_i, x_j) = 1 - \frac{\textbf{EditDistance}(x_i, x_j)}{|x_i| + |x_j|}$$

DiPS

Induce Diversity while not compromising on Fidelity

Diversity Components

'where' , 'can' , 'film', 'I' , 'How',
'find that' , 'that picture',
.....
'I get' , 'can I' , 'Where can I'

$\mathcal{D}_1(X) \equiv$ Rewards unique n-grams

Where ----- can I find --- that picture
------- How can I ---- get that picture

$\mathcal{D}_2(X) \equiv$ Rewards Structural Coverage

Fidelity Components

$V^{(t)}$: 3k Candidate Subsequences

Where can I find that film?

How can I get that picture?

Where can I get that film?

$\mathcal{L}_1(X, s)$

Source Sentence

Where can I get that movie?

Synonym (similar embeddings)

$\mathcal{L}_2(X, s)$

$argmax_{X \subset V^{(t)}} \mathcal{F}(X)$

$V^{(t)}$ X^*

"Where can I get that movie?"

'Where' 'can' 'I' <eos>

<sos>

k- sequences

ENCODER

DECODER

DiPS

Induce Diversity while not compromising on Fidelity

Diversity Components

'where' , 'can' , 'film', 'I' , 'How',
'find that' , 'that picture',
.....
'I get' , 'can I' , 'Where can I'

$\mathcal{D}_1(X) \equiv$ Rewards unique n-grams

Where ----- can I find --- that picture
------- How can I ---- get that picture

$\mathcal{D}_2(X) \equiv$ Rewards Structural Coverage

Fidelity Components

$V^{(t)}$: 3k Candidate Subsequences

Where can I find that film?
How can I get that picture?
⋮
Where can I get that film?

$\mathcal{L}_1(X, s)$

Source Sentence

Where can I get that movie?

Synonym (similar embeddings)

$\mathcal{L}_2(X, s)$

$argmax_{X \subset V^{(t)}} \mathcal{F}(X)$

$V^{(t)}$ X^*

"Where can I get that movie?"

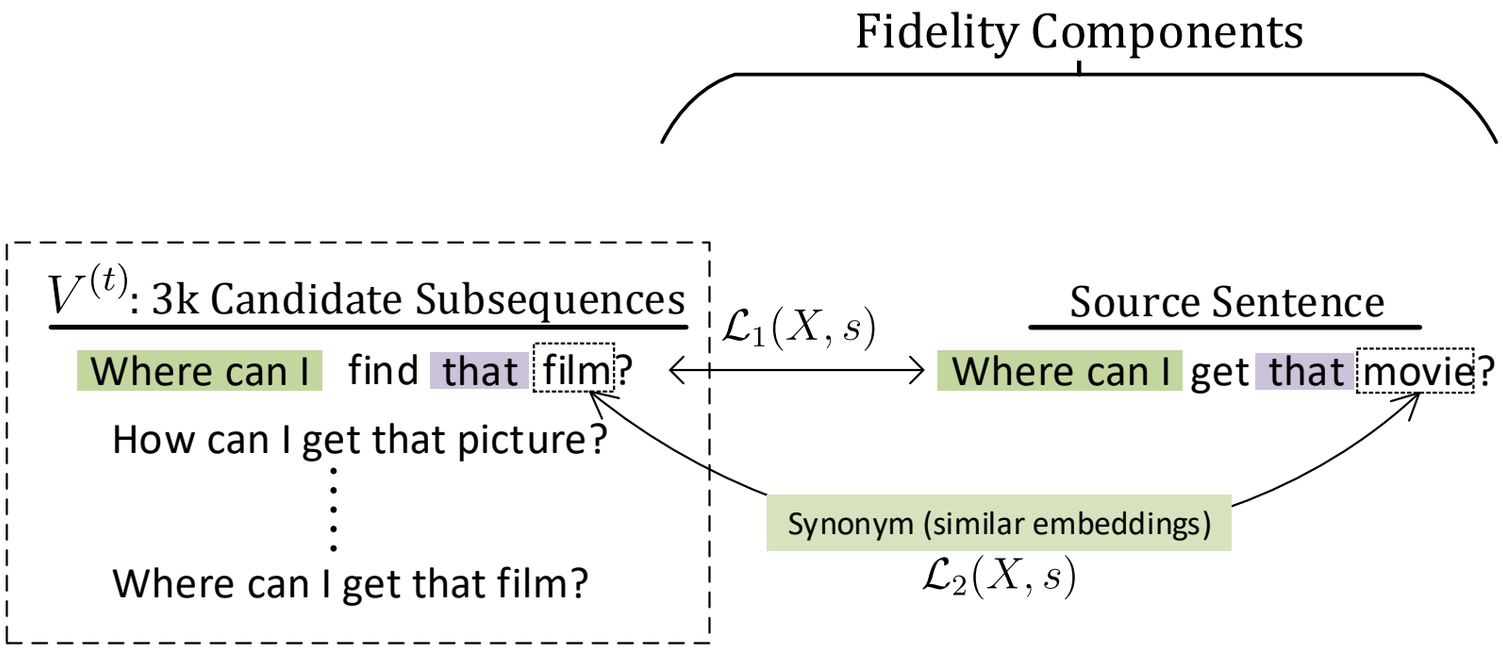
'Where' 'can' 'I' <eos> <sos>

ENCODER

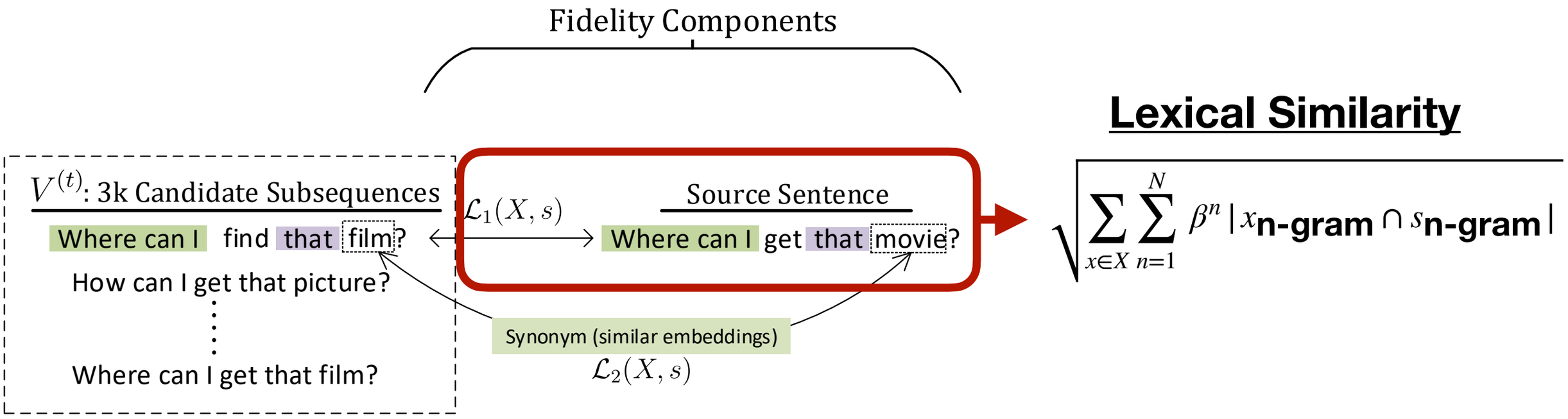
DECODER

k- sequences

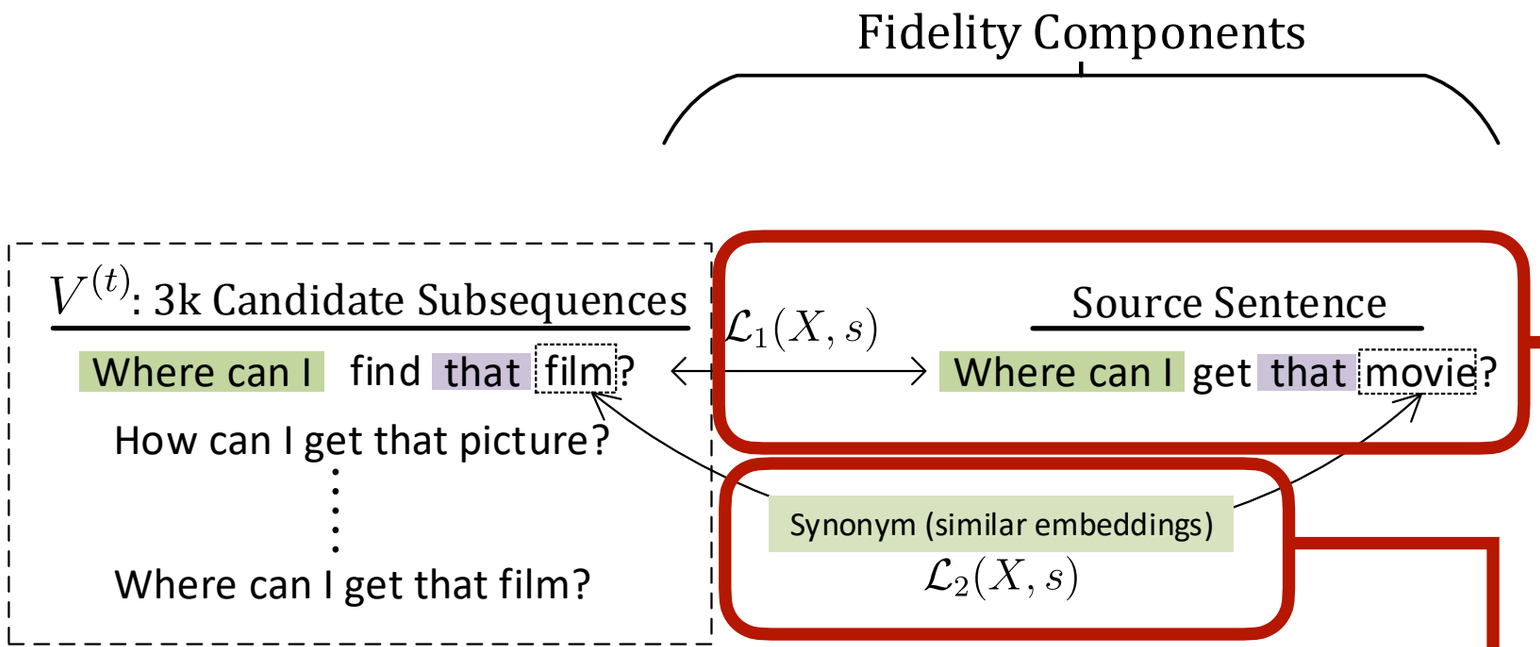
Fidelity Components



Fidelity Components



Fidelity Components



Lexical Similarity

$$\sqrt{\sum_{x \in X} \sum_{n=1}^{N} \beta^n \, |\, x_{\textbf{n-gram}} \cap s_{\textbf{n-gram}} \,|}$$

Embedding based Similarity

$$\sqrt{\sum_{x \in X} \mathcal{S}(x, s)}$$

$$\mathcal{S}(x, s) = \frac{1}{|x|} \sum_{w_i \in x} \textbf{max}_{w_j \in s} \, \psi(\mathbf{v}_{w_i}, \mathbf{v}_{w_j})$$

DiPS Objective

Diversity Components

'where' , 'can' , 'film', 'I' , 'How',
'find that' , 'that picture',
.....
'I get' , 'can I' , 'Where can I'

$\mathcal{D}_1(X) \equiv$ Rewards unique n-grams

Where ----- can I find --- that picture
------- How can I ---- get that picture

$\mathcal{D}_2(X) \equiv$ Rewards Structural Coverage

Fidelity Components

DiPS Objective

$$\textbf{argmax}_{X \subseteq V,\, X = |k|} F(X)$$

$$F(X) = \lambda(\mu_1 D_1(X) + \mu_2 D_2(X)) + (1 - \lambda)(\nu_1 L_1(X, s) + \nu_2 L_2(X, s))$$

Diversity Components

'where' , 'can' , 'film', 'I' , 'How',
'find that' , 'that picture',
.....
'I get' , 'can I' , 'Where can I'

$\mathcal{D}_1(X) \equiv$ Rewards unique n-grams

Where ----- can I find --- that picture
------- How can I ---- get that picture

$\mathcal{D}_2(X) \equiv$ Rewards Structural Coverage

Fidelity Components

$V^{(t)}$: 3k Candidate Subsequences

Where can I find that film?

How can I get that picture?

Where can I get that film?

$\mathcal{L}_1(X, s)$

Source Sentence

Where can I get that movie?

Synonym (similar embeddings)

$\mathcal{L}_2(X, s)$

DiPS Objective

$$\textbf{argmax}_{X \subseteq V,\, X=|k|} F(X)$$

$$F(X) = \lambda(\mu_1 D_1(X) + \mu_2 D_2(X)) + (1-\lambda)(\nu_1 L_1(X,s) + \nu_2 L_2(X,s))$$

Diversity Components

'where' , 'can' , 'film', 'I' , 'How',
'find that' , 'that picture',
.....
'I get' , 'can I' , 'Where can I'

$\mathcal{D}_1(X) \equiv$ Rewards unique n-grams

Where ----- can I find --- that picture
------- How can I ---- get that picture

$\mathcal{D}_2(X) \equiv$ Rewards Structural Coverage

Fidelity Components

$V^{(t)}$: 3k Candidate Subsequences

Where can I find that film?

How can I get that picture?

Where can I get that film?

$\mathcal{L}_1(X,s)$

Source Sentence

Where can I get that movie?

Synonym (similar embeddings)

$\mathcal{L}_2(X,s)$

DiPS Objective

$$\textbf{argmax}_{X \subseteq V,\, X=|k|} F(X)$$

$$F(X) = \lambda(\mu_1 D_1(X) + \mu_2 D_2(X)) + (1-\lambda)(\nu_1 L_1(X,s) + \nu_2 L_2(X,s))$$

Diversity Components

'where' , 'can' , 'film', 'I' , 'How',
'find that' , 'that picture',
.....
'I get' , 'can I' , 'Where can I'

$\mathcal{D}_1(X) \equiv$ Rewards unique n-grams

Where ----- can I find --- that picture
------- How can I ---- get that picture

$\mathcal{D}_2(X) \equiv$ Rewards Structural Coverage

Fidelity Components

$V^{(t)}$: 3k Candidate Subsequences

Where can I find that film?

How can I get that picture?

Where can I get that film?

$\mathcal{L}_1(X,s)$

Source Sentence

Where can I get that movie?

Synonym (similar embeddings)

$\mathcal{L}_2(X,s)$

Experimental Setup

Datasets

Evaluation

Experimental Setup

Datasets

1. Quora Question Pairs
2. Twitter URL
3. Yahoo-L31
4. Snips

Evaluation

Experimental Setup

Datasets

1. Quora Question Pairs
2. Twitter URL
3. Yahoo-L31
4. Snips

Evaluation

Fidelity BLEU, METEOR, TERplus

Experimental Setup

Datasets

1. Quora Question Pairs
2. Twitter URL
3. Yahoo-L31
4. Snips

Evaluation

Fidelity BLEU, METEOR, TERplus

Diversity n-distinct score

Experimental Setup

Datasets

1. Quora Question Pairs
2. Twitter URL
3. Yahoo-L31
4. Snips

Evaluation

Fidelity BLEU, METEOR, TERplus

Diversity n-distinct score

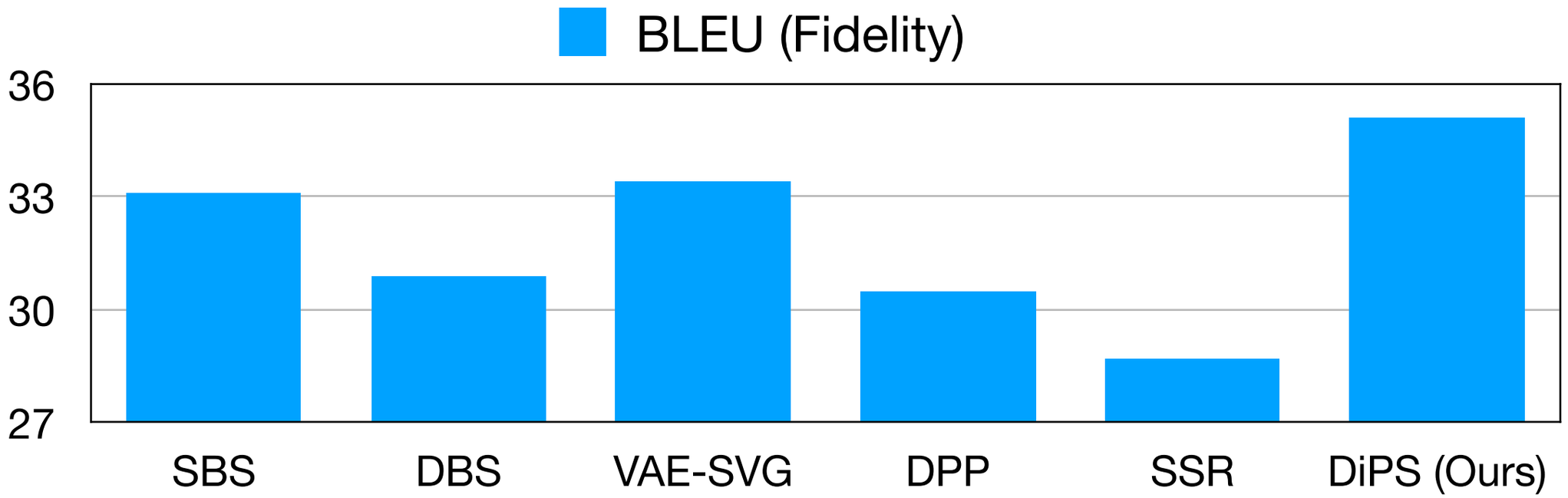
Data Augmentation Accuracy

Fidelity & Diversity

(Quora Dataset)

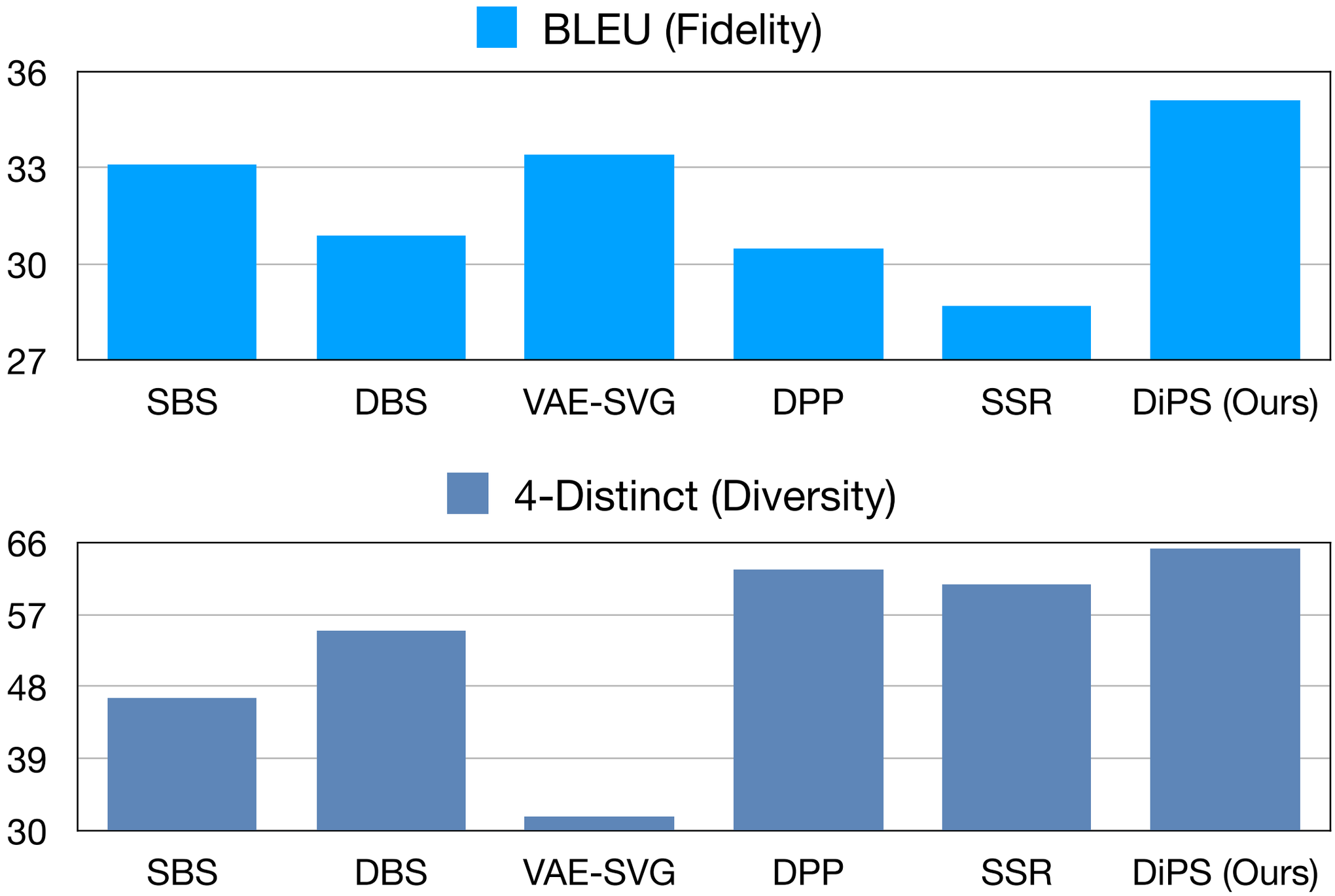
Fidelity & Diversity

(Quora Dataset)



Fidelity & Diversity

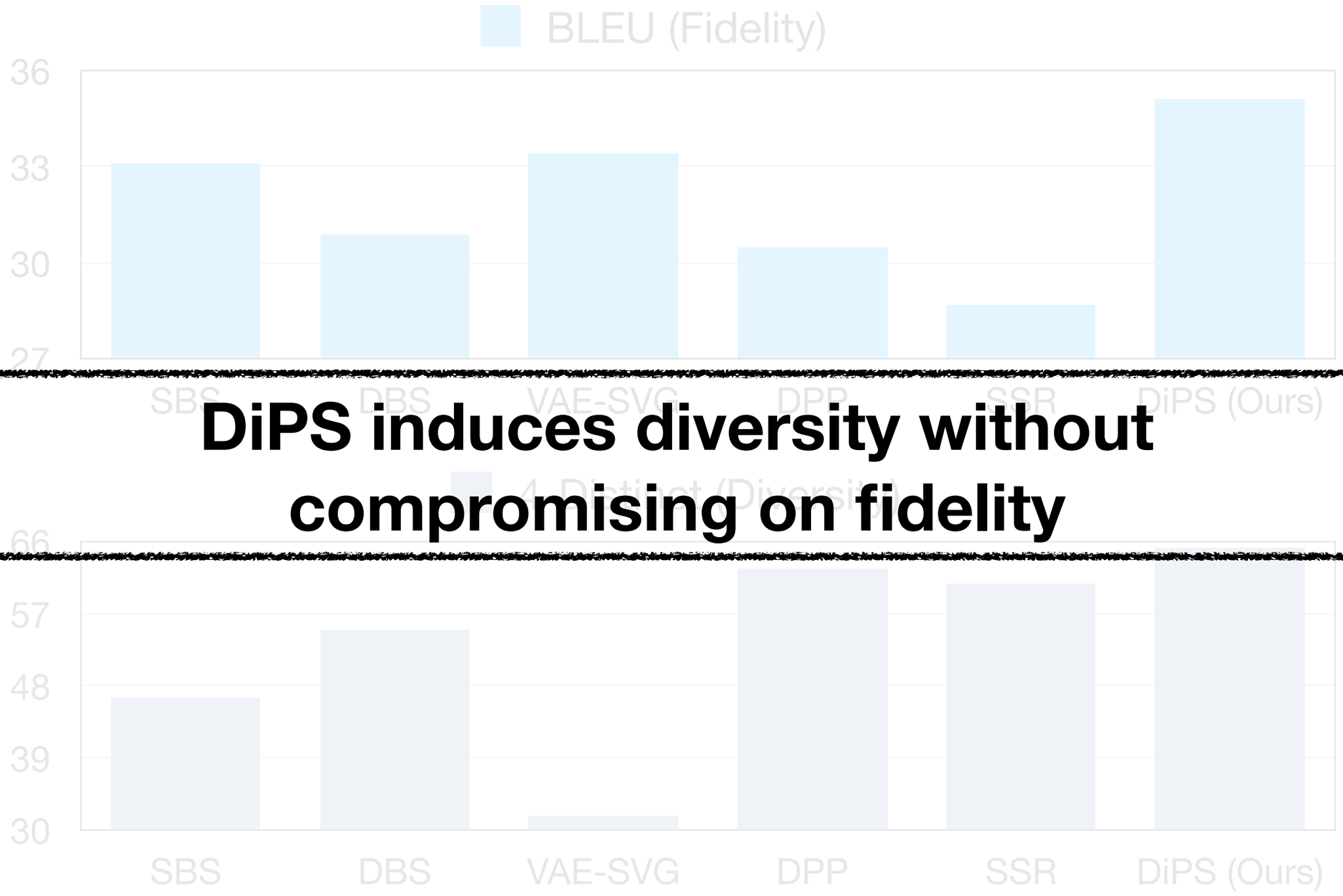
(Quora Dataset)



Models

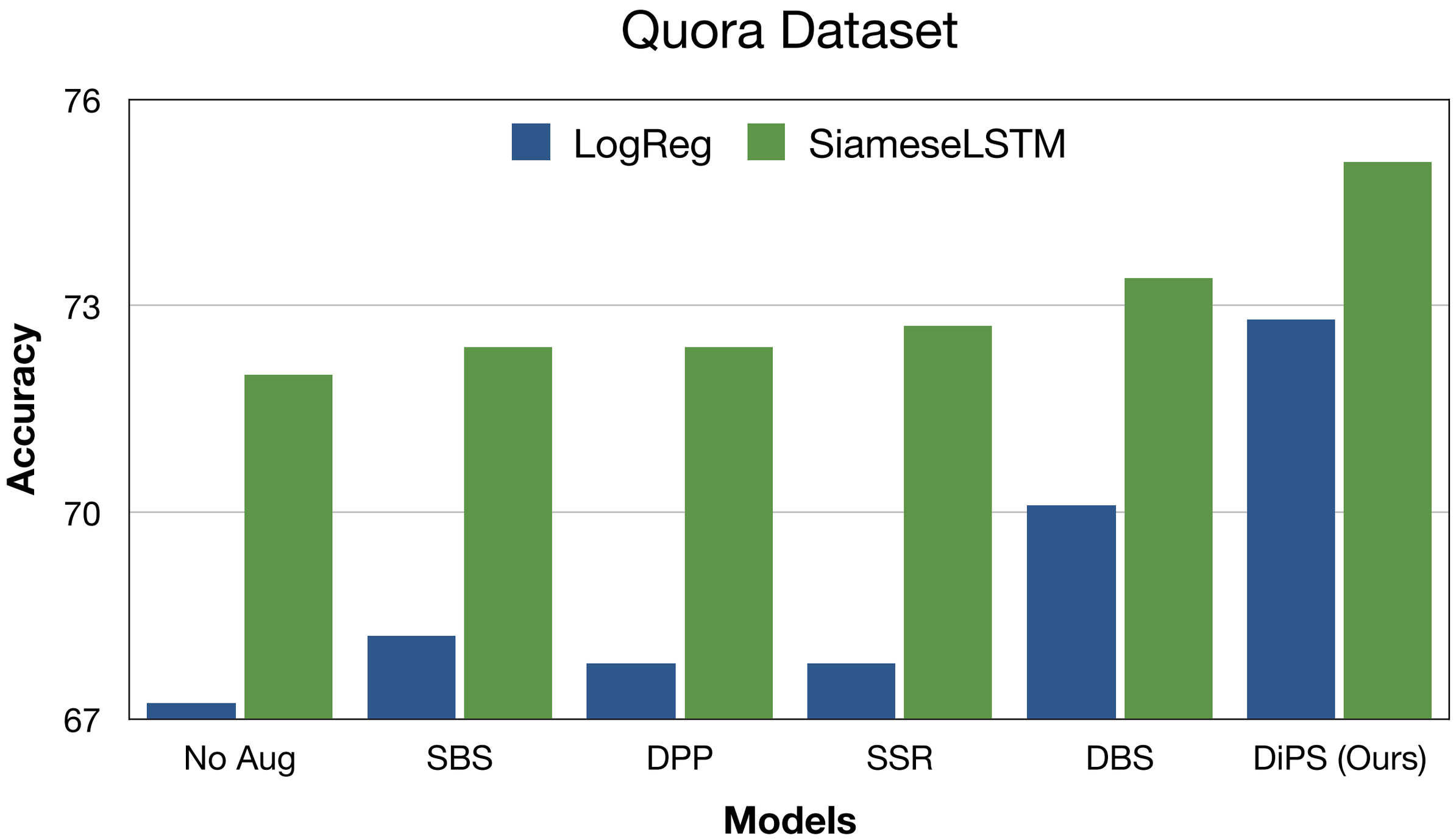
Fidelity & Diversity

(Quora Dataset)

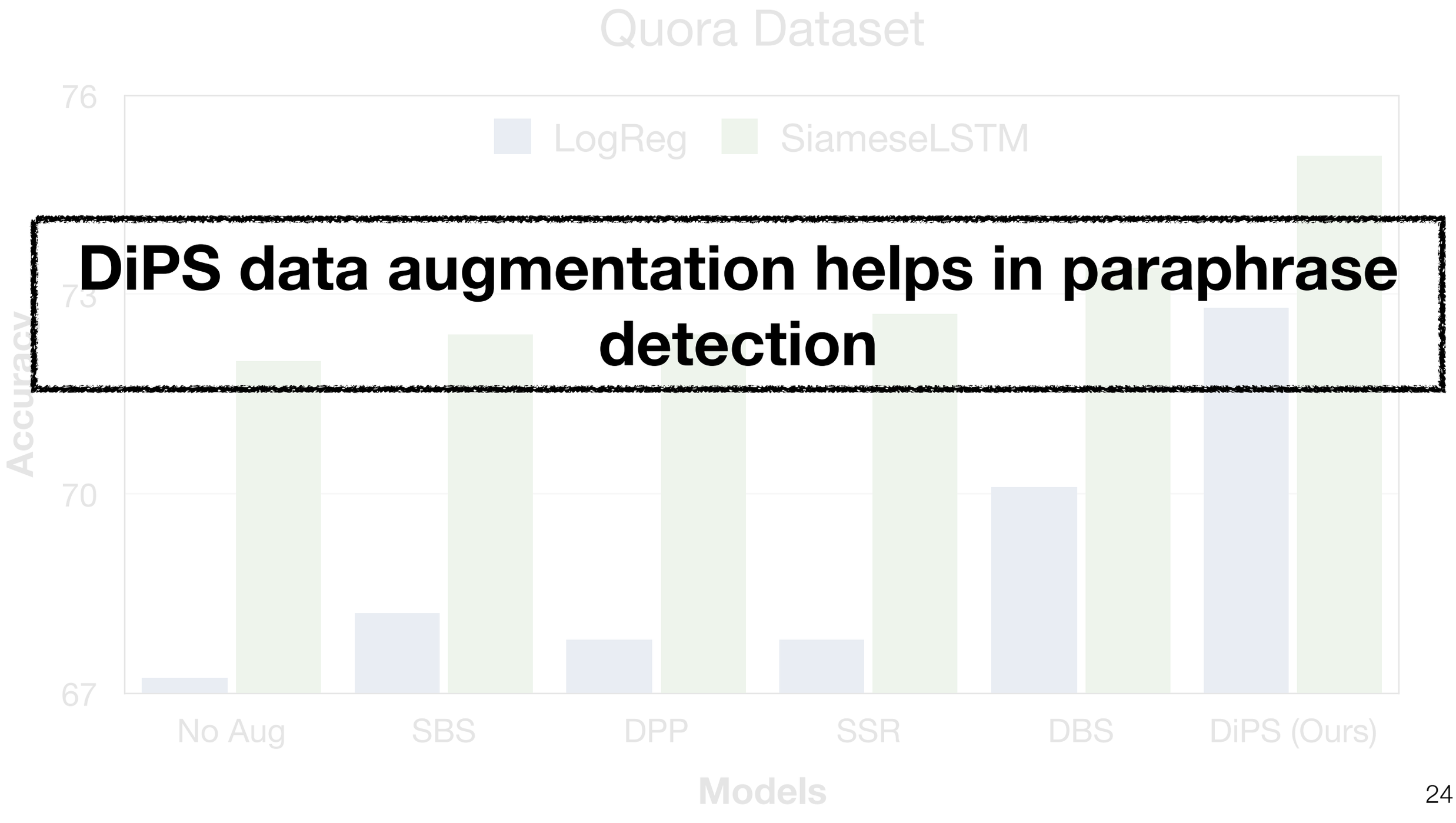


DiPS induces diversity without compromising on fidelity

Data Augmentation Paraphrase Detection

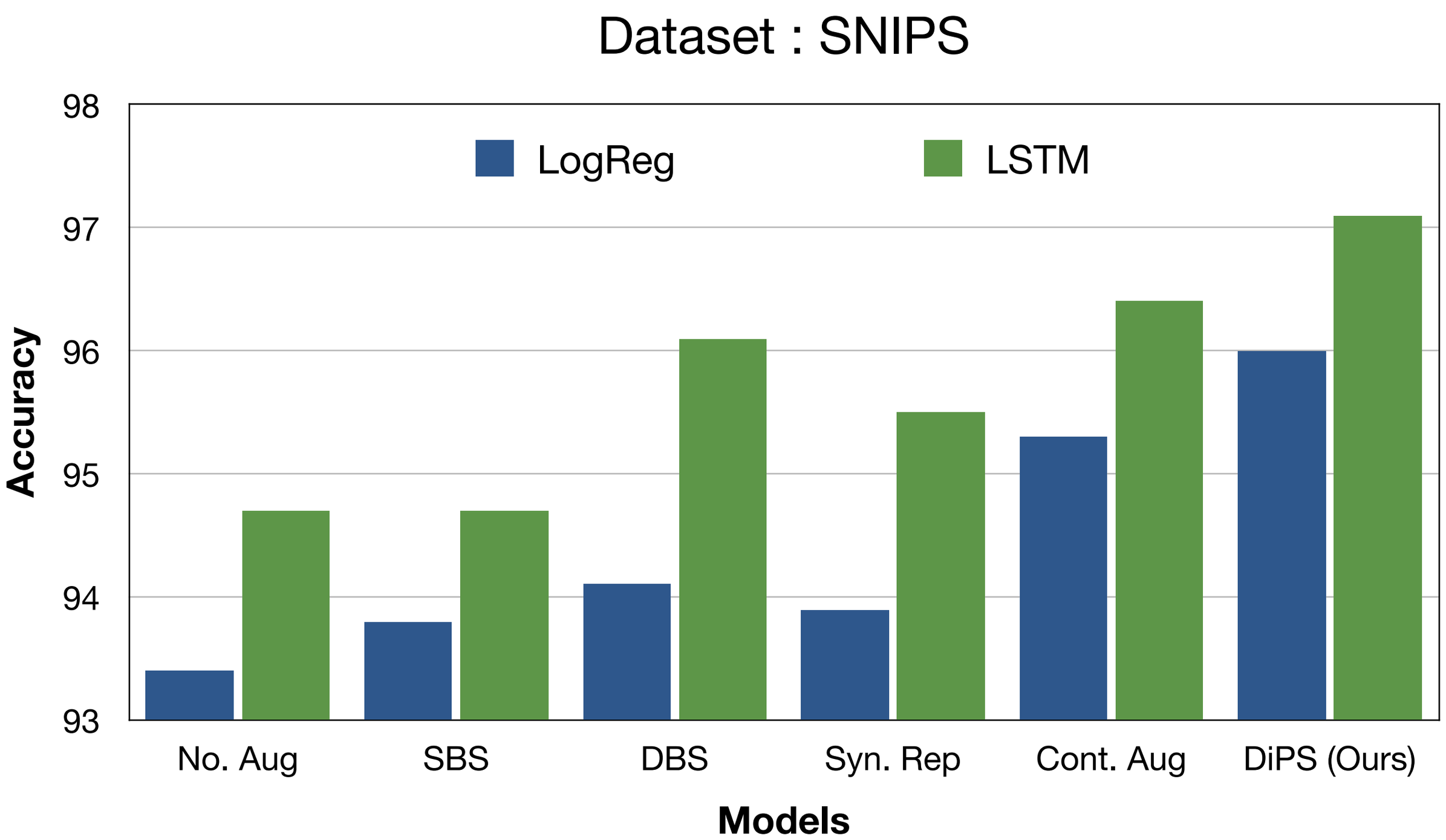


Data Augmentation Paraphrase Detection

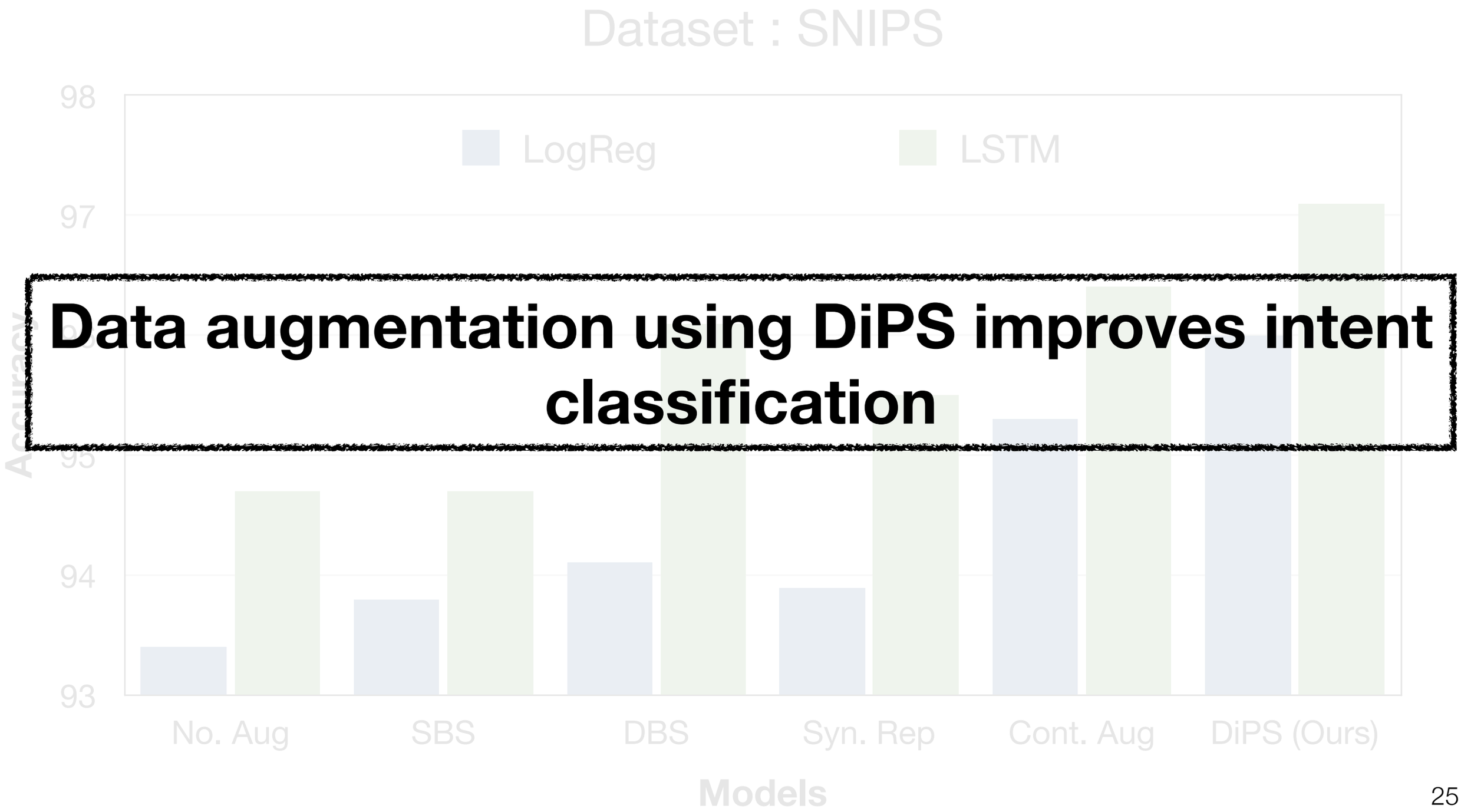


DiPS data augmentation helps in paraphrase detection

Data Augmentation for Intent Classification

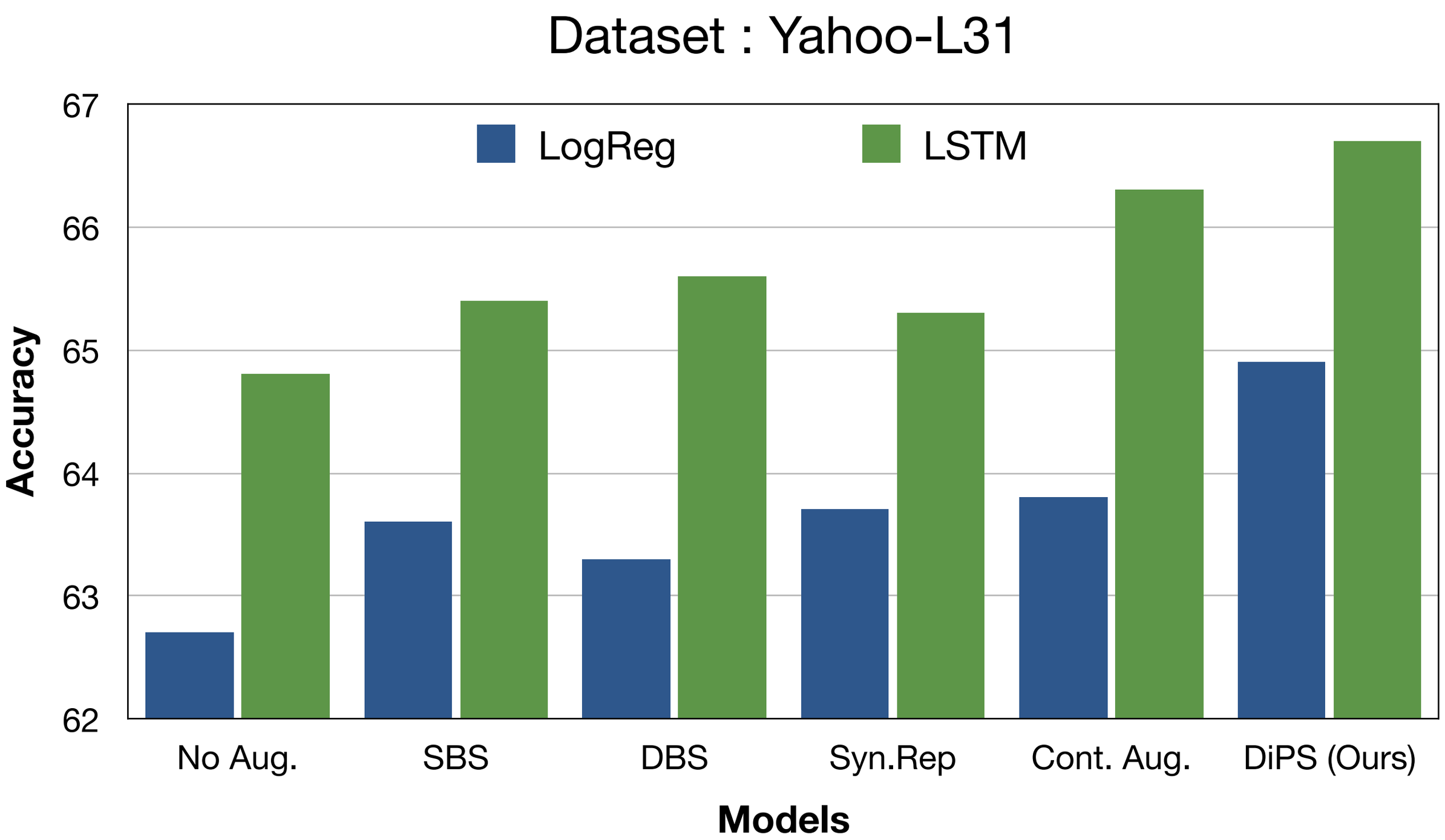


Data Augmentation for Intent Classification

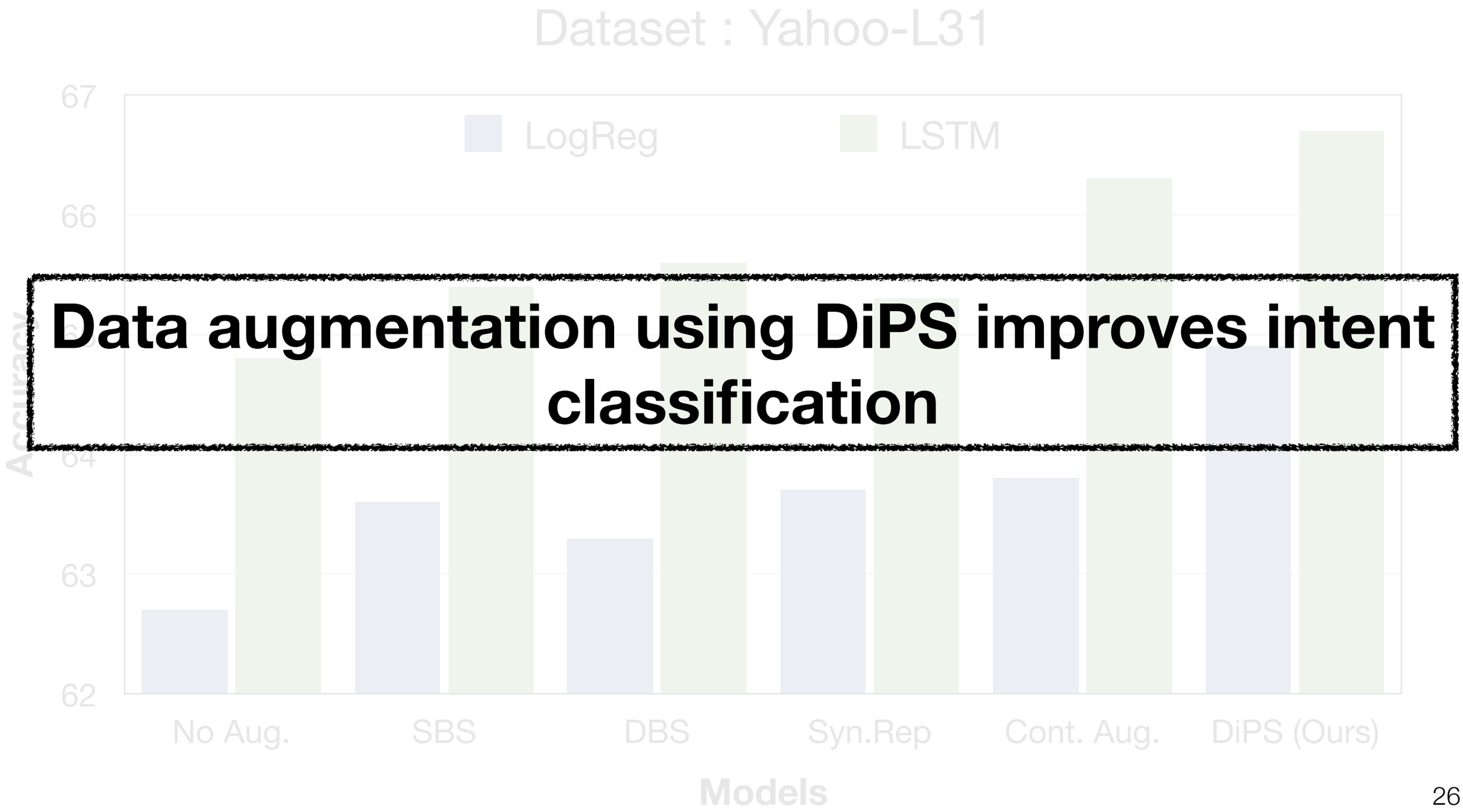


Data augmentation using DiPS improves intent classification

Data Augmentation for Intent Classification



Data Augmentation for Intent Classification



Data augmentation using DiPS improves intent classification

Diversity in Paraphrase Generation: Summary

Diversity in Paraphrase Generation: Summary

Problem

Diversity in Paraphrases

Without compromising on fidelity

Diversity in Paraphrase Generation: Summary

Problem

Diversity in Paraphrases

Without compromising on fidelity

Method

DiPS

Sub-modular optimisation

Diversity in Paraphrase Generation: Summary

Problem

Diversity in Paraphrases

Without compromising on fidelity

Method

DiPS

Sub-modular optimisation

Take-Aways

Seq2Seq + Diversity

Data Augmentation Using Paraphrasing

Syntacticality
In
Paraphrase Generation

Motivation for Syntax-Guided Paraphrasing

S1 : Because it is raining today, you should carry an umbrella

S2 : You should carry an umbrella today, because it is raining

	Fifth Graders	Adults
Preference		

Motivation for Syntax-Guided Paraphrasing

S1 : Because it is raining today, you should carry an umbrella

S2 : You should carry an umbrella today, because it is raining

	Fifth Graders	Adults
Preference	**S2**	

Motivation for Syntax-Guided Paraphrasing

S1 : Because it is raining today, you should carry an umbrella

S2 : You should carry an umbrella today, because it is raining

	Fifth Graders	Adults
Preference	**S2**	**S1**

Motivation for Syntax-Guided Paraphrasing

S1 : Because it is raining today, you should carry an umbrella

S2 : You should carry an umbrella today, because it is raining

Task : Syntax-guided Paraphrasing

	Fifth Graders	Adults
Preference	S2	S1

Syntactic Paraphrase Generation

Constraining paraphrases to conform to a given syntactic exemplar

Syntactic Paraphrase Generation

Constraining paraphrases to conform to a given syntactic exemplar

SOURCE	what are pure substances ? what are some examples ?

Syntactic Paraphrase Generation

Constraining paraphrases to conform to a given syntactic exemplar

SOURCE	what are pure substances ? what are some examples ?
EXEMPLAR	what are the characteristics of the elizabethan theatre ?

Syntactic Paraphrase Generation

Constraining paraphrases to conform to a given syntactic exemplar

SOURCE	what are pure substances ? what are some examples ?
EXEMPLAR	what are the characteristics of the elizabethan theatre ?
PARAPHRASE	what are some examples of pure substances ?

Syntactic Paraphrase Generation

Constraining paraphrases to conform to a given syntactic exemplar

SOURCE	what are pure substances ? what are some examples ?
EXEMPLAR	what are the characteristics of the elizabethan theatre ?
PARAPHRASE	what are some examples of pure substances ?

Fidelity
(Meaning preserving)

Syntactic Paraphrase Generation

Constraining paraphrases to conform to a given syntactic exemplar

SOURCE	what are pure substances ? what are some examples ?
EXEMPLAR	what are the characteristics of the elizabethan theatre ?
PARAPHRASE	what are some examples of pure substances ?

Fidelity
(Meaning preserving)

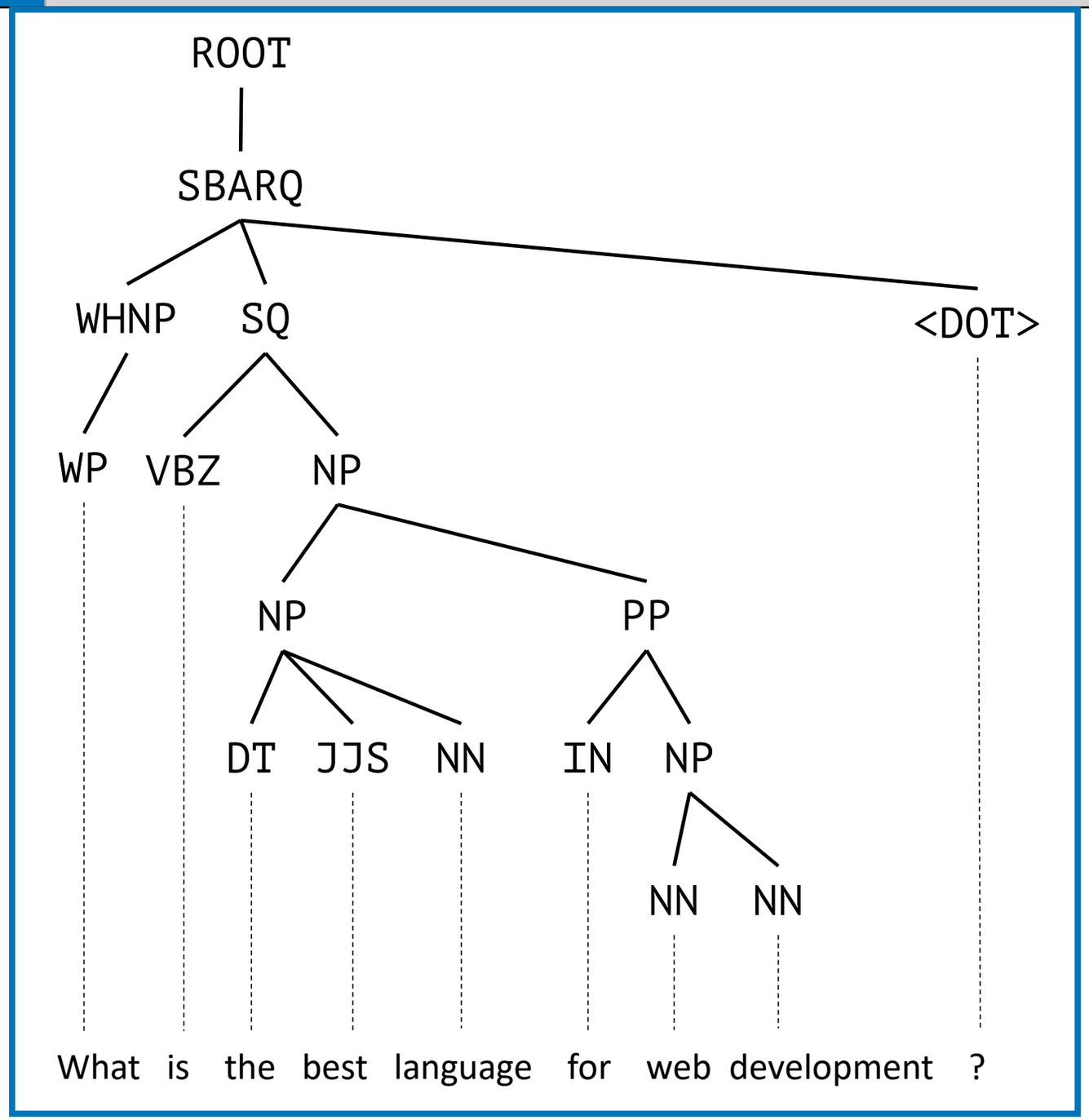
Syntacticality
(Adherence to exemplar syntax)

Constituency-based parse tree syntactic information

EXEMPLAR	what is the best language for web development ?

Constituency-based parse tree syntactic information

EXEMPLAR	what is the best language for web development ?



Utilisation of Syntactic Information

Utilisation of Syntactic Information

SOURCE	what are some of the mobile apps you can’t live without and why ?

Utilisation of Syntactic Information

SOURCE	what are some of the mobile apps you can’t live without and why ?
EXEMPLAR	what is the best language for web development ?

Utilisation of Syntactic Information

SOURCE	what are some of the mobile apps you can’t live without and why ?
EXEMPLAR	what is the best language for web development ?

	SYNTACTICAL SIGNAL	SINGLE-PASS	GRANULARITY

Utilisation of Syntactic Information

SOURCE	what are some of the mobile apps you can’t live without and why ?
EXEMPLAR	what is the best language for web development ?

	SYNTACTICAL SIGNAL	SINGLE-PASS	GRANULARITY
SCPN*	Linearized Tree		

* Adversarial Example Generation with Syntactically Controlled Paraphrase Networks, Iyyer et. al. 2018

Utilisation of Syntactic Information

SOURCE	what are some of the mobile apps you can’t live without and why ?
EXEMPLAR	what is the best language for web development ?

	SYNTACTICAL SIGNAL	SINGLE-PASS	GRANULARITY
SCPN*	Linearized Tree	✗	

* Adversarial Example Generation with Syntactically Controlled Paraphrase Networks, Iyyer et. al. 2018

Utilisation of Syntactic Information

SOURCE	what are some of the mobile apps you can’t live without and why ?
EXEMPLAR	what is the best language for web development ?

	SYNTACTICAL SIGNAL	SINGLE-PASS	GRANULARITY
SCPN*	Linearized Tree	✗	✓

* Adversarial Example Generation with Syntactically Controlled Paraphrase Networks, Iyyer et. al. 2018

Utilisation of Syntactic Information

SOURCE	what are some of the mobile apps you can't live without and why ?
EXEMPLAR	what is the best language for web development ?

	SYNTACTICAL SIGNAL	SINGLE-PASS	GRANULARITY
SCPN*	Linearized Tree	✗	✓
CGEN**	POS Tags (During Training)		

* Adversarial Example Generation with Syntactically Controlled Paraphrase Networks, Iyyer et. al. 2018

** Controllable Paraphrase Generation with a Syntactic Exemplar, Chen et. al. 2019

Utilisation of Syntactic Information

SOURCE	what are some of the mobile apps you can't live without and why ?
EXEMPLAR	what is the best language for web development ?

	SYNTACTICAL SIGNAL	SINGLE-PASS	GRANULARITY
SCPN*	Linearized Tree	✗	✓
CGEN**	POS Tags (During Training)	✓	

* Adversarial Example Generation with Syntactically Controlled Paraphrase Networks, Iyyer et. al. 2018

** Controllable Paraphrase Generation with a Syntactic Exemplar, Chen et. al. 2019

Utilisation of Syntactic Information

SOURCE	what are some of the mobile apps you can't live without and why ?
EXEMPLAR	what is the best language for web development ?

	SYNTACTICAL SIGNAL	SINGLE-PASS	GRANULARITY
SCPN*	Linearized Tree	✗	✓
CGEN**	POS Tags (During Training)	✓	✗

* Adversarial Example Generation with Syntactically Controlled Paraphrase Networks, Iyyer et. al. 2018

** Controllable Paraphrase Generation with a Syntactic Exemplar, Chen et. al. 2019

Utilisation of Syntactic Information

SOURCE	what are some of the mobile apps you can't live without and why ?
EXEMPLAR	what is the best language for web development ?

	SYNTACTICAL SIGNAL	SINGLE-PASS	GRANULARITY
SCPN*	Linearized Tree	✗	✓
CGEN**	POS Tags (During Training)	✓	✗
SGCP	Constituency Parse Tree		

* Adversarial Example Generation with Syntactically Controlled Paraphrase Networks, Iyyer et. al. 2018

** Controllable Paraphrase Generation with a Syntactic Exemplar, Chen et. al. 2019

Utilisation of Syntactic Information

SOURCE	what are some of the mobile apps you can't live without and why ?
EXEMPLAR	what is the best language for web development ?

	SYNTACTICAL SIGNAL	SINGLE-PASS	GRANULARITY
SCPN*	Linearized Tree	✗	✓
CGEN**	POS Tags (During Training)	✓	✗
SGCP	Constituency Parse Tree	✓	

* Adversarial Example Generation with Syntactically Controlled Paraphrase Networks, Iyyer et. al. 2018

** Controllable Paraphrase Generation with a Syntactic Exemplar, Chen et. al. 2019

Utilisation of Syntactic Information

SOURCE	what are some of the mobile apps you can't live without and why ?
EXEMPLAR	what is the best language for web development ?

	SYNTACTICAL SIGNAL	SINGLE-PASS	GRANULARITY
SCPN*	Linearized Tree	✗	✓
CGEN**	POS Tags (During Training)	✓	✗
SGCP	Constituency Parse Tree	✓	✓

* Adversarial Example Generation with Syntactically Controlled Paraphrase Networks, Iyyer et. al. 2018

** Controllable Paraphrase Generation with a Syntactic Exemplar, Chen et. al. 2019

Qualitative Results based on Syntactic Signals

SOURCE	what are some of the mobile apps you can’t live without and why ?
EXEMPLAR	what is the best language for web development ?

Qualitative Results based on Syntactic Signals

SOURCE	what are some of the mobile apps you can’t live without and why ?
EXEMPLAR	what is the best language for web development ?
SCPN*	what are the best ways to lose weight ?

* Adversarial Example Generation with Syntactically Controlled Paraphrase Networks, Iyyer et. al. 2018

Qualitative Results based on Syntactic Signals

SOURCE	what are some of the mobile apps you can’t live without and why ?
EXEMPLAR	what is the best language for web development ?
SCPN*	what are the best ways to lose weight ?
CGEN**	which is the best mobile app you can’t ?

* Adversarial Example Generation with Syntactically Controlled Paraphrase Networks, Iyyer et. al. 2018

** Controllable Paraphrase Generation with a Syntactic Exemplar, Chen et. al. 2019

Qualitative Results based on Syntactic Signals

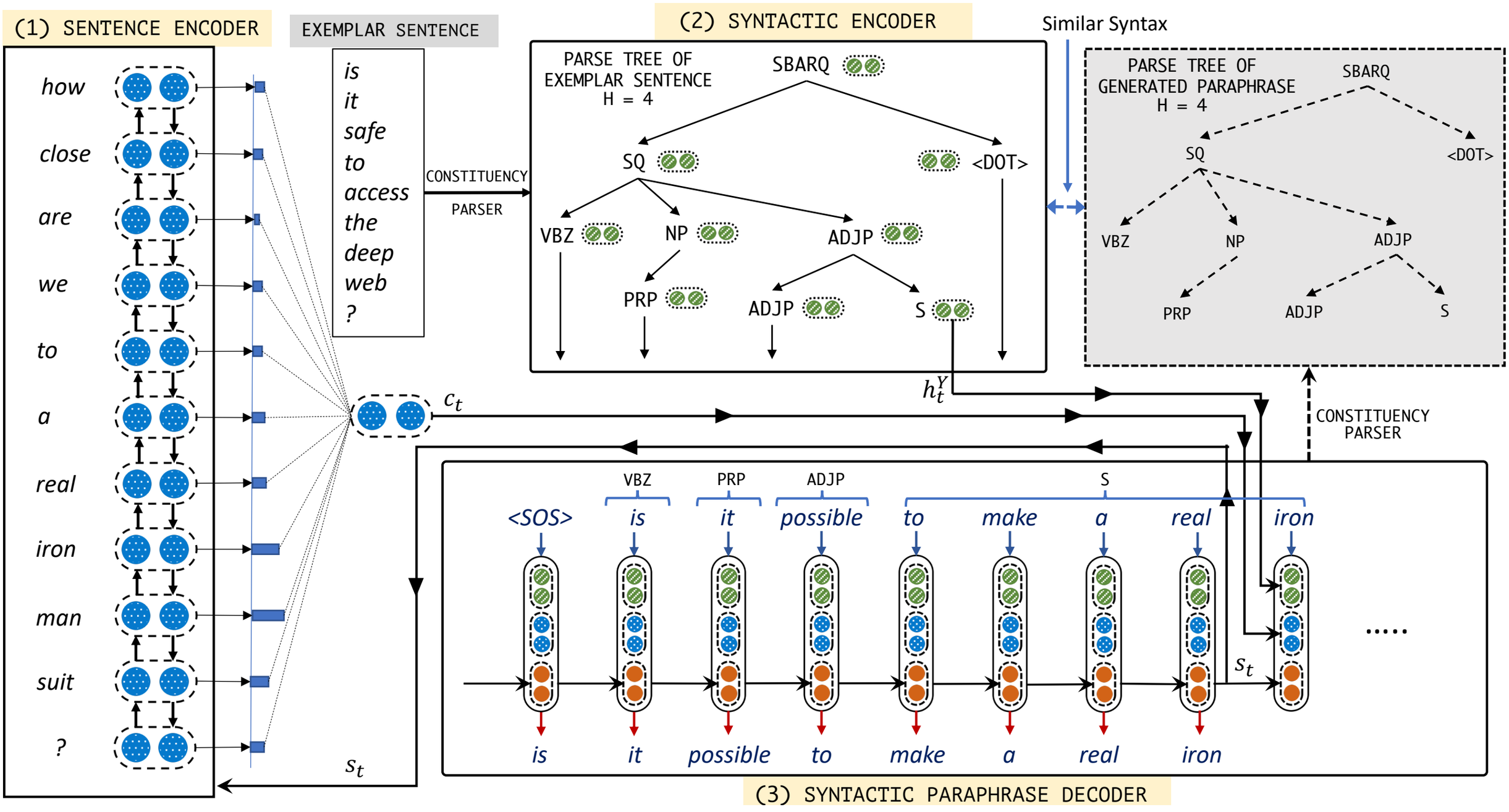
SOURCE	what are some of the mobile apps you can’t live without and why ?
EXEMPLAR	what is the best language for web development ?
SCPN*	what are the best ways to lose weight ?
CGEN**	which is the best mobile app you can’t ?
SGCP (Ours)	which is the best app you can’t live without and why ?

* Adversarial Example Generation with Syntactically Controlled Paraphrase Networks, Iyyer et. al. 2018

** Controllable Paraphrase Generation with a Syntactic Exemplar, Chen et. al. 2019

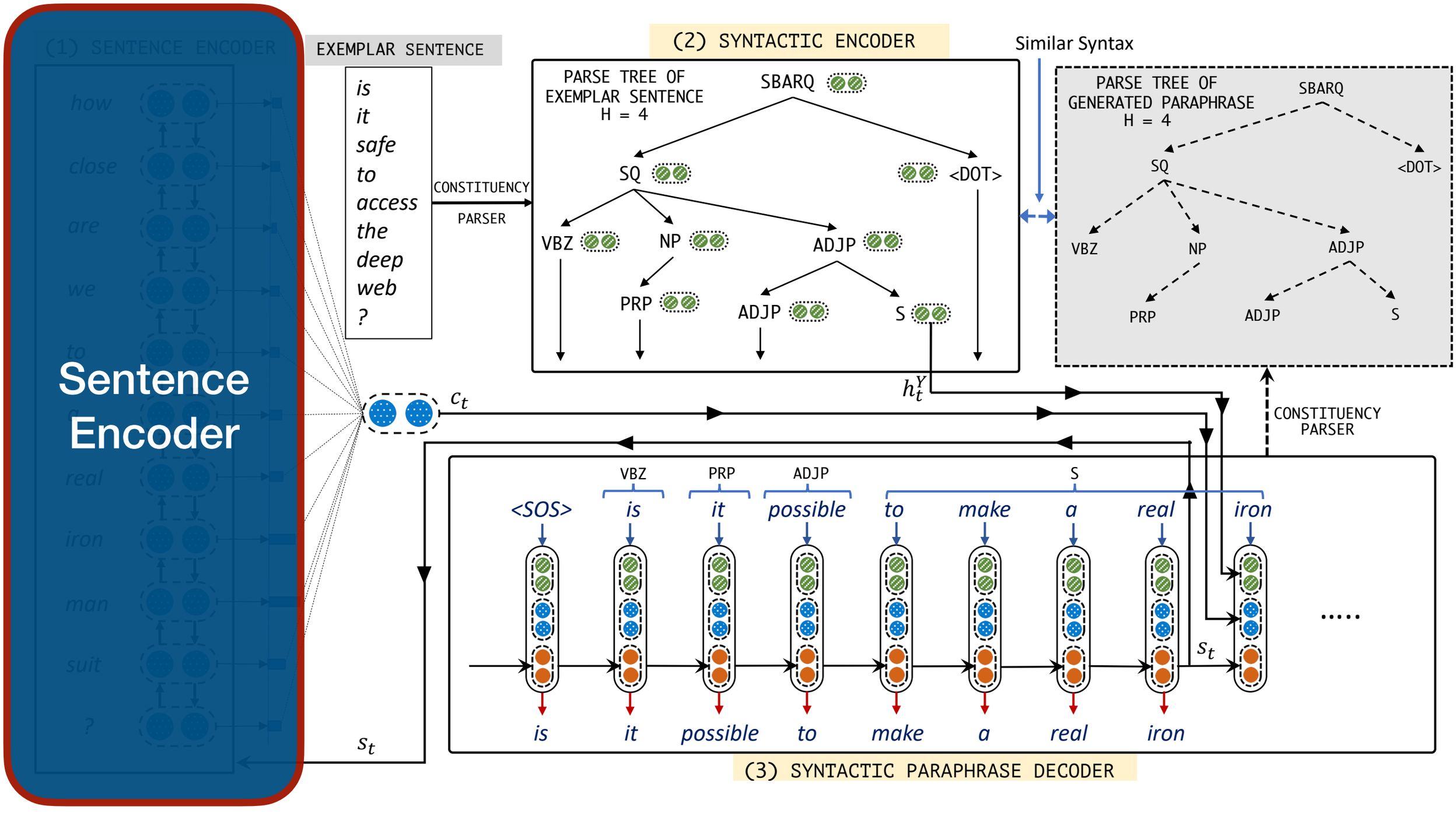
SGCP: Syntax Guided Controlled Paraphraser

Use Syntactic Tree Structure to Guide Paraphrase Generation model



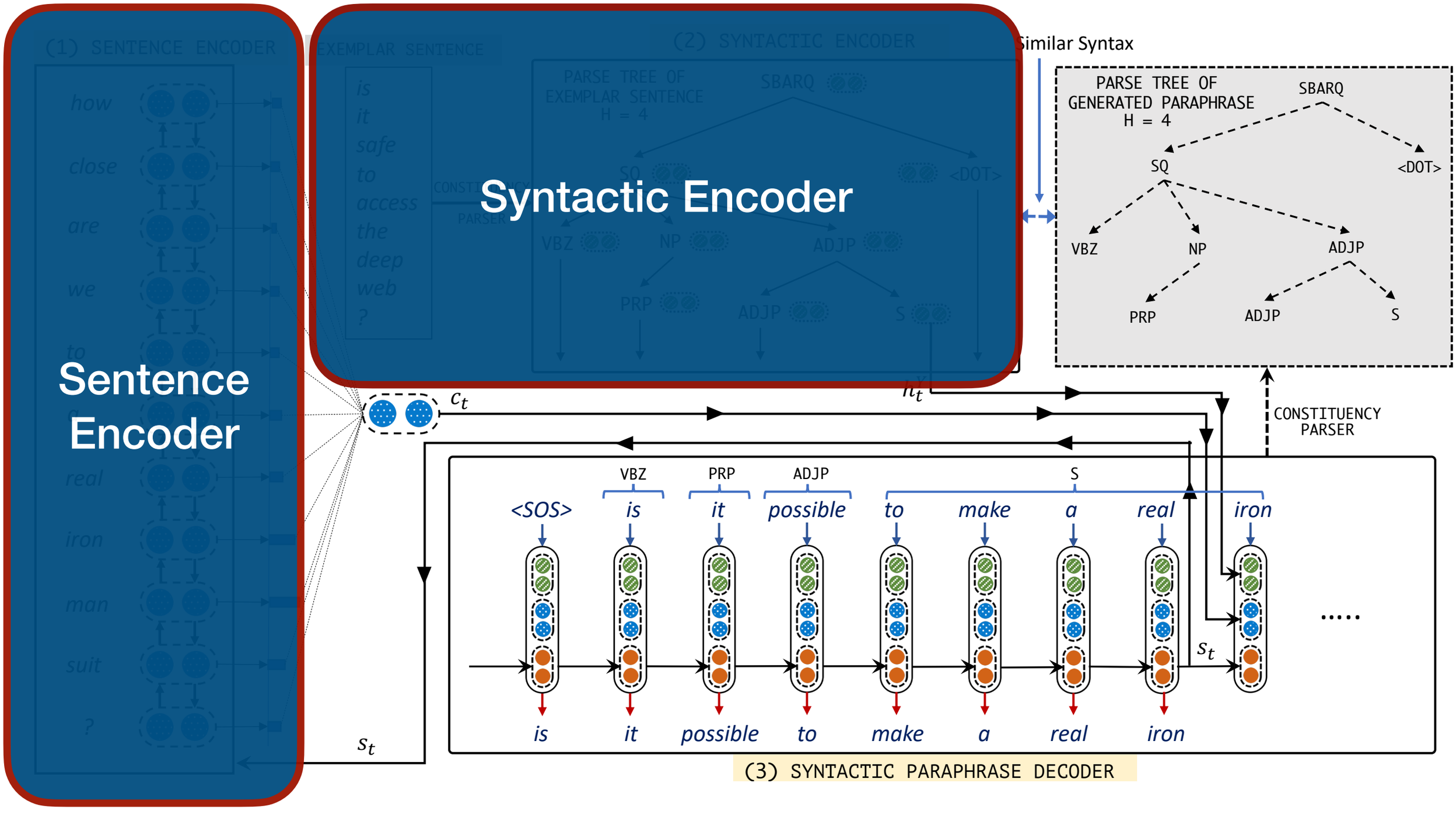
SGCP: Syntax Guided Controlled Paraphraser

Use Syntactic Tree Structure to Guide Paraphrase Generation model



SGCP: Syntax Guided Controlled Paraphraser

Use Syntactic Tree Structure to Guide Paraphrase Generation model



SGCP: Syntax Guided Controlled Paraphraser

Use Syntactic Tree Structure to Guide Paraphrase Generation model

Sentence Encoder

Syntactic Encoder

Decoder

Similar Syntax

PARSE TREE OF GENERATED PARAPHRASE H = 4

SBARQ

SQ

<DOT>

VBZ

NP

ADJP

PRP

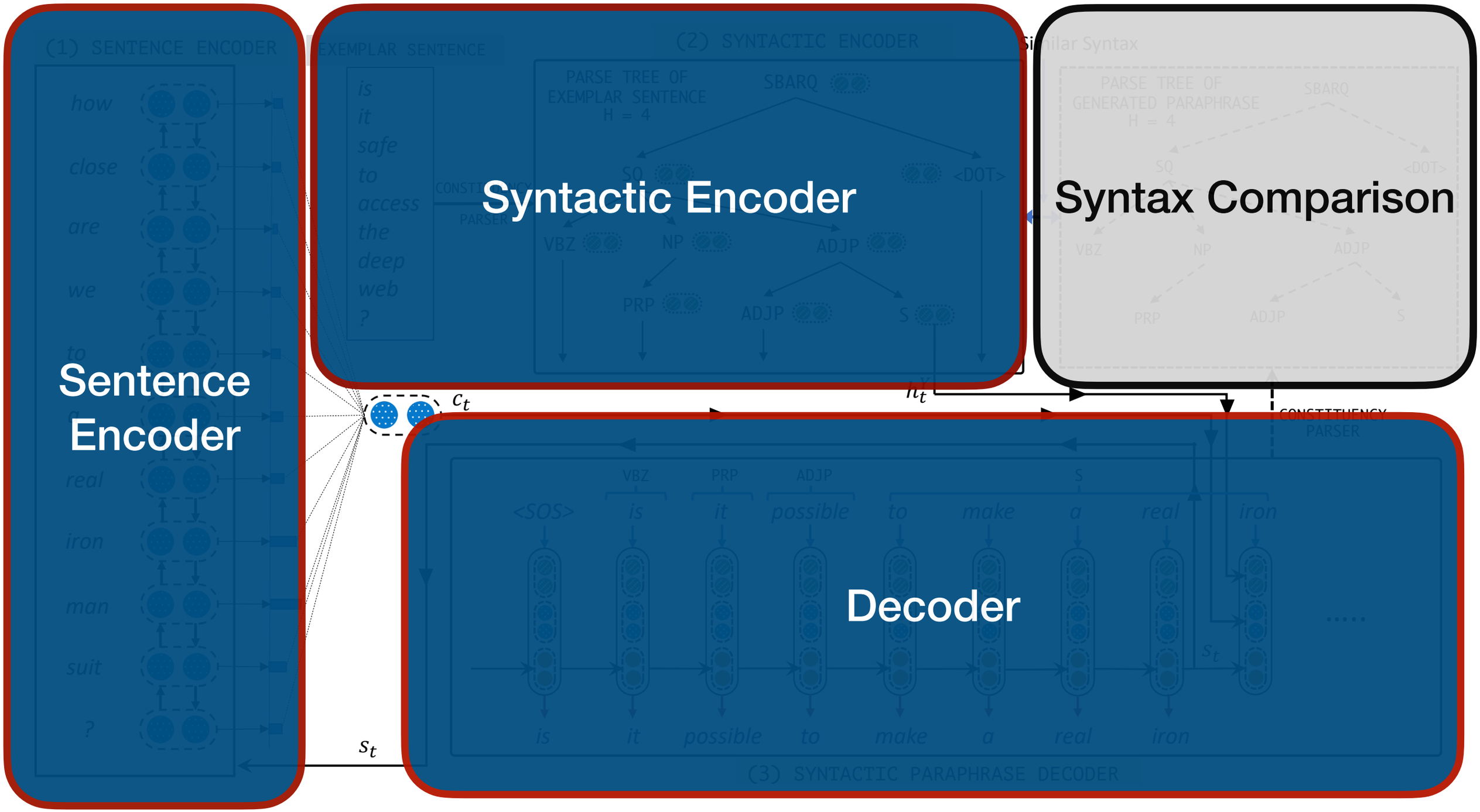
ADJP

S

CONSTITUENCY PARSER

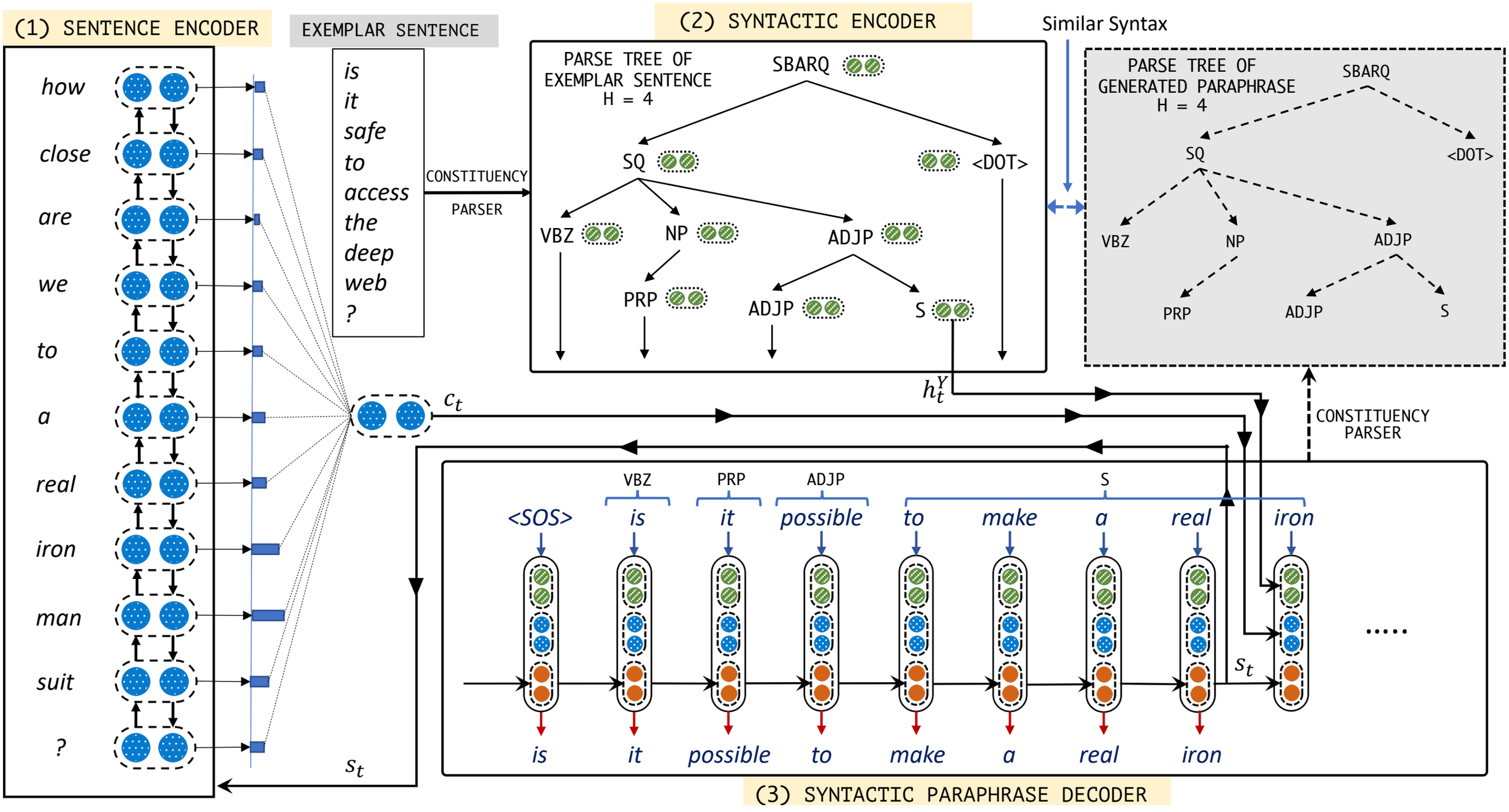
SGCP: Syntax Guided Controlled Paraphraser

Use Syntactic Tree Structure to Guide Paraphrase Generation model



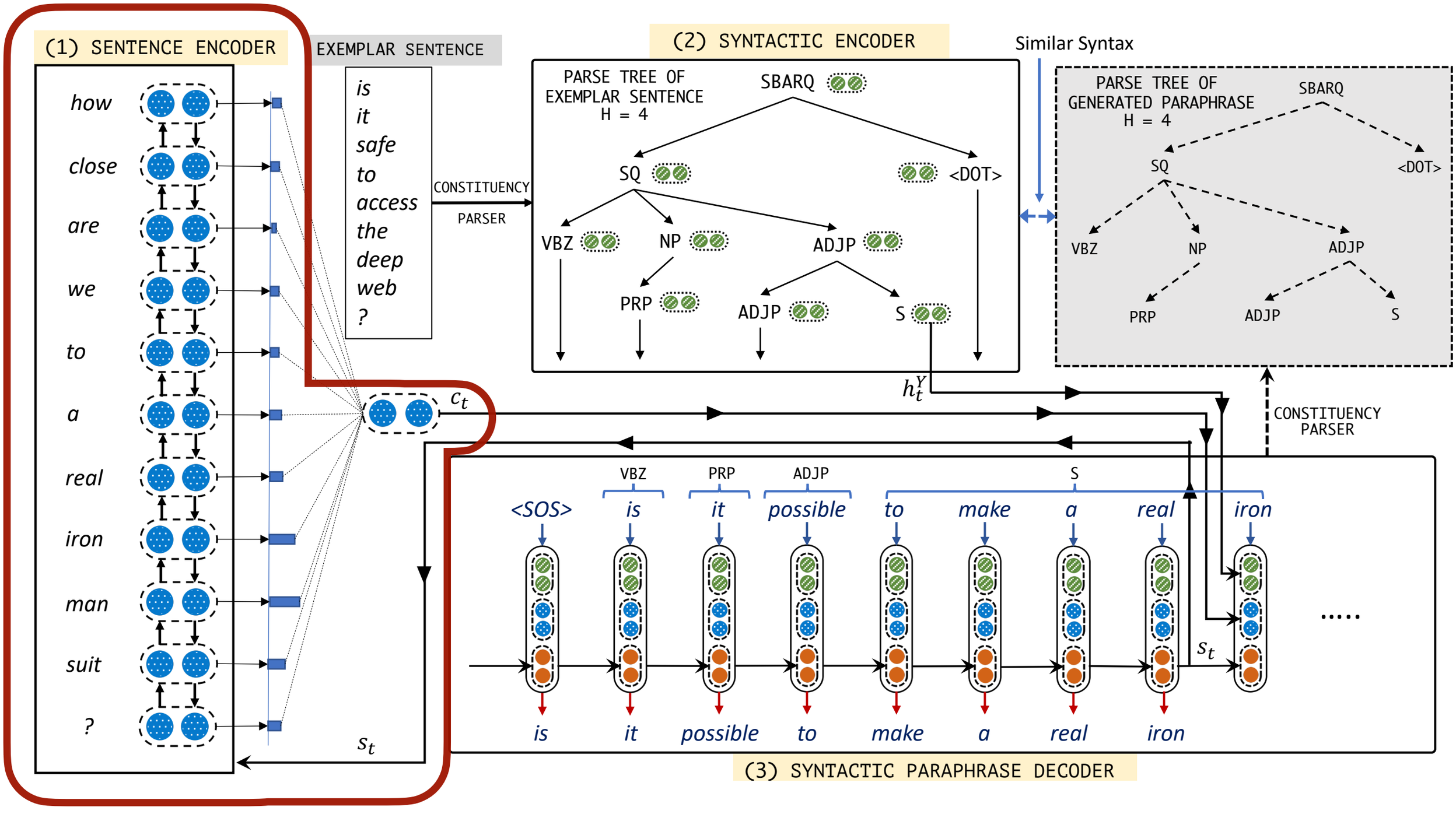
SGCP: Syntax Guided Controlled Paraphraser

Use Syntactic Tree Structure to Guide Paraphrase Generation model



SGCP: Syntax Guided Controlled Paraphraser

Use Syntactic Tree Structure to Guide Paraphrase Generation model



SGCP : Sentence Encoder

(1) SENTENCE ENCODER

how

close

are

we

to

a

real

iron

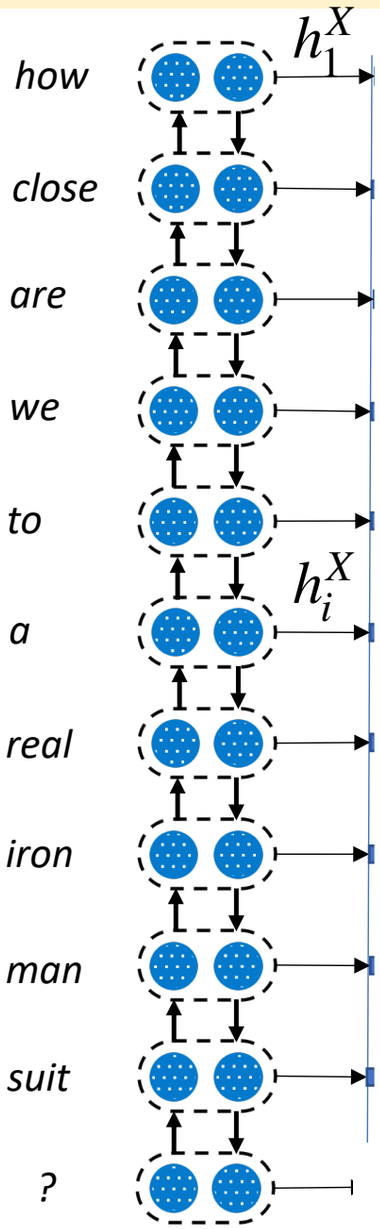
man

suit

?

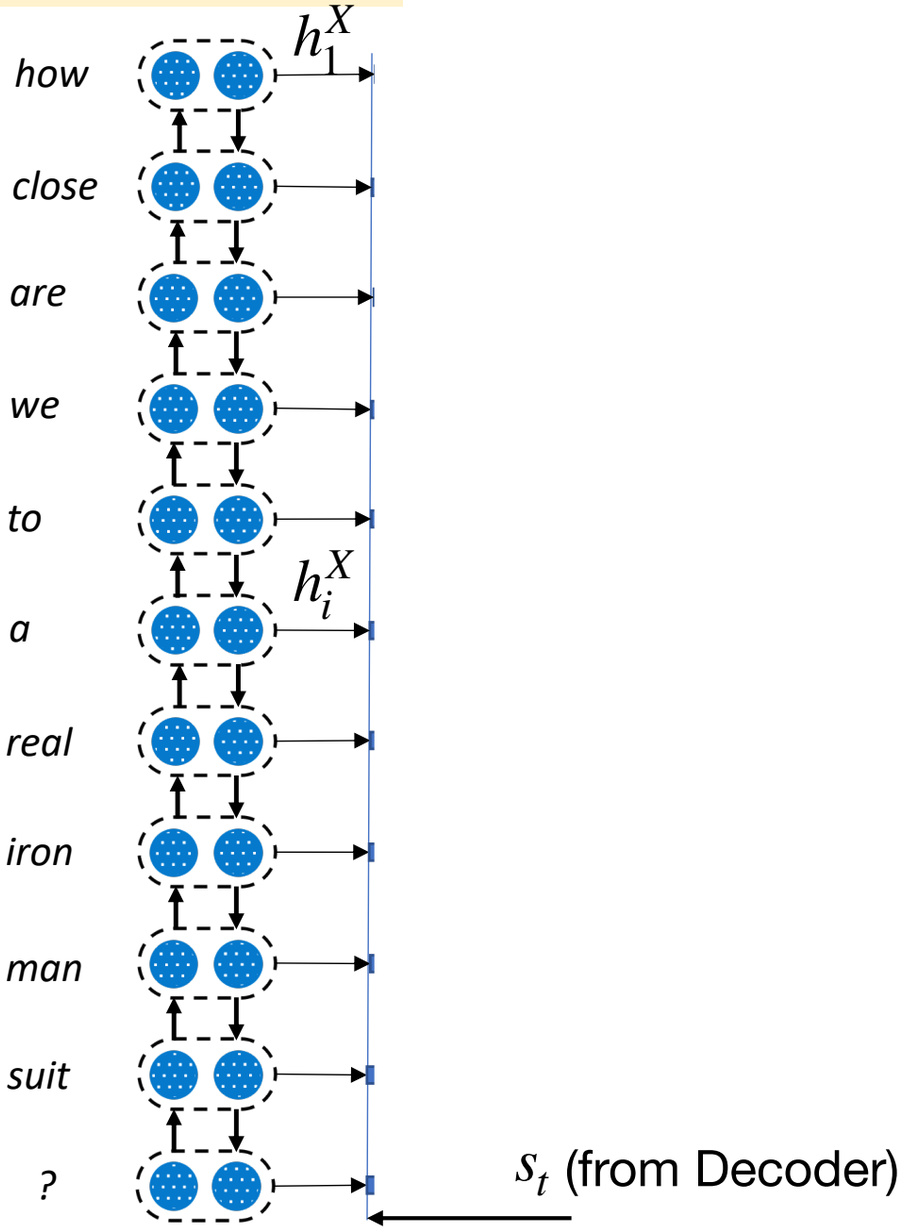
SGCP : Sentence Encoder

(1) SENTENCE ENCODER

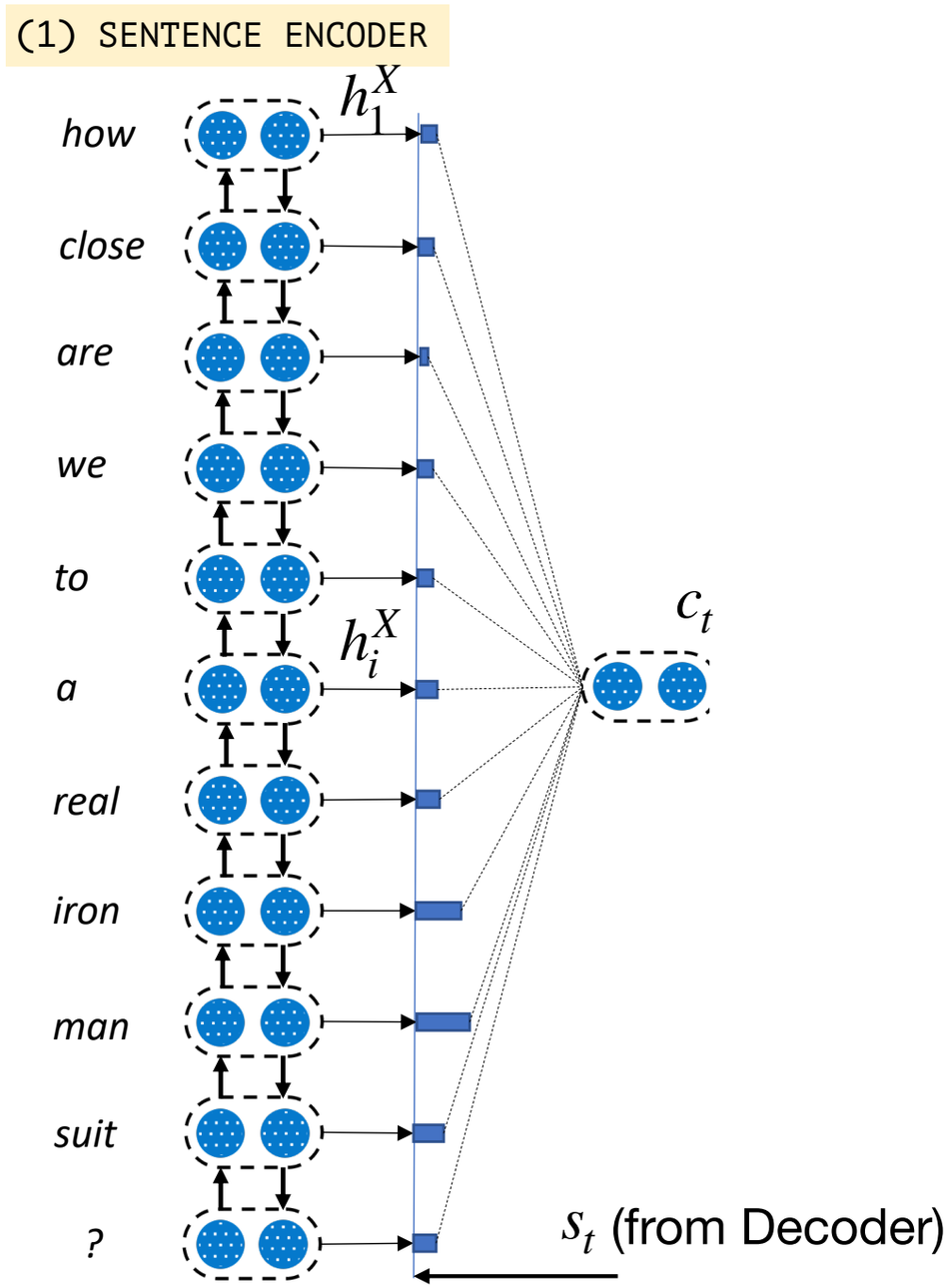


SGCP : Sentence Encoder

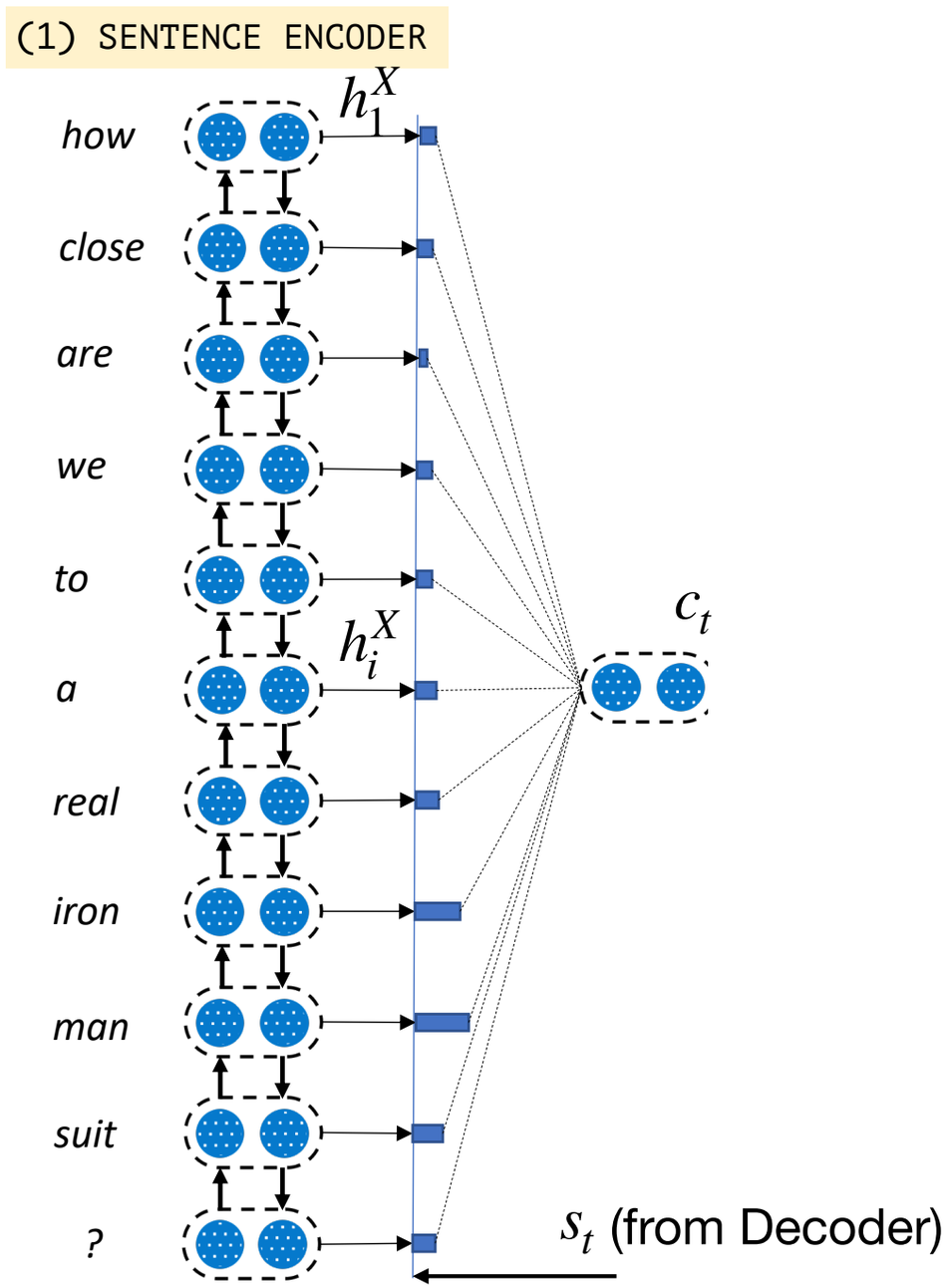
(1) SENTENCE ENCODER



SGCP : Sentence Encoder



SGCP : Sentence Encoder



$$h_i^X = \textbf{GRU}(h_{i-1}^X, e(x_i))$$

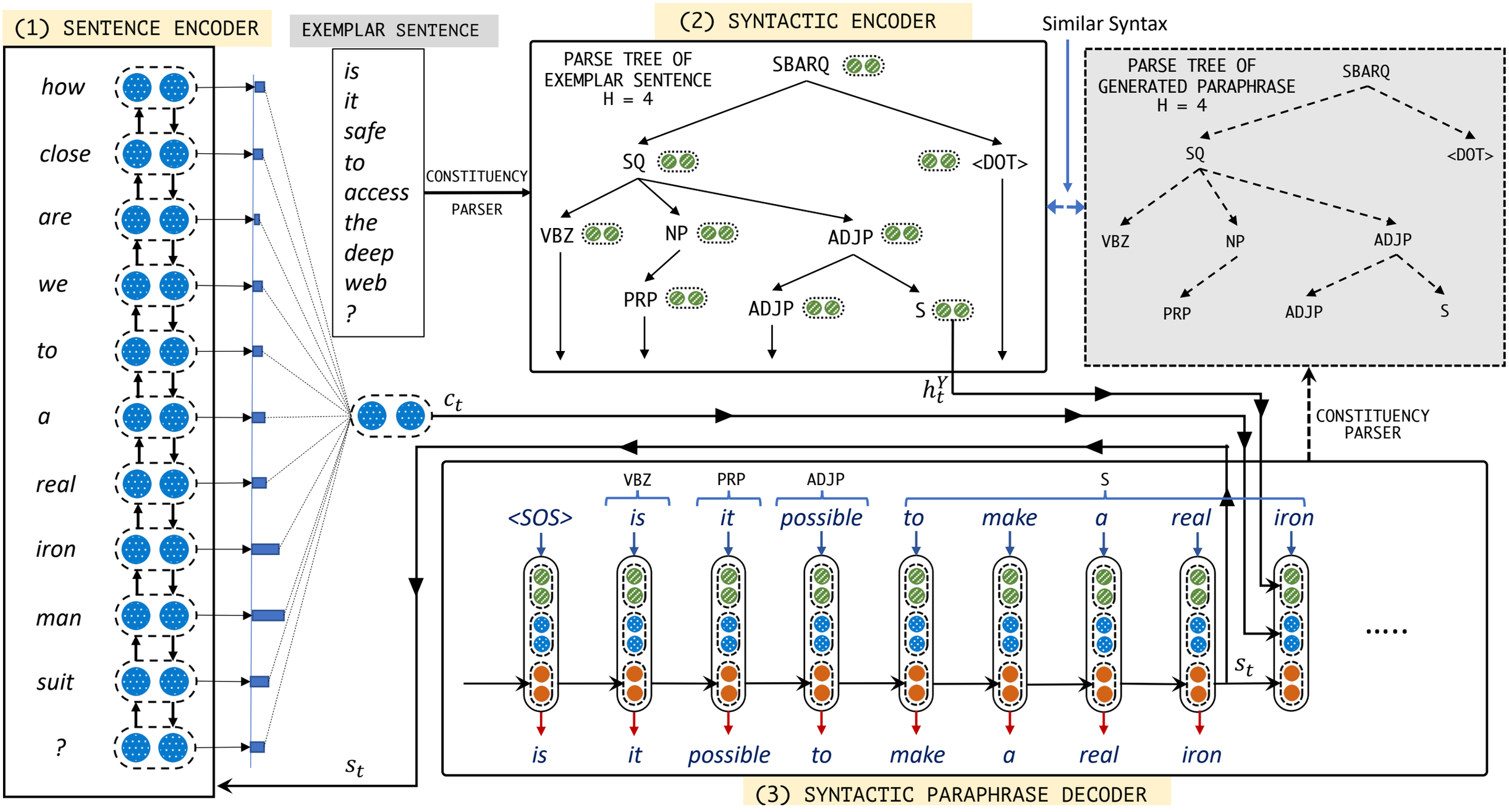
$$e_i^t = v^\intercal \textbf{tanh}(W_h h_i^X + W_s s_t + b_{attn})$$

$$\alpha^t = \textbf{softmax}(e^t)$$

$$c_t = \sum_i \alpha_i^t h_i^X$$

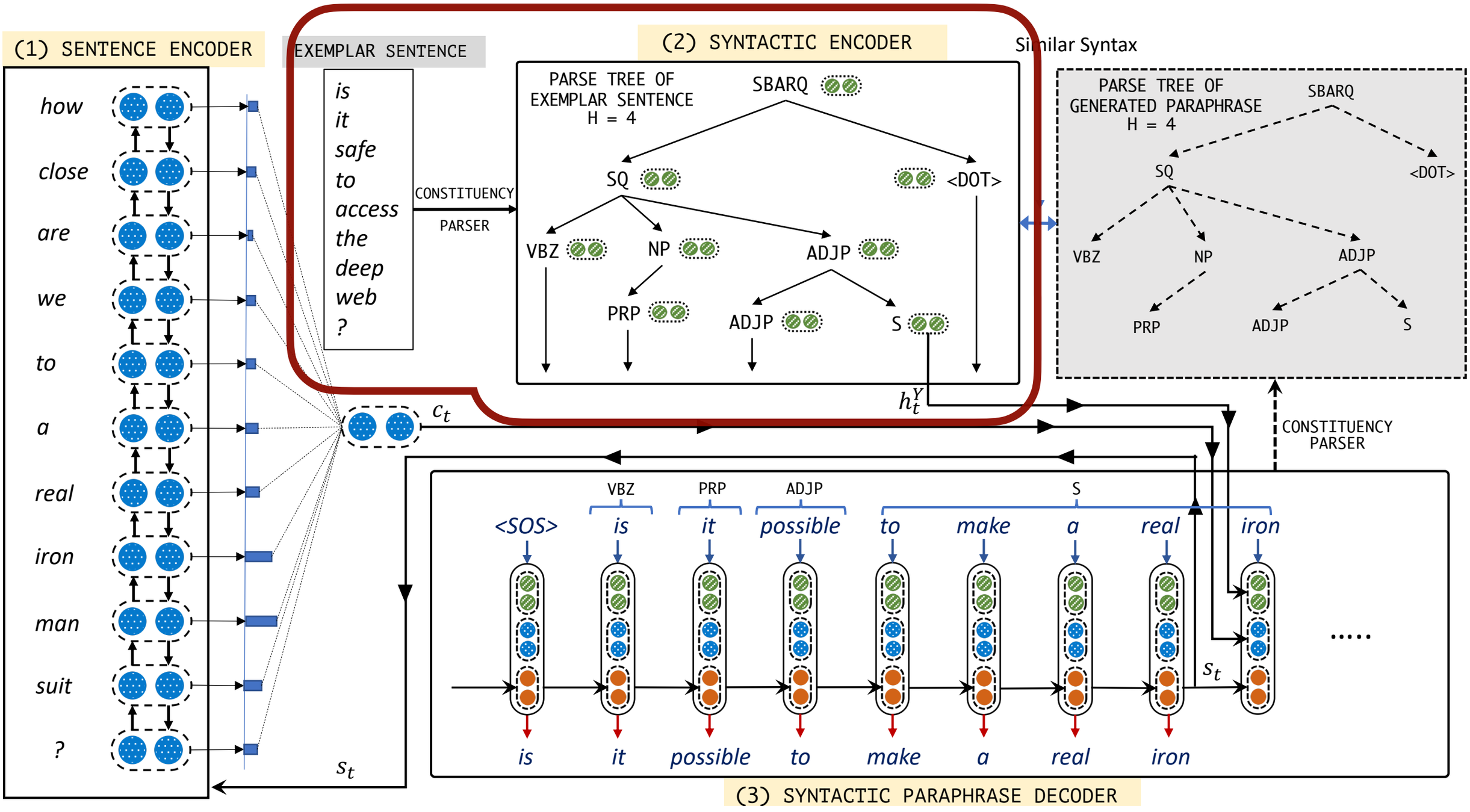
SGCP: Syntax Guided Controlled Paraphraser

Use Syntactic Tree Structure to Guide Paraphrase Generation model

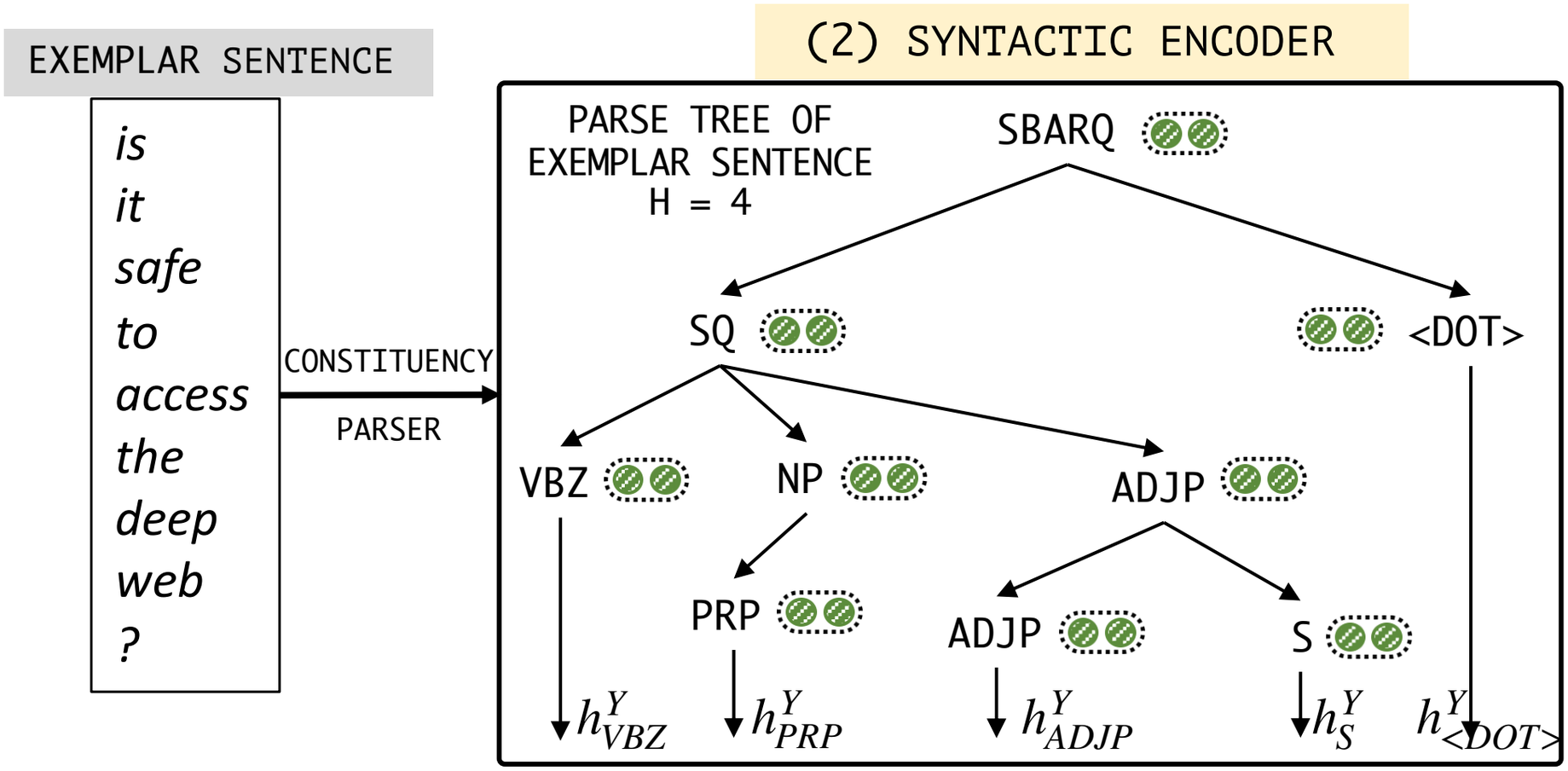


SGCP: Syntax Guided Controlled Paraphraser

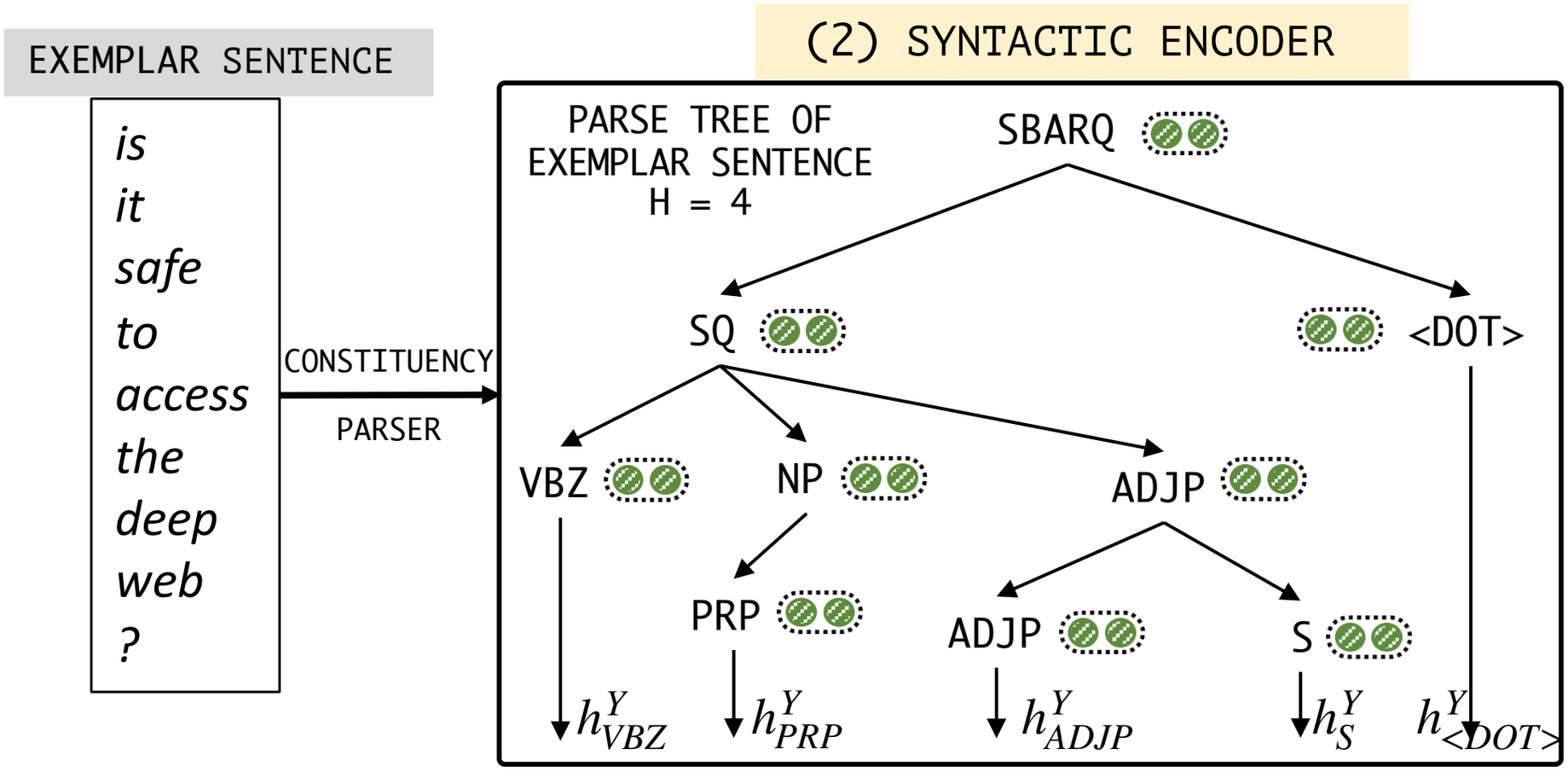
Use Syntactic Tree Structure to Guide Paraphrase Generation model



SGCP: Syntactic Encoder



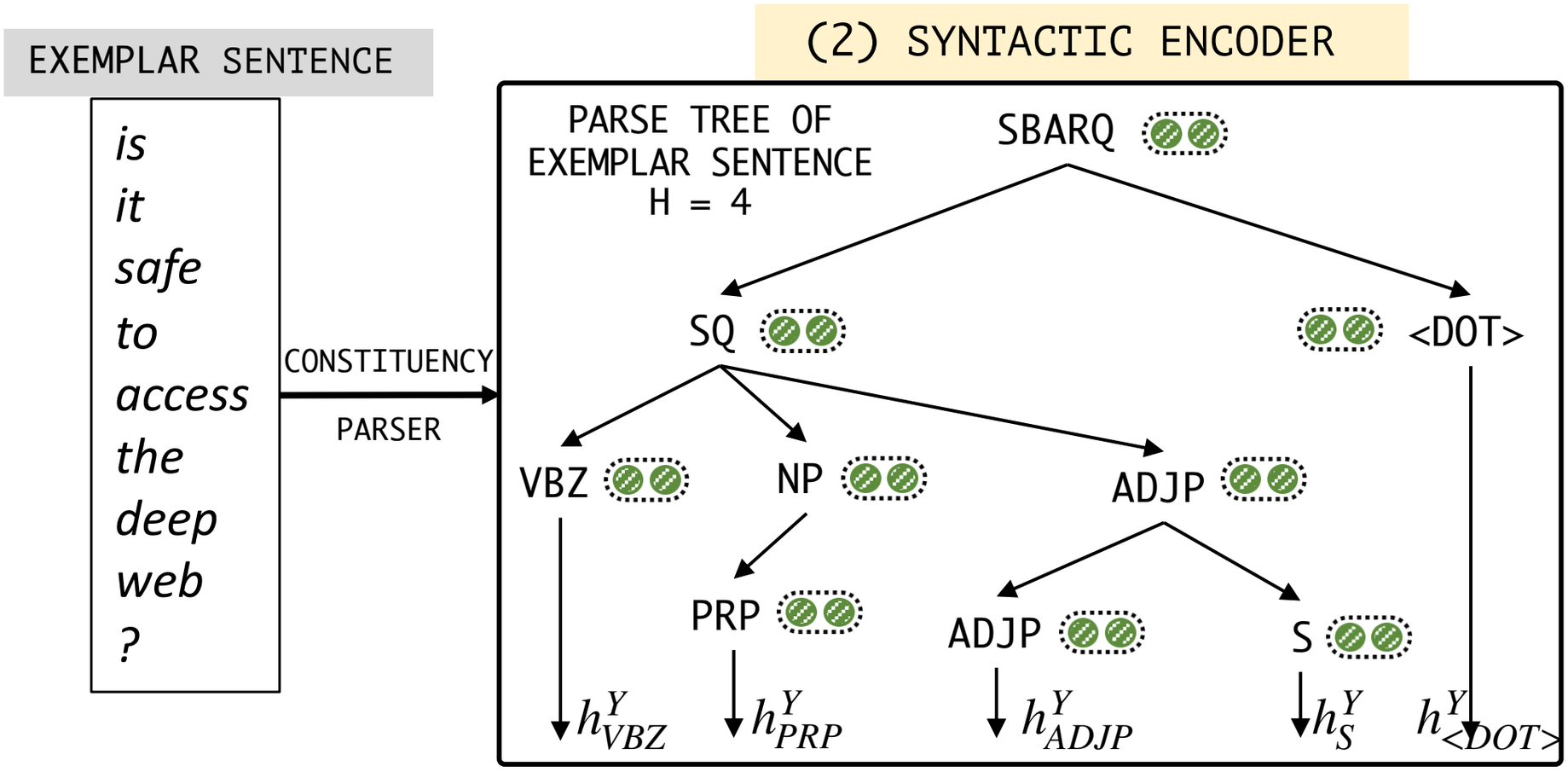
SGCP: Syntactic Encoder



$$h_v^Y = \textbf{GeLU}(W_{pa} h_{pa(v)}^Y + W_v e(y_v) + b_v)$$

$$\mathbb{L}_H^Y = [h_{VBZ}^Y, h_{PRP}^Y, h_{ADJP}^Y, h_S^Y, h_{<DOT>}^Y]$$

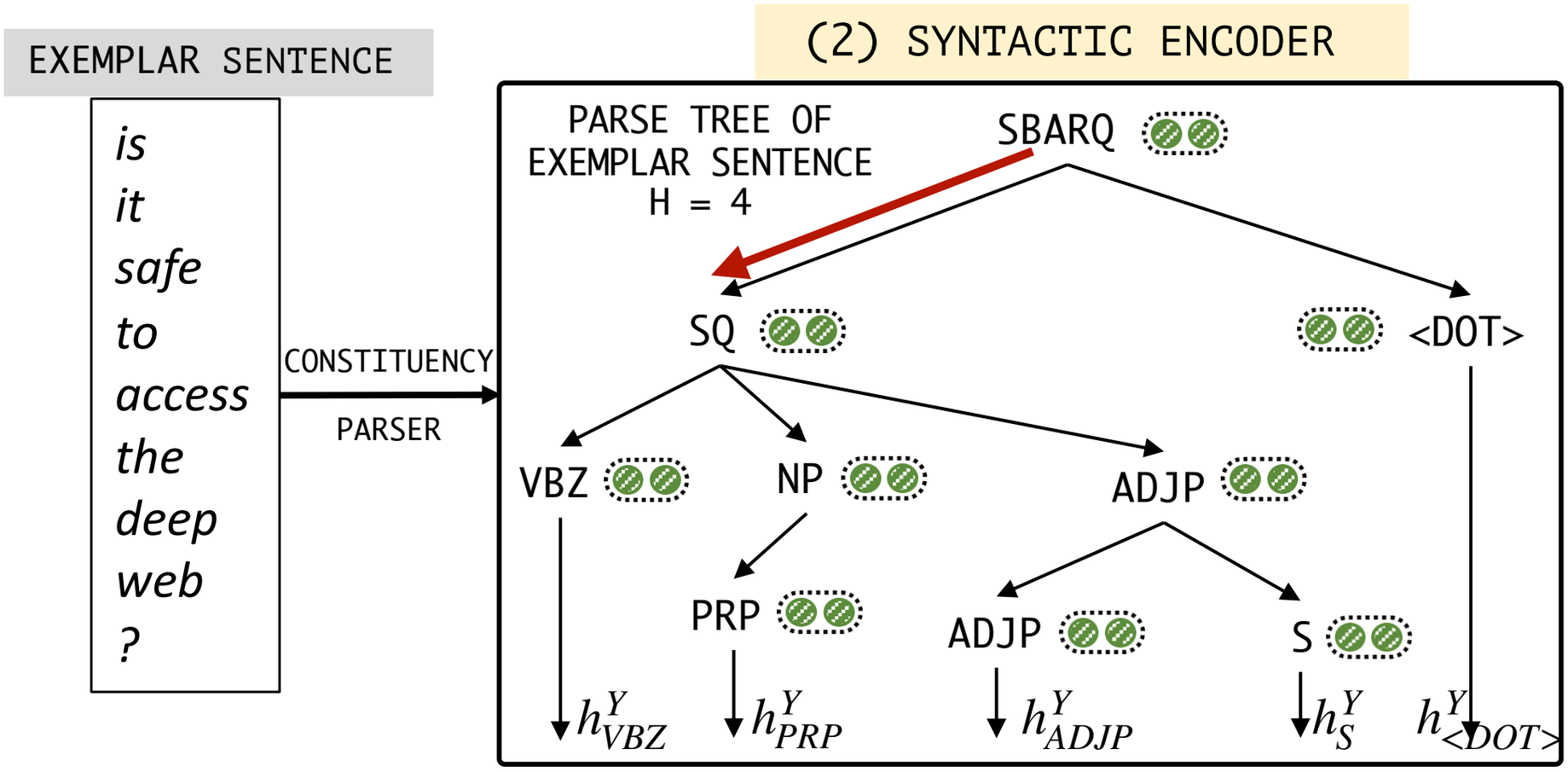
SGCP: Syntactic Encoder



$$h_v^Y = \textbf{GeLU}(W_{pa} h_{pa(v)}^Y + W_v e(y_v) + b_v)$$

$$\mathbb{L}_H^Y = [h_{VBZ}^Y, h_{PRP}^Y, h_{ADJP}^Y, h_S^Y, h_{<DOT>}^Y]$$

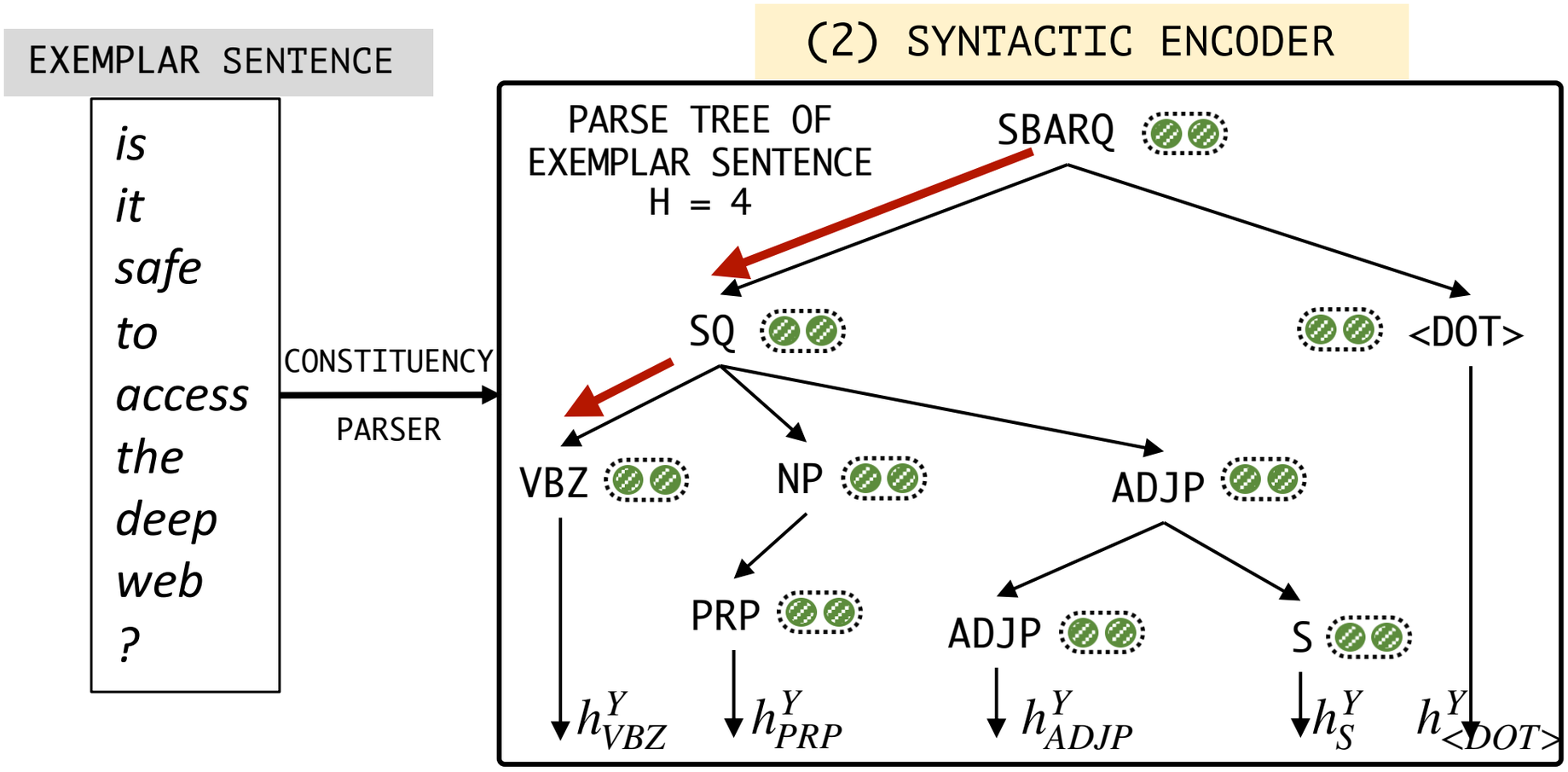
SGCP: Syntactic Encoder



$$h_v^Y = \textbf{GeLU}(W_{pa} h_{pa(v)}^Y + W_v e(y_v) + b_v)$$

$$\mathbb{L}_H^Y = [h_{VBZ}^Y, h_{PRP}^Y, h_{ADJP}^Y, h_S^Y, h_{<DOT>}^Y]$$

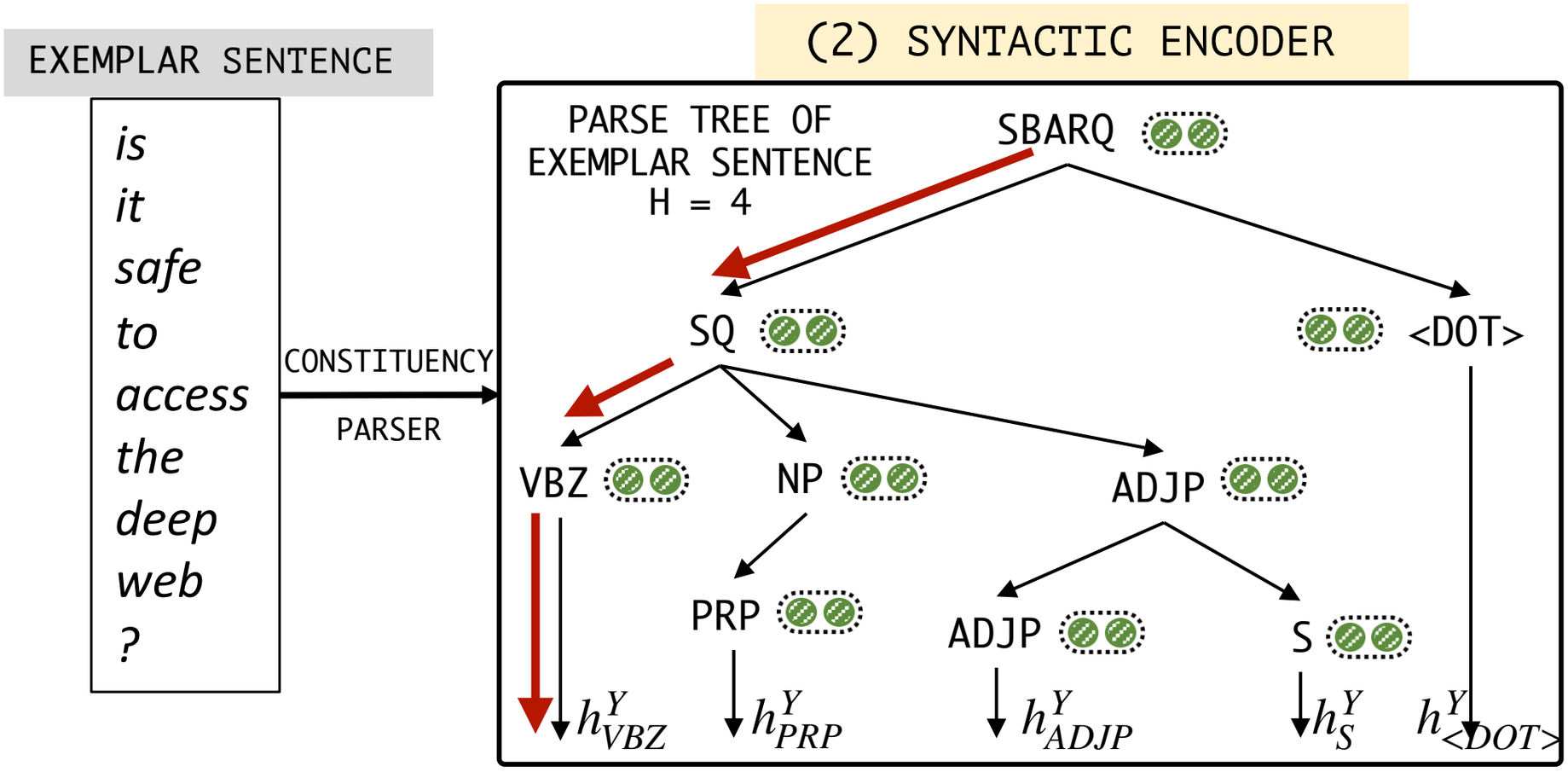
SGCP: Syntactic Encoder



$$h_v^Y = \textbf{GeLU}(W_{pa} h_{pa(v)}^Y + W_v e(y_v) + b_v)$$

$$\mathbb{L}_H^Y = [h_{VBZ}^Y, h_{PRP}^Y, h_{ADJP}^Y, h_S^Y, h_{<DOT>}^Y]$$

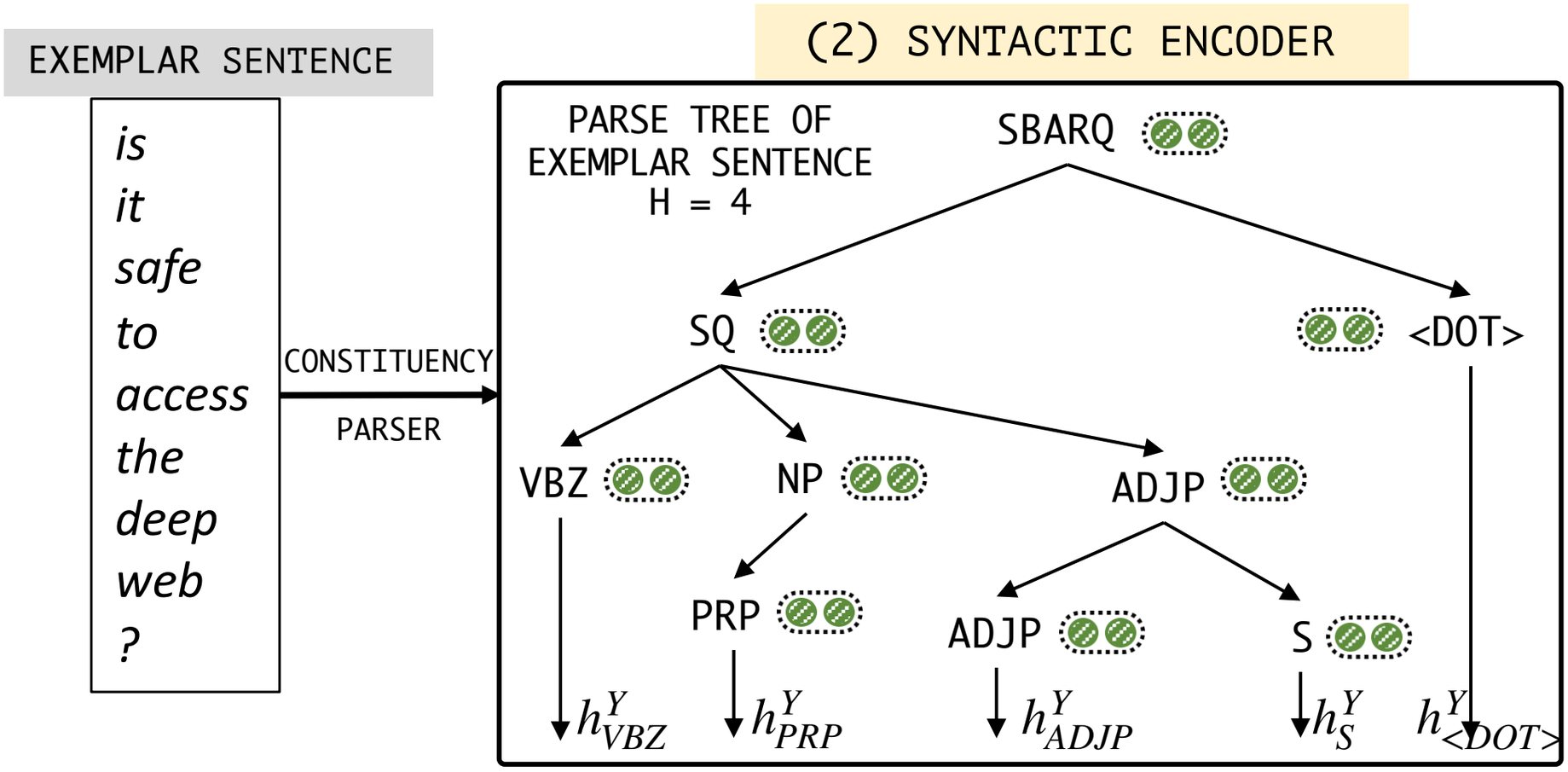
SGCP: Syntactic Encoder



$$h_v^Y = \textbf{GeLU}(W_{pa} h_{pa(v)}^Y + W_v e(y_v) + b_v)$$

$$\mathbb{L}_H^Y = [h_{VBZ}^Y, h_{PRP}^Y, h_{ADJP}^Y, h_S^Y, h_{<DOT>}^Y]$$

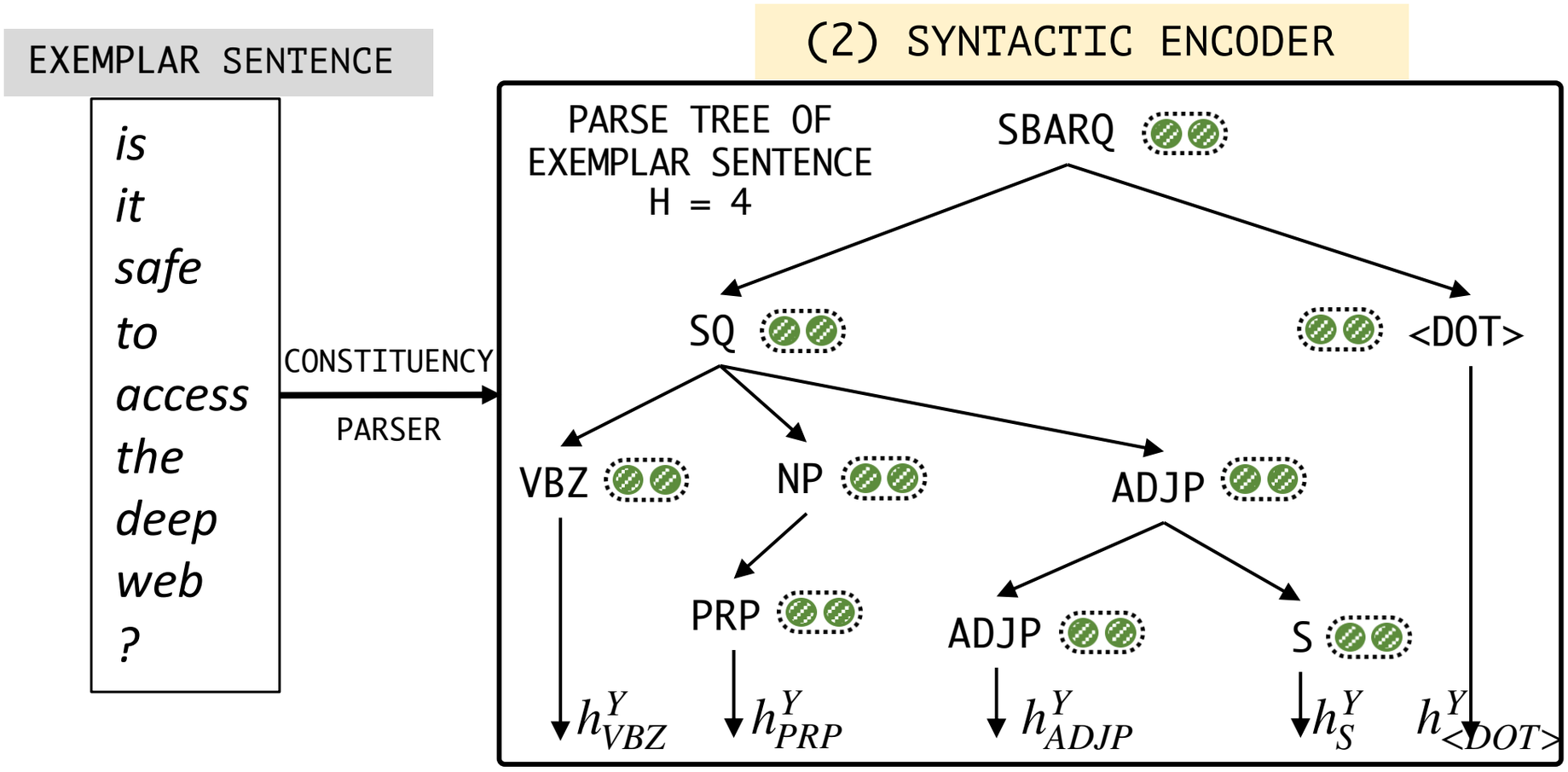
SGCP: Syntactic Encoder



$$h_v^Y = \textbf{GeLU}(W_{pa} h_{pa(v)}^Y + W_v e(y_v) + b_v)$$

$$\mathbb{L}_H^Y = [h_{VBZ}^Y, h_{PRP}^Y, h_{ADJP}^Y, h_S^Y, h_{<DOT>}^Y]$$

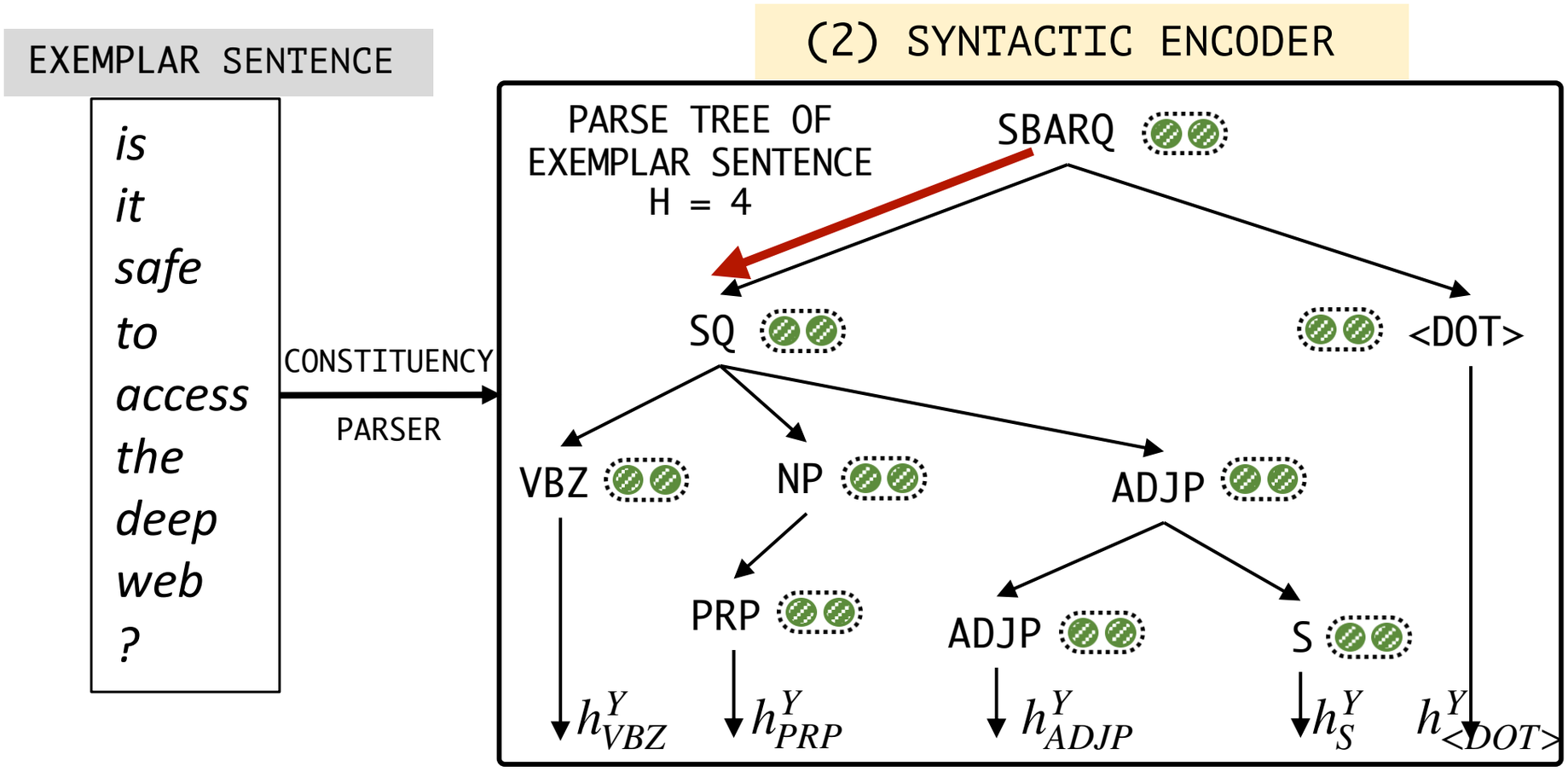
SGCP: Syntactic Encoder



$$h_v^Y = \textbf{GeLU}(W_{pa} h_{pa(v)}^Y + W_v e(y_v) + b_v)$$

$$\mathbb{L}_H^Y = [h_{VBZ}^Y, h_{PRP}^Y, h_{ADJP}^Y, h_S^Y, h_{<DOT>}^Y]$$

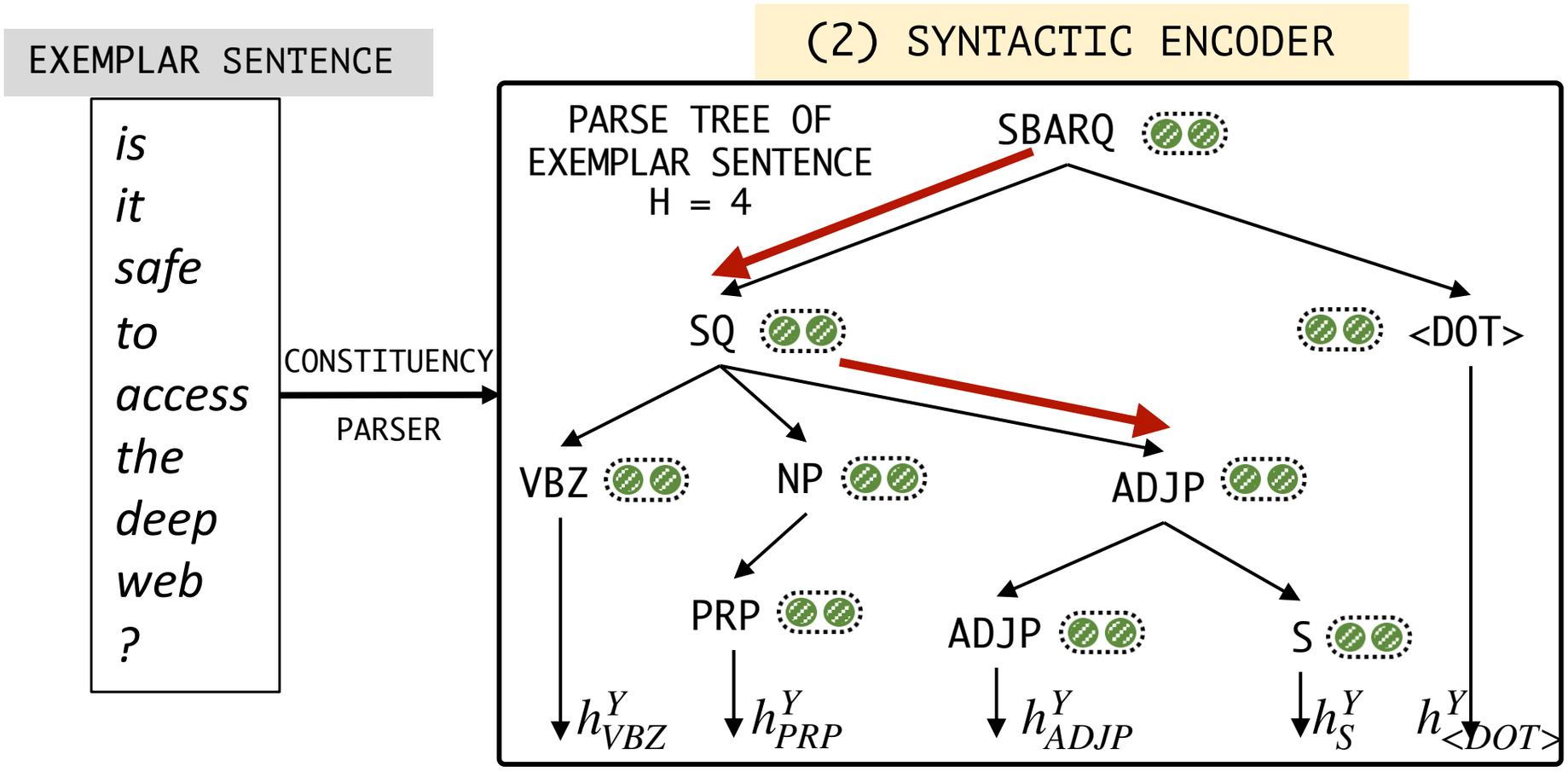
SGCP: Syntactic Encoder



$$h_v^Y = \textbf{GeLU}(W_{pa} h_{pa(v)}^Y + W_v e(y_v) + b_v)$$

$$\mathbb{L}_H^Y = [h_{VBZ}^Y, h_{PRP}^Y, h_{ADJP}^Y, h_S^Y, h_{<DOT>}^Y]$$

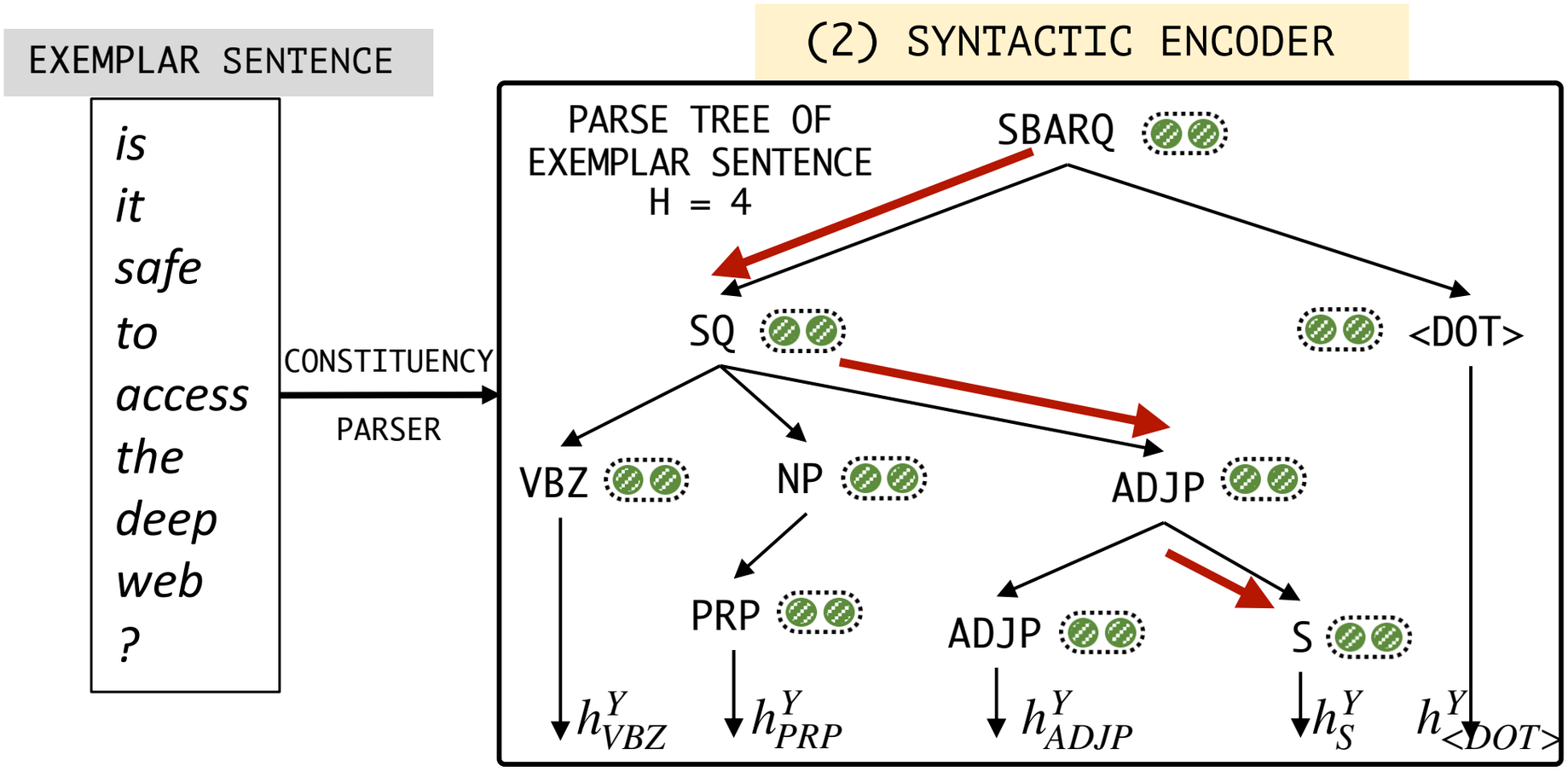
SGCP: Syntactic Encoder



$$h_v^Y = \textbf{GeLU}(W_{pa} h_{pa(v)}^Y + W_v e(y_v) + b_v)$$

$$\mathbb{L}_H^Y = [h_{VBZ}^Y, h_{PRP}^Y, h_{ADJP}^Y, h_S^Y, h_{<DOT>}^Y]$$

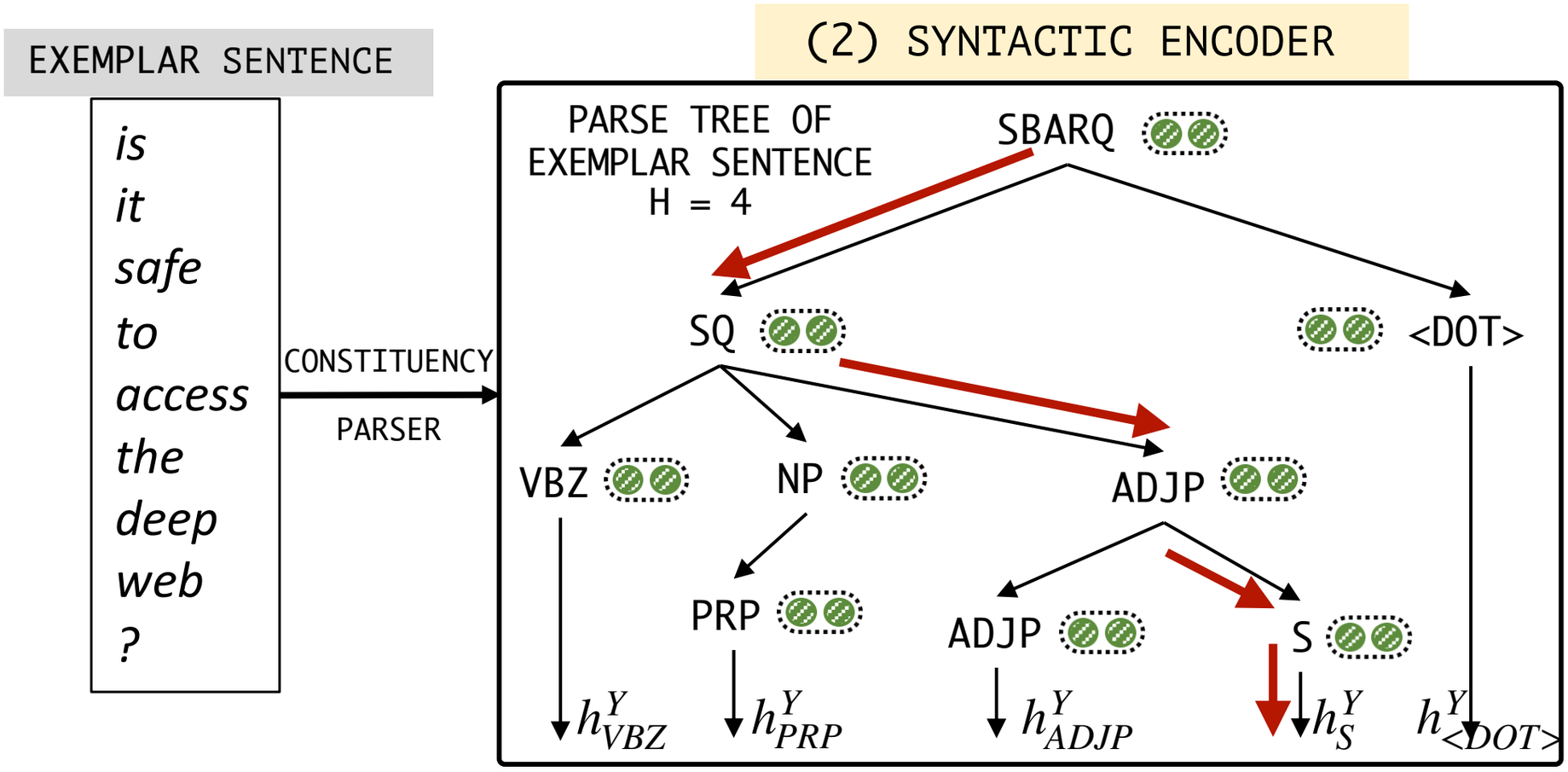
SGCP: Syntactic Encoder



$$h_v^Y = \textbf{GeLU}(W_{pa} h_{pa(v)}^Y + W_v e(y_v) + b_v)$$

$$\mathbb{L}_H^Y = [h_{VBZ}^Y, h_{PRP}^Y, h_{ADJP}^Y, h_S^Y, h_{<DOT>}^Y]$$

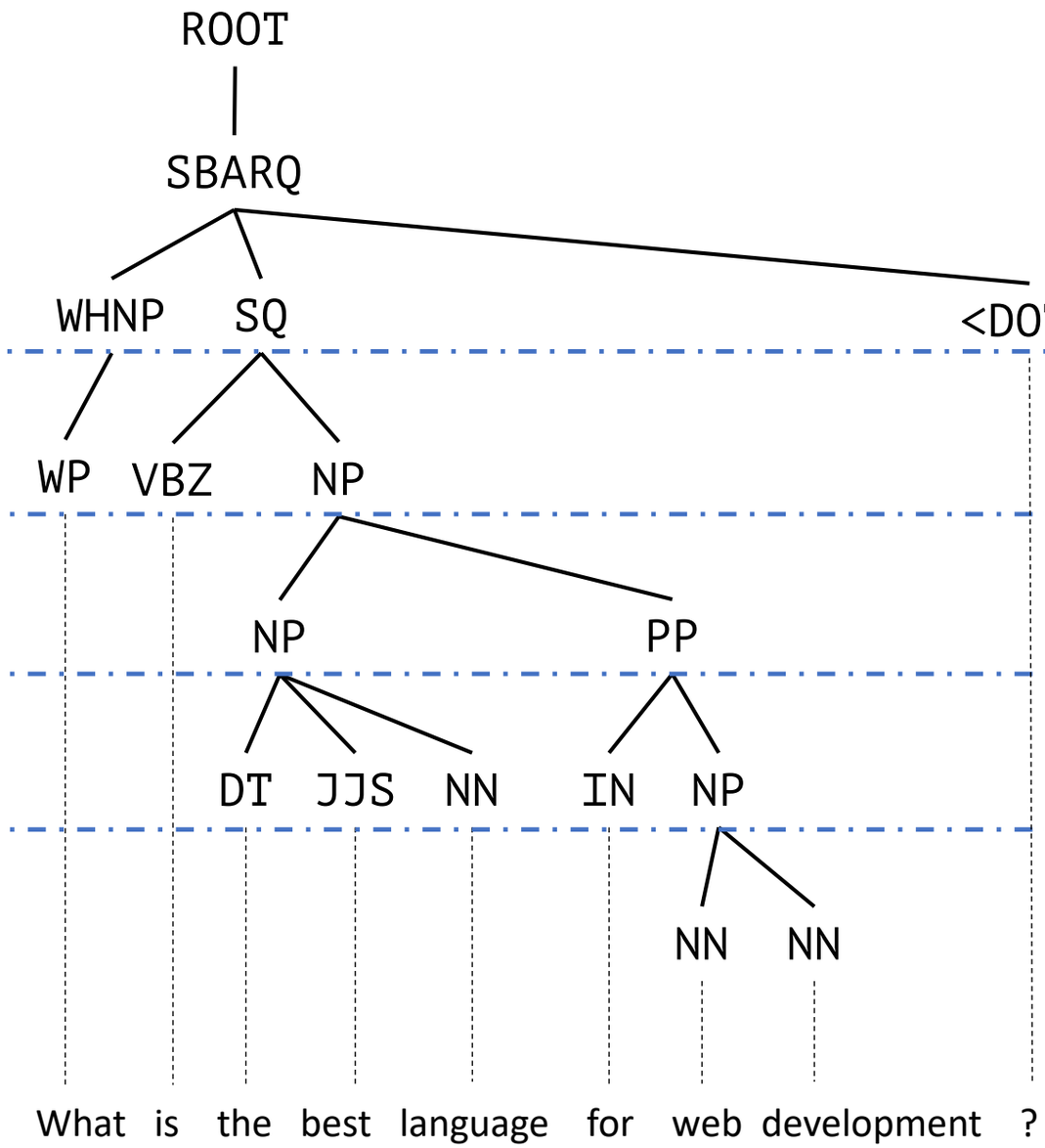
SGCP: Syntactic Encoder



$$h_v^Y = \textbf{GeLU}(W_{pa} h_{pa(v)}^Y + W_v e(y_v) + b_v)$$

$$\mathbb{L}_H^Y = [h_{VBZ}^Y, h_{PRP}^Y, h_{ADJP}^Y, h_S^Y, h_{<DOT>}^Y]$$

Syntactic Tree to Syntactic Signalling Vector (Only during Training)



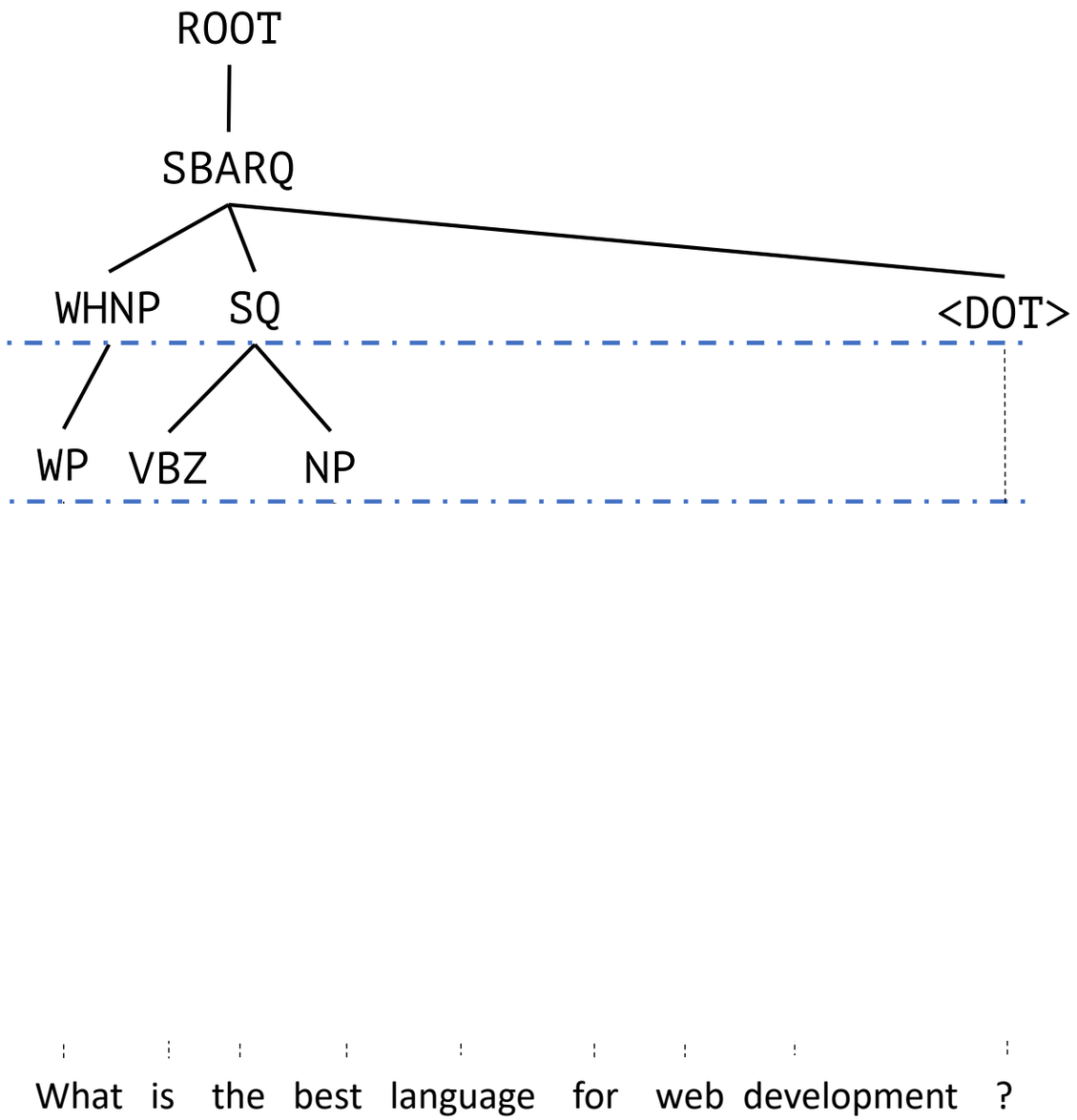
$H = 2; a$ = (1,1,0,0,0,0,0,0,1)

$H = 3; a$ = (1,1,1,0,0,0,0,0,1)

$H = 4; a$ = (1,1,1,0,0,1,0,0,1)

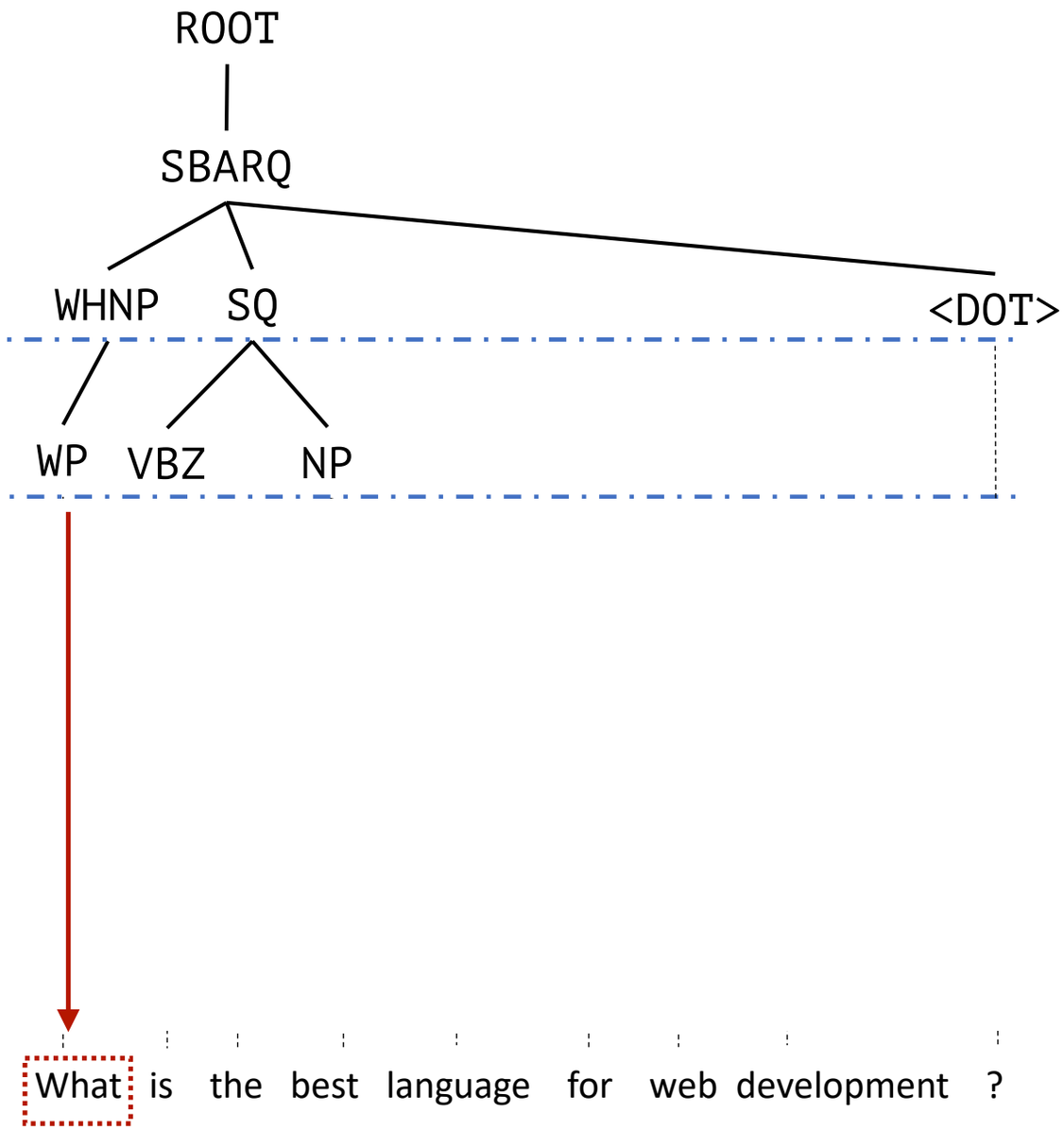
$H = 5; a$ = (1,1,1,1,1,1,1,0,1)

Syntactic Tree to Syntactic Signalling Vector (Only during Training)



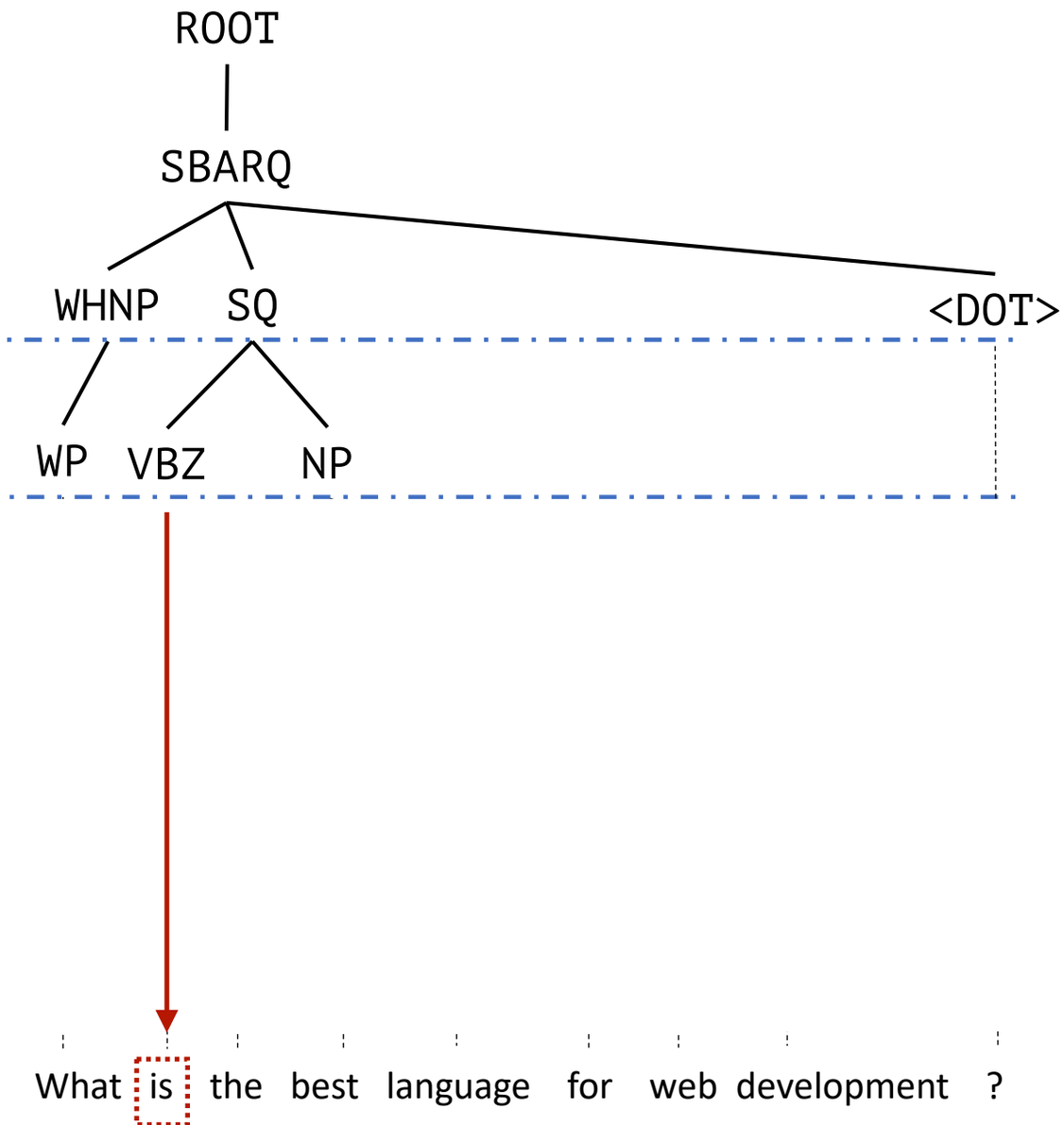
$H = 3; a = (1,1,1,0,0,0,0,0,1)$

Syntactic Tree to Syntactic Signalling Vector (Only during Training)



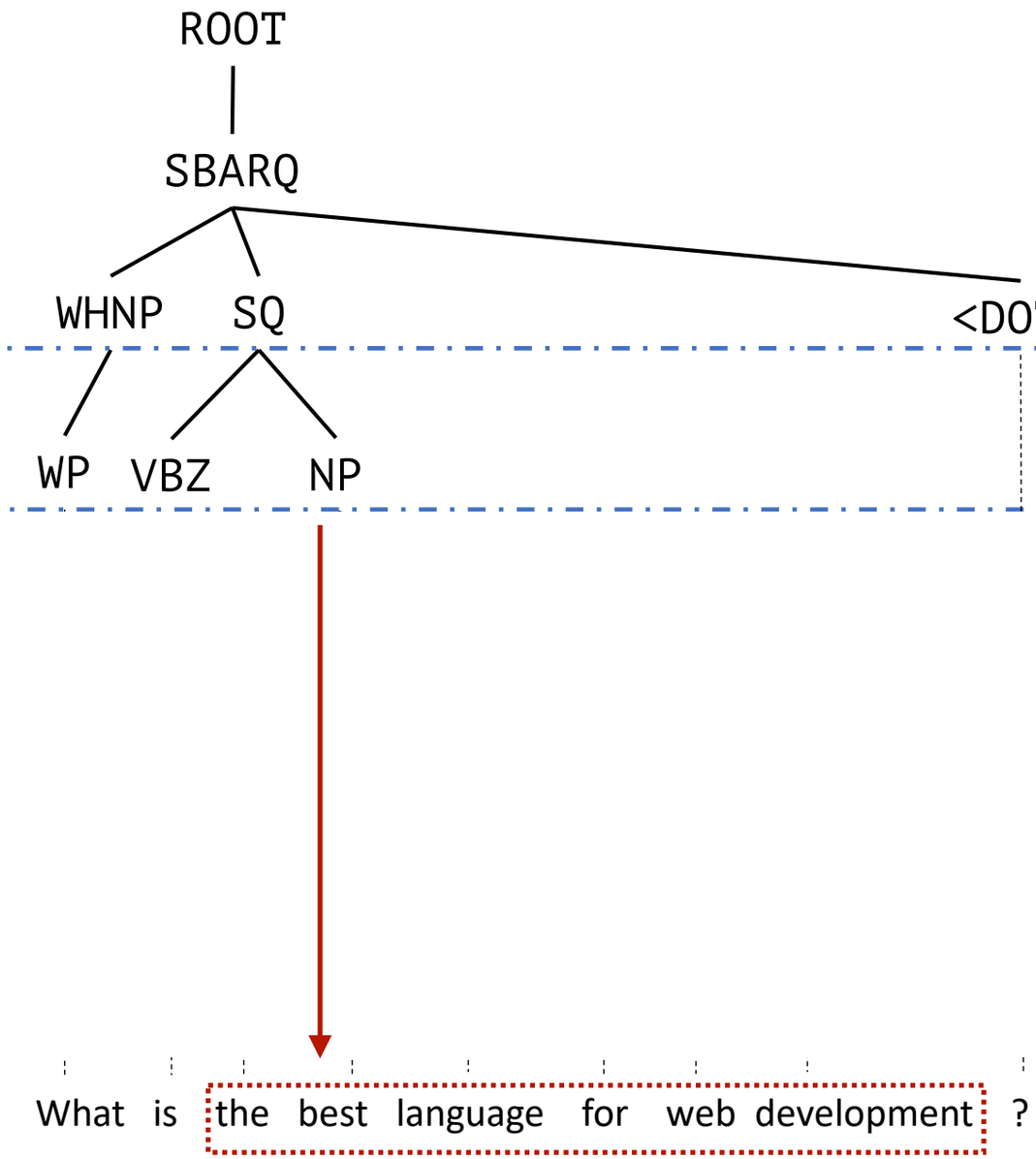
$H = 3; a = (1,1,1,0,0,0,0,0,1)$

Syntactic Tree to Syntactic Signalling Vector (Only during Training)



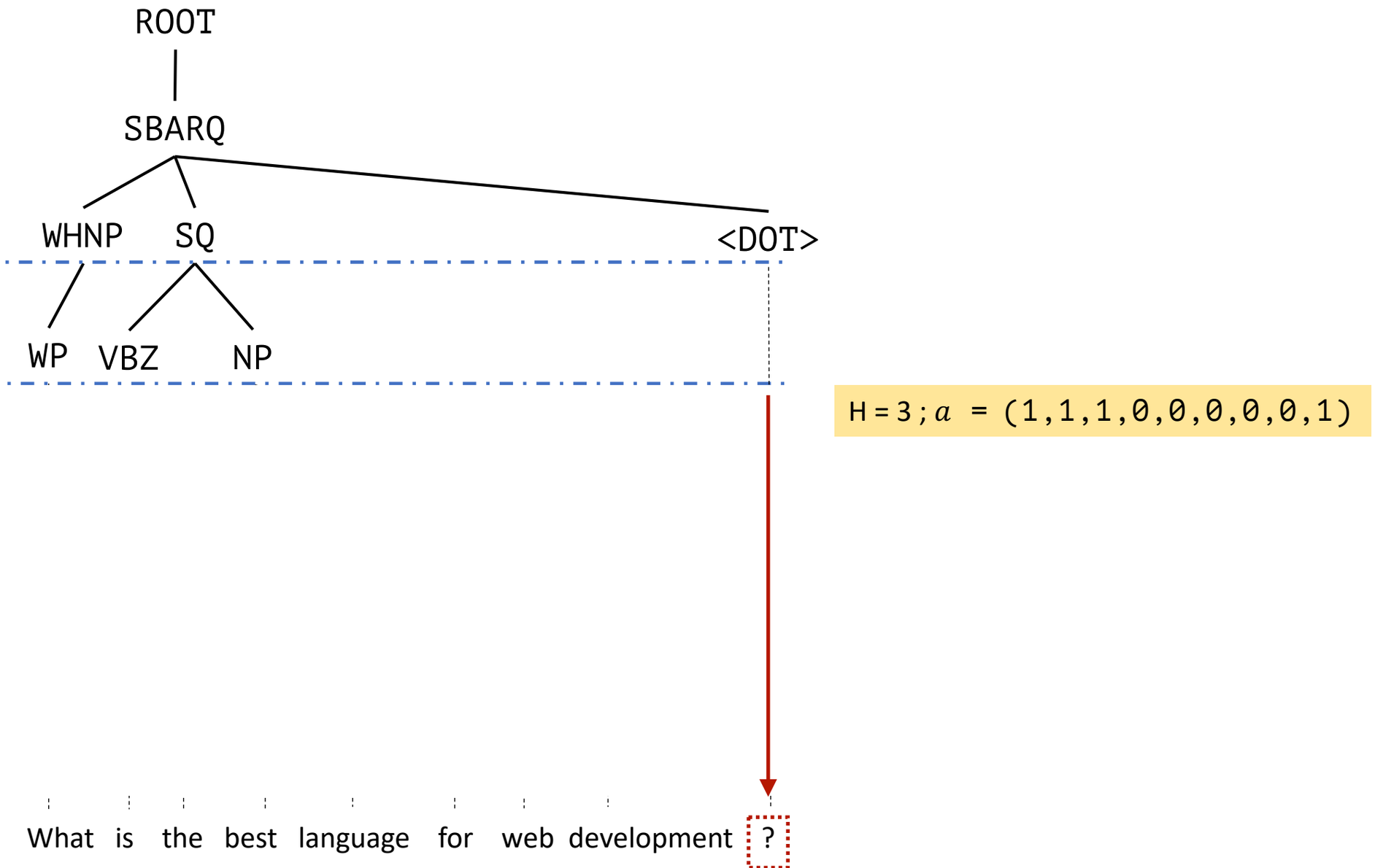
$H = 3; a = (1,1,1,0,0,0,0,0,1)$

Syntactic Tree to Syntactic Signalling Vector (Only during Training)



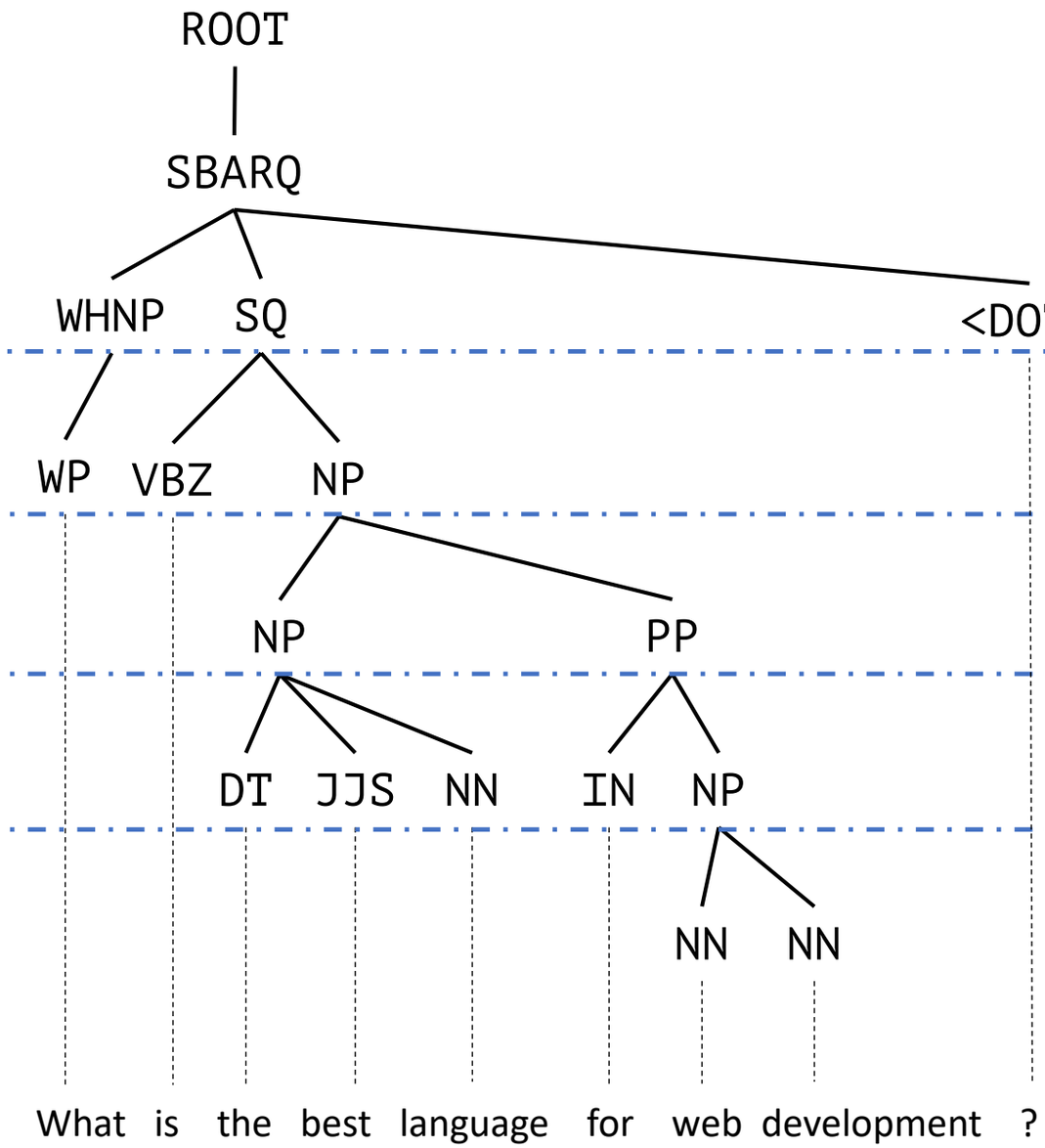
$H = 3; a = (1,1,1,0,0,0,0,0,1)$

Syntactic Tree to Syntactic Signalling Vector (Only during Training)



$H = 3; a = (1,1,1,0,0,0,0,0,1)$

Syntactic Tree to Syntactic Signalling Vector (Only during Training)



$H = 2; a$ = (1,1,0,0,0,0,0,0,1)

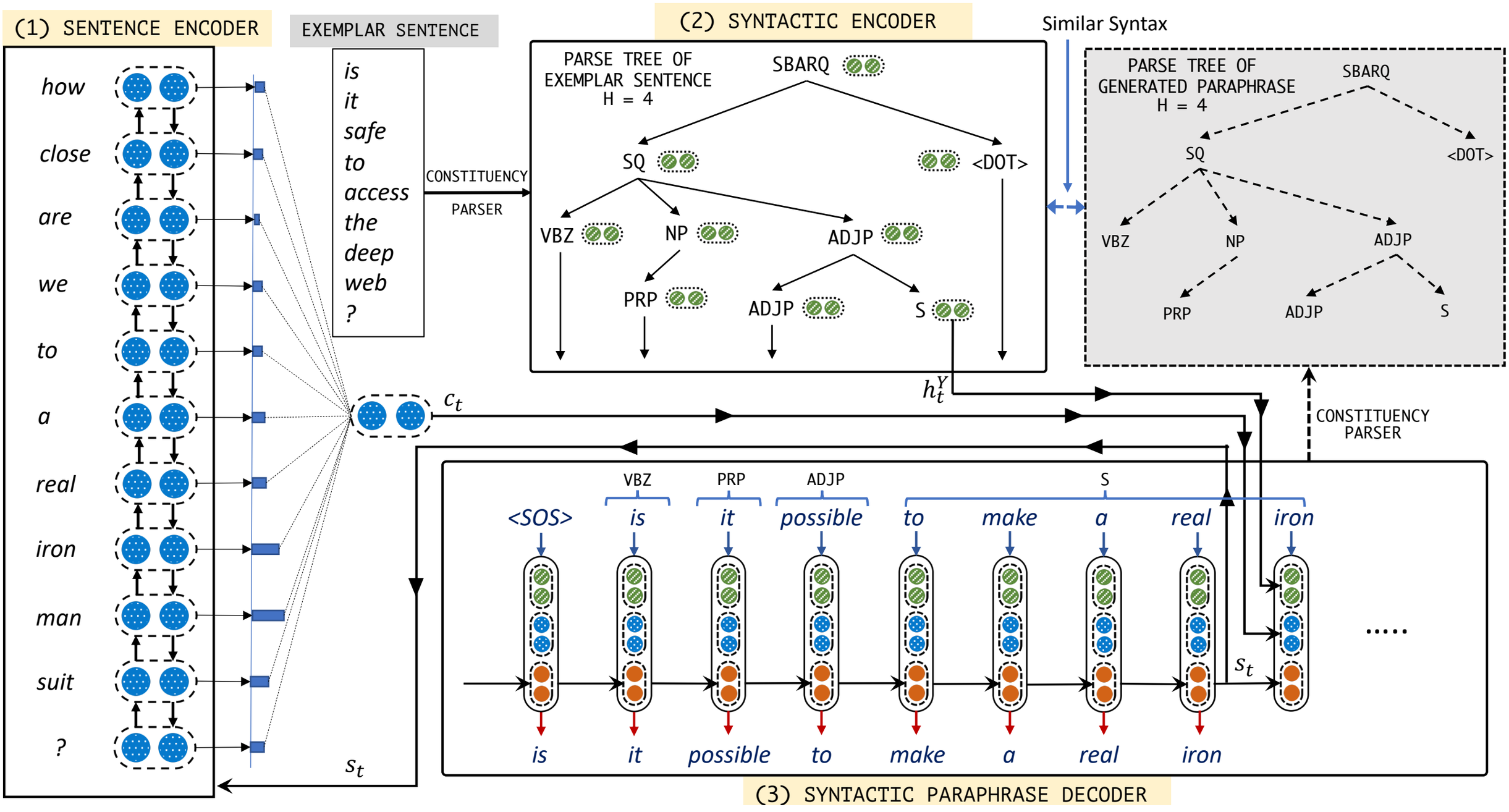
$H = 3; a$ = (1,1,1,0,0,0,0,0,1)

$H = 4; a$ = (1,1,1,0,0,1,0,0,1)

$H = 5; a$ = (1,1,1,1,1,1,1,0,1)

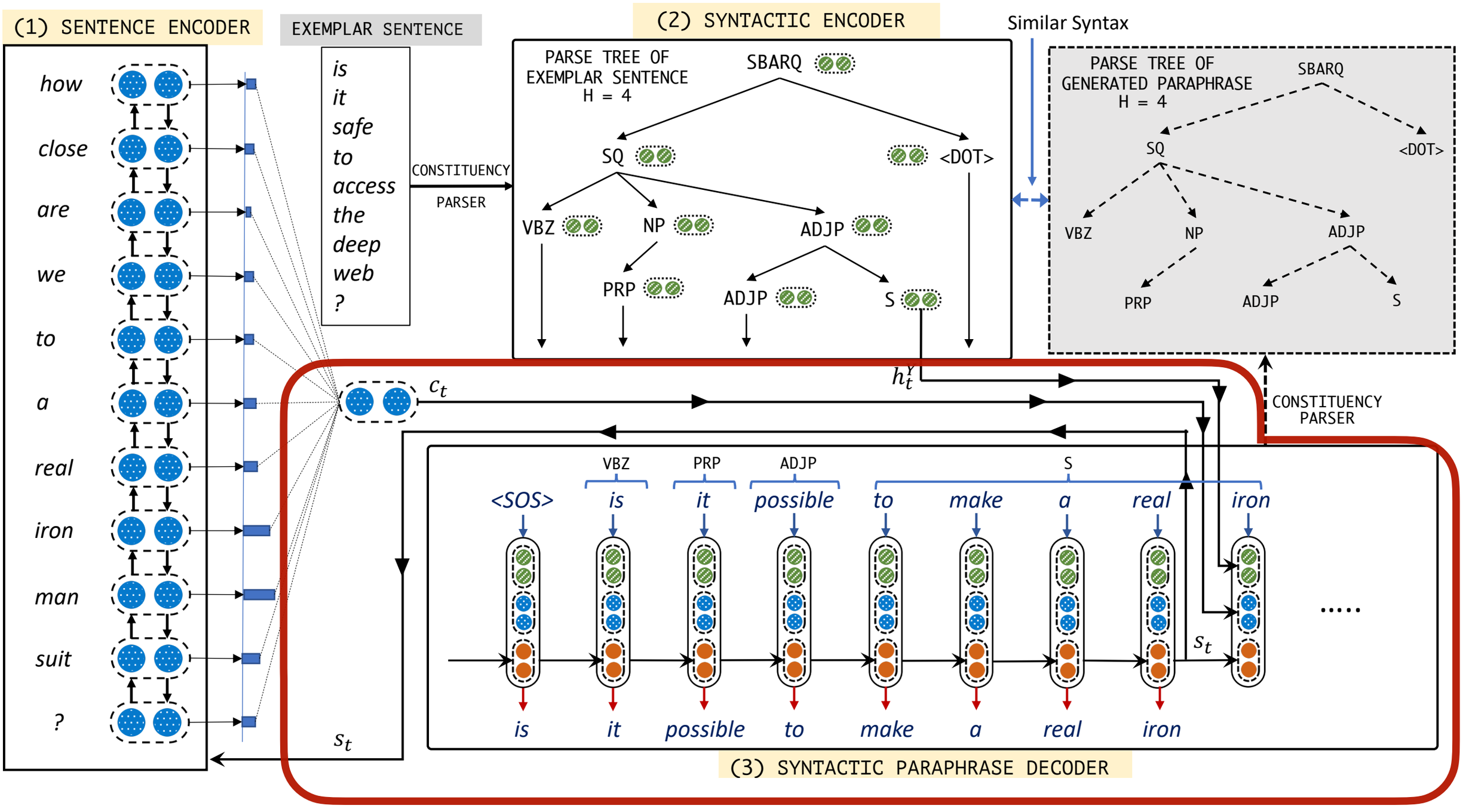
SGCP: Syntax Guided Controlled Paraphraser

Use Syntactic Tree Structure to Guide Paraphrase Generation model

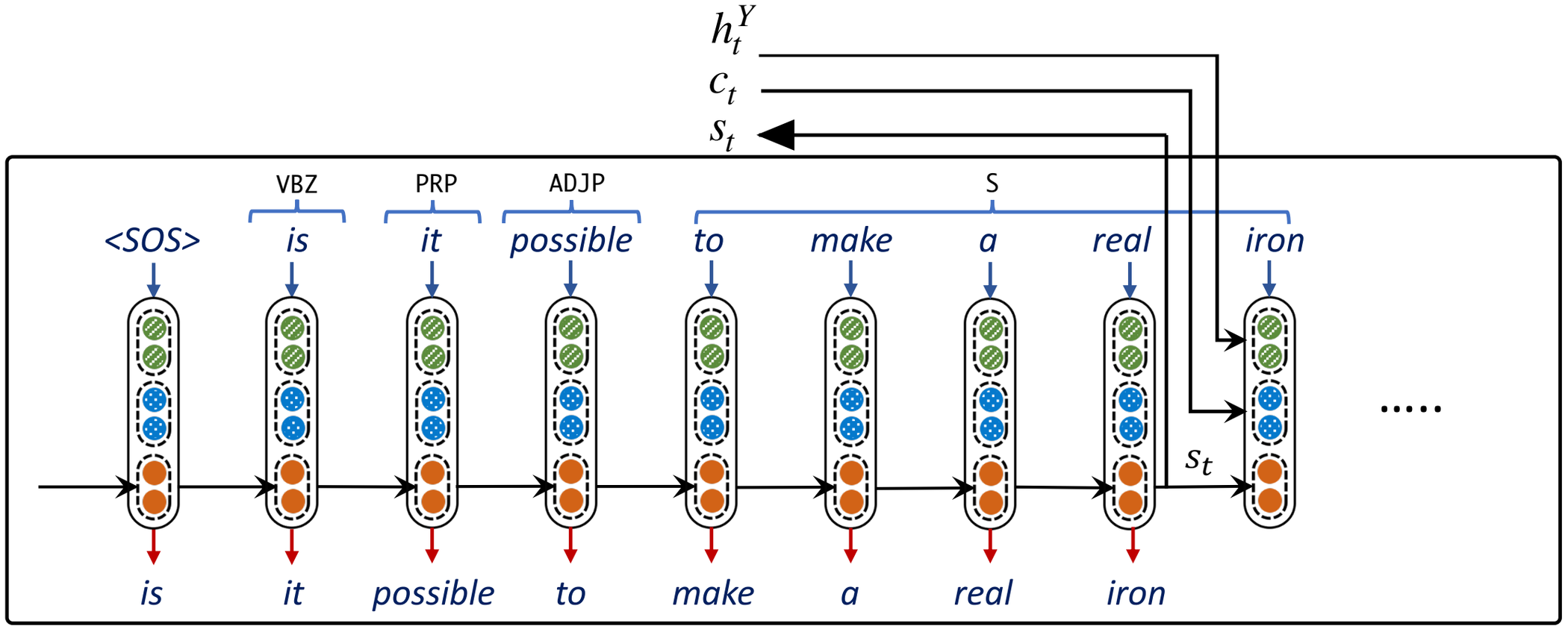


SGCP: Syntax Guided Controlled Paraphraser

Use Syntactic Tree Structure to Guide Paraphrase Generation model

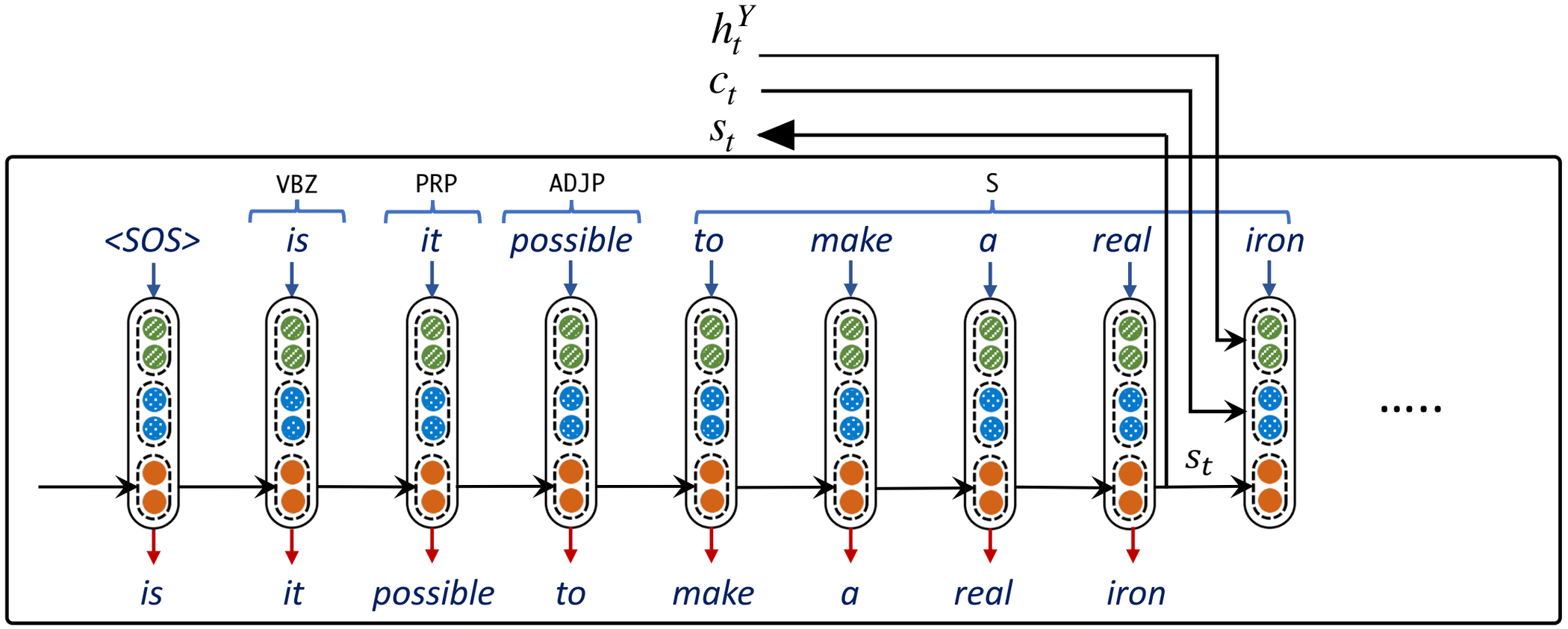


SGCP: Decoder



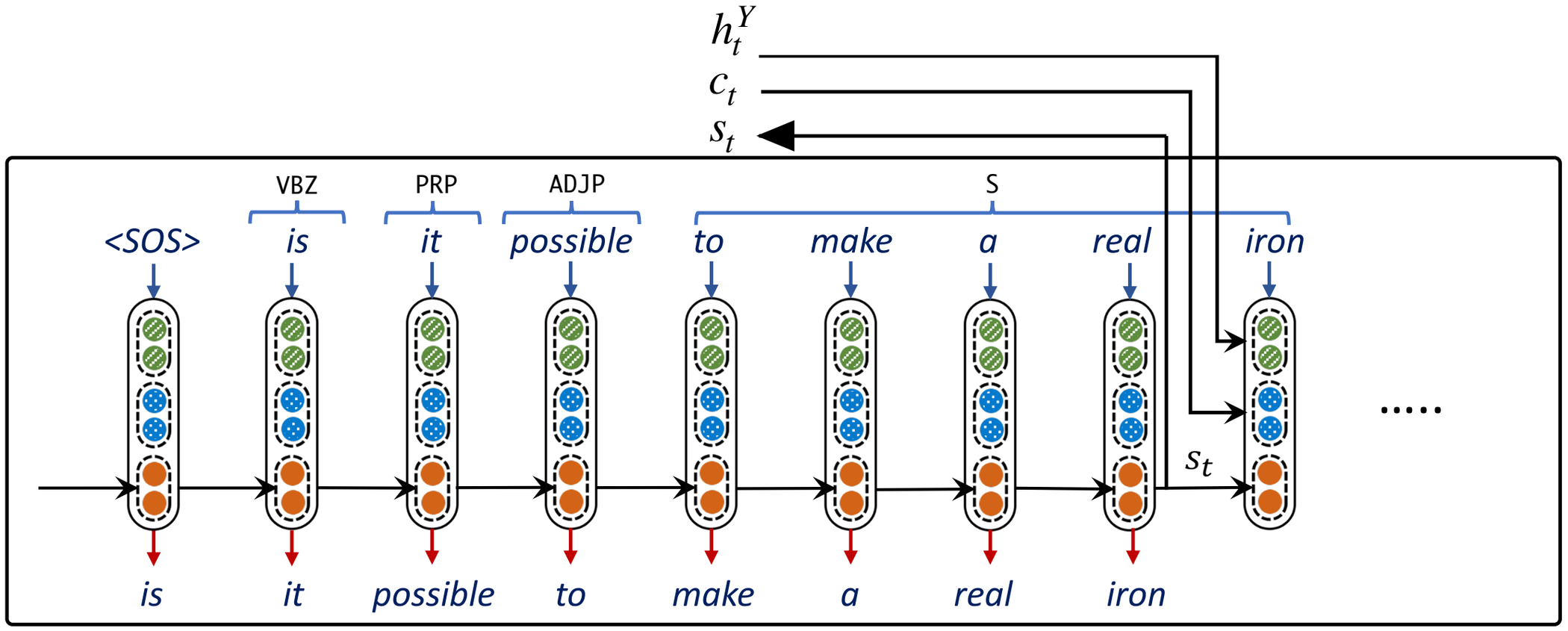
(3) SYNTACTIC PARAPHRASE DECODER

SGCP: Decoder



(3) SYNTACTIC PARAPHRASE DECODER

SGCP: Decoder

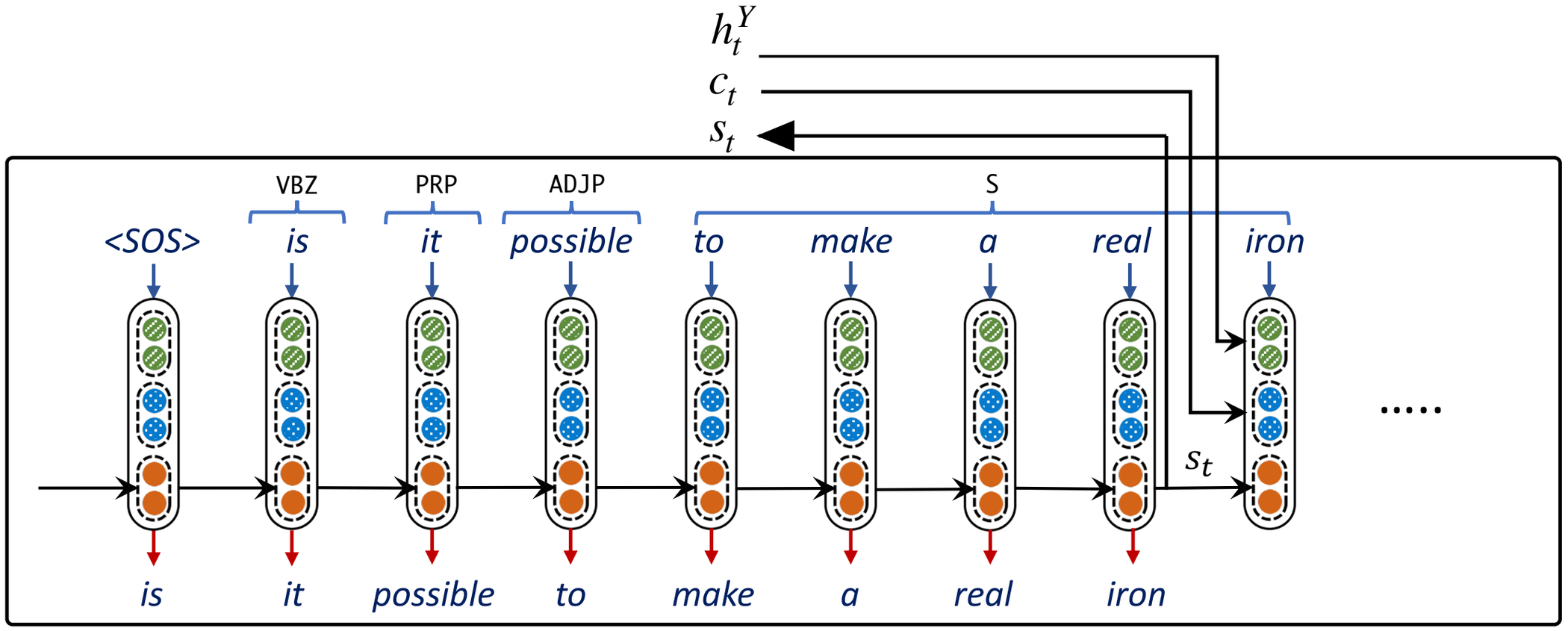


(3) SYNTACTIC PARAPHRASE DECODER

$$p_t = \sigma(W_{bop}([c_t; h_t^Y; s_t; e(z_t')]) + b_{bop})$$

$$h_{t+1}^Y = \begin{cases} h_t^Y & p_t < 0.5 \\ \textbf{pop}(\mathbb{L}_H^Y) & otherwise \end{cases}$$

SGCP: Decoder



(3) SYNTACTIC PARAPHRASE DECODER

$$p_t = \sigma(W_{bop}([c_t; h_t^Y; s_t; e(z_t')]) + b_{bop})$$

$$h_{t+1}^Y = \begin{cases} h_t^Y & p_t < 0.5 \\ \textbf{pop}(\mathbb{L}_H^Y) & otherwise \end{cases}$$

$$\mathbb{P}(z) = \textbf{softmax}(W([c_t; h_t^Y; s_t; e(z_t')]) + b)$$

SGCP Objective

SGCP Objective

$$\mathscr{L} = -\frac{1}{T}\sum_{t=0}^{T}[\log \mathbb{P}(z_t^*) + a_t \log(p_t) + (1 - a_t)\log(1 - p_t)]$$

a_t : Signalling vector, p_t: Transition probability,

T : Generation Time-step, z_t^*: Ground Truth token

Dataset Statistics

Dataset Statistics

	Triples (Sentence, Exemplar, Reference)		
	Train*	Dev.	Test
ParaNMT-small	4,92,878	500	800
QQP-Pos	1,37,185	3000	3000

* During Training: Exemplar = Reference Paraphrase

Syntactic Granularity & SGCP-Variations

GRANULARITY	
SOURCE	what are pure substances ? what are some examples ?
EXEMPLAR	what are the characteristics of the elizabethan theatre ?

Syntactic Granularity & SGCP-Variations

GRANULARITY	
SOURCE	**what are pure substances ? what are some examples ?**
EXEMPLAR	**what are the characteristics of the elizabethan theatre ?**
H = 4	what are pure substances ?
H = 5	what are some of pure substances ?
H = 6	what are some examples of pure substances ?
H = 7	what are some examples of a pure substance ?

Syntactic Granularity & SGCP-Variations

GRANULARITY	
SOURCE	what are pure substances ? what are some examples ?
EXEMPLAR	what are the characteristics of the elizabethan theatre ?
H = 4	~~what are pure substances ?~~
H = 5	~~what are some of pure substances ?~~
H = 6	~~what are some examples of pure substances ?~~
H = 7	what are some examples of a pure substance ? ✓

SGCP VARIATIONS	
SGCP-F (Full Tree)	what are some examples of a pure substance ?

Syntactic Granularity & SGCP-Variations

GRANULARITY	
SOURCE	what are pure substances ? what are some examples ?
EXEMPLAR	what are the characteristics of the elizabethan theatre ?
H = 4	~~what are pure substances ?~~
H = 5	~~what are some of pure substances ?~~
H = 6	what are some examples of pure substances ? ✓
H = 7	~~what are some examples of a pure substance ?~~

SGCP VARIATIONS	
SGCP-F (Full Tree)	what are some examples of a pure substance ?
SGCP-R (ROUGE)	what are some examples of pure substances ?

Experimental Setup

Datasets

Evaluation

Experimental Setup

Datasets

1. Quora Question Pairs - Positives
2. ParaNMT-small

Evaluation

Experimental Setup

Datasets

1. Quora Question Pairs - Positives
2. ParaNMT-small

Evaluation

Fidelity BLEU, METEOR, PDS(model)

Experimental Setup

Datasets

1. Quora Question Pairs - Positives
2. ParaNMT-small

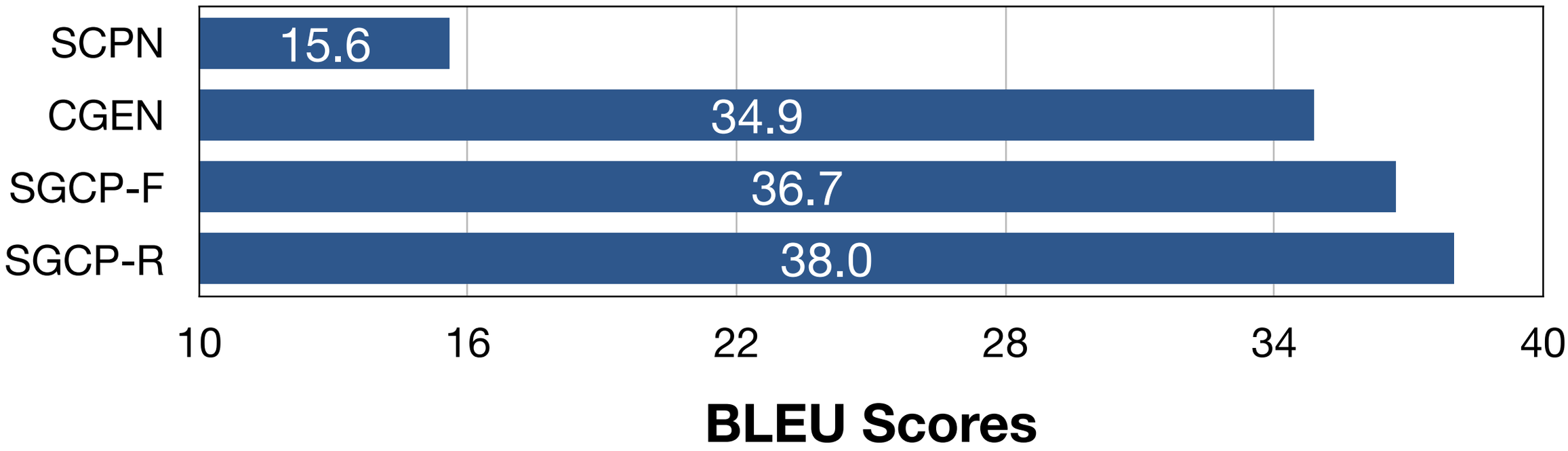
Evaluation

Fidelity BLEU, METEOR, PDS(model)

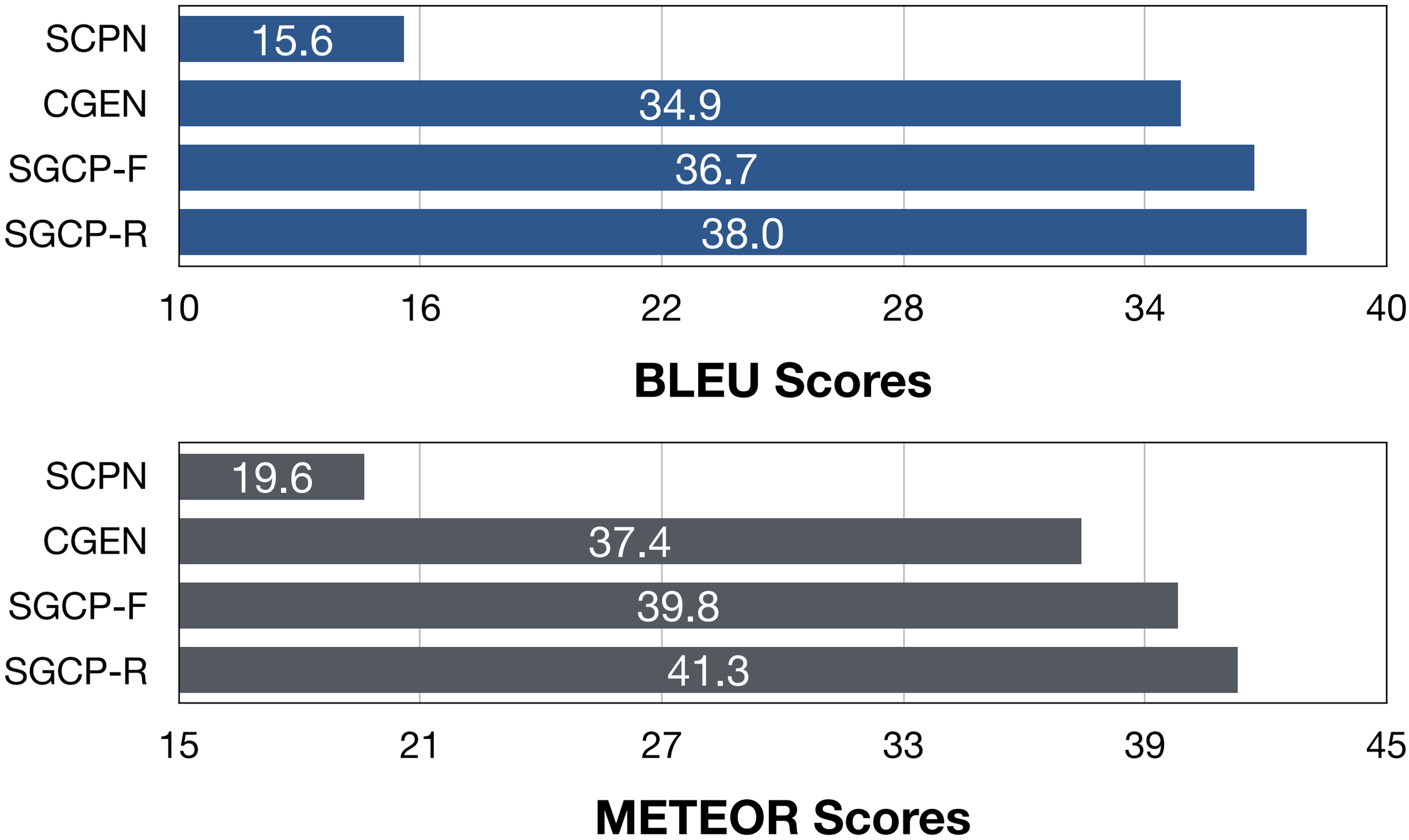
Syntax Tree Edit Distance

Fidelity
QQP-Pos Dataset

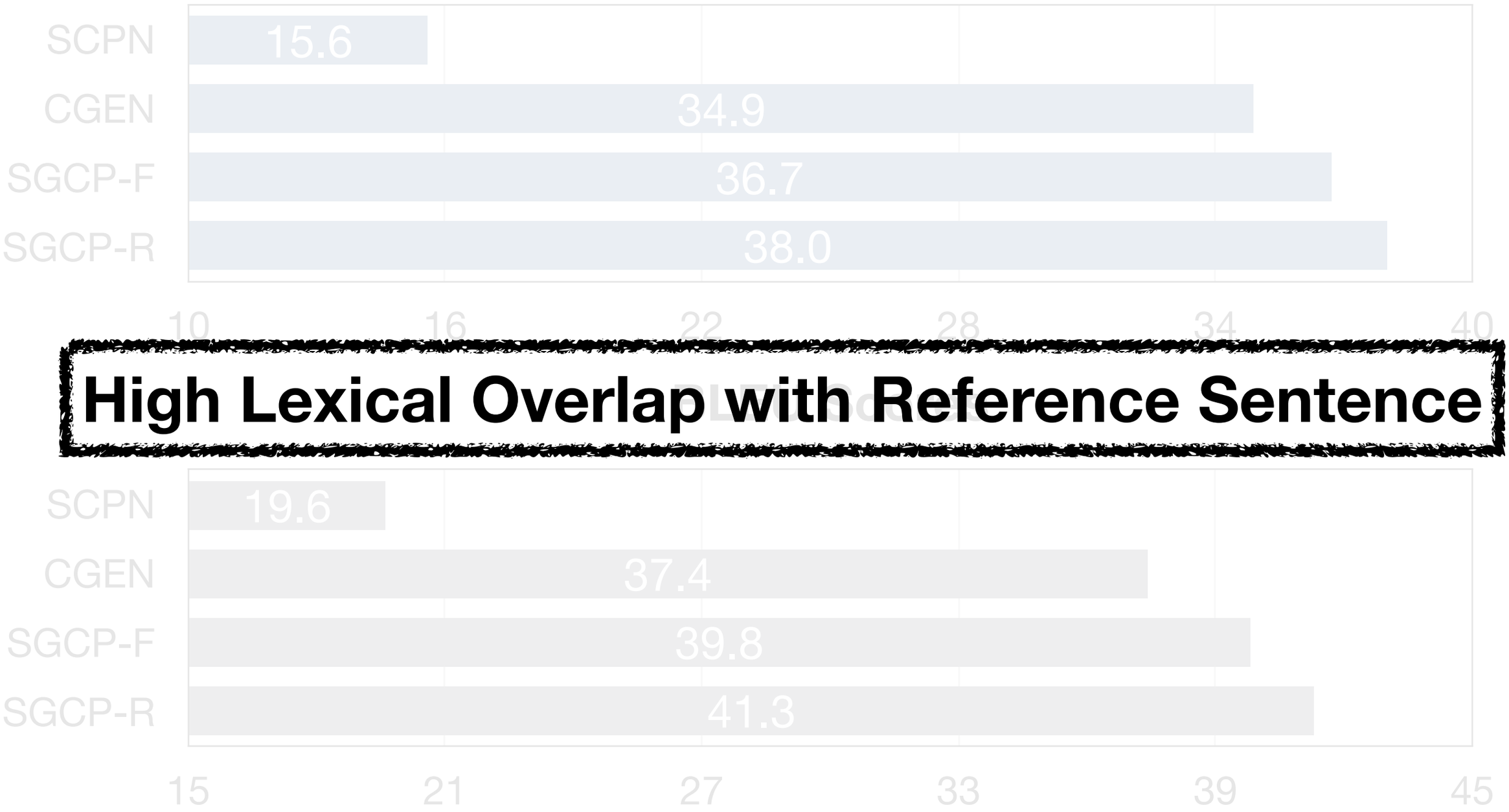
Fidelity
QQP-Pos Dataset



Fidelity
QQP-Pos Dataset



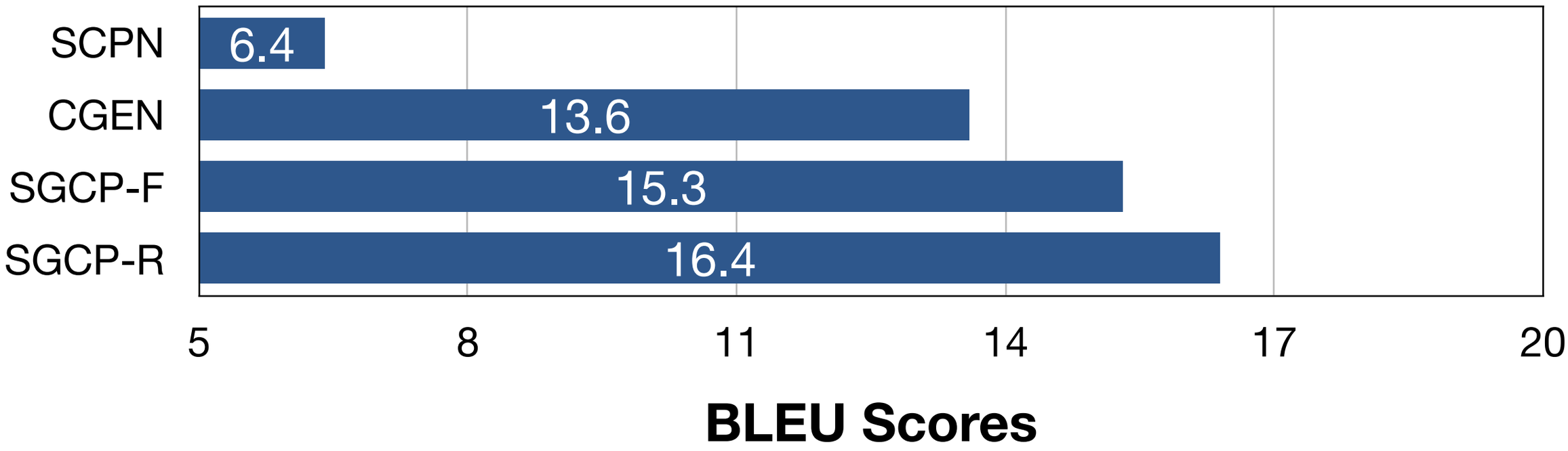
Fidelity
QQP-Pos Dataset



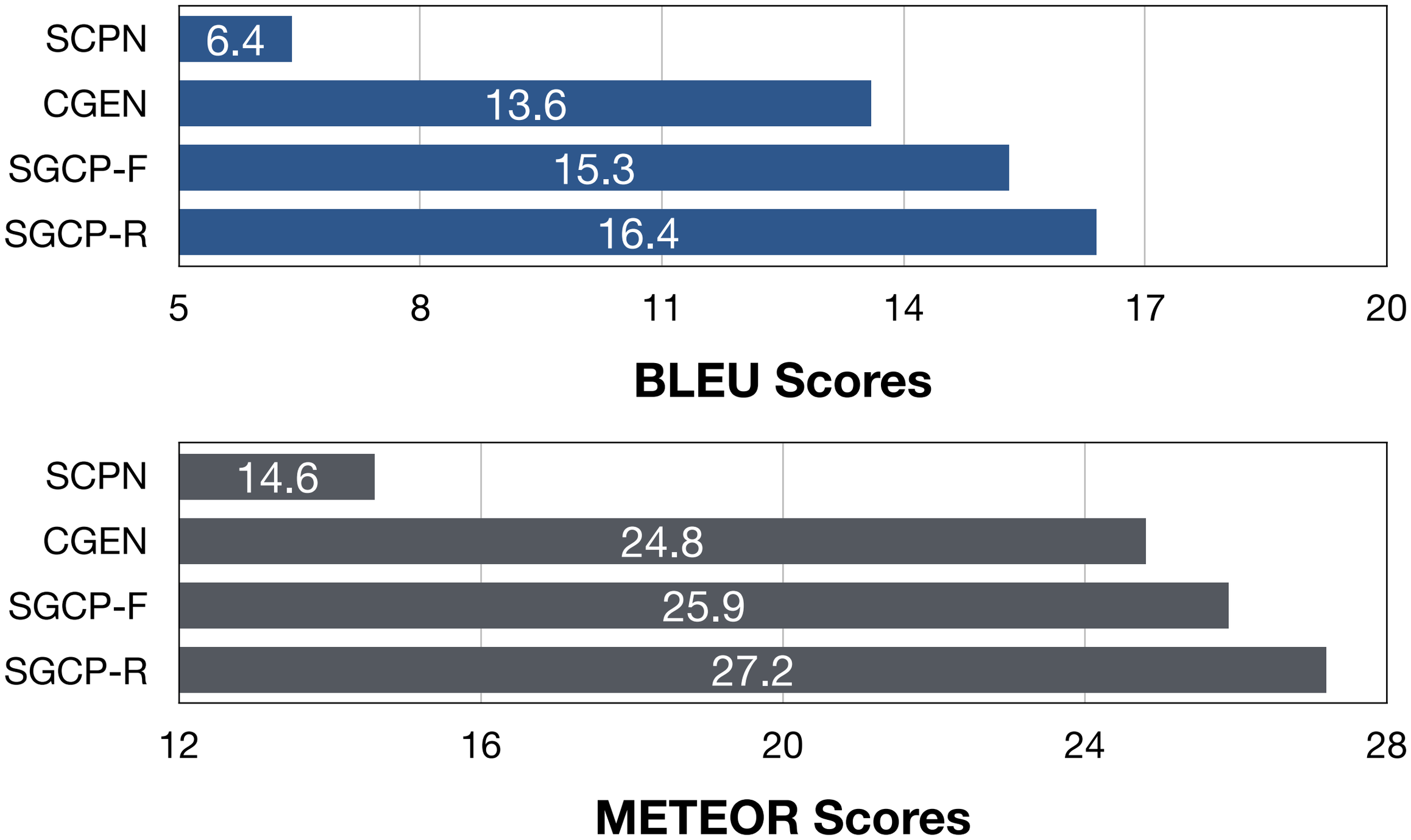
High Lexical Overlap with Reference Sentence

Fidelity
ParaNMT-small Dataset

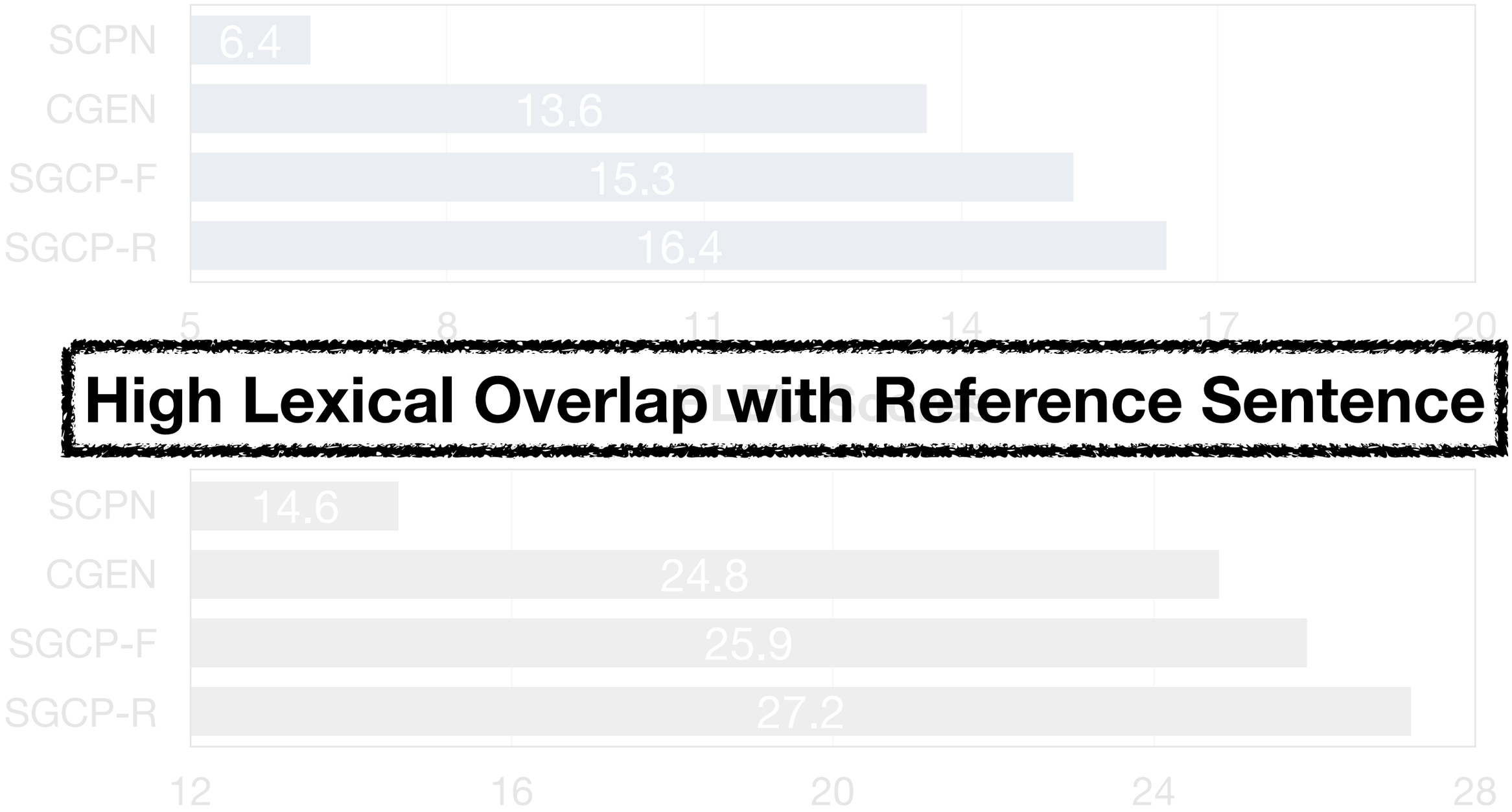
Fidelity
ParaNMT-small Dataset



Fidelity
ParaNMT-small Dataset



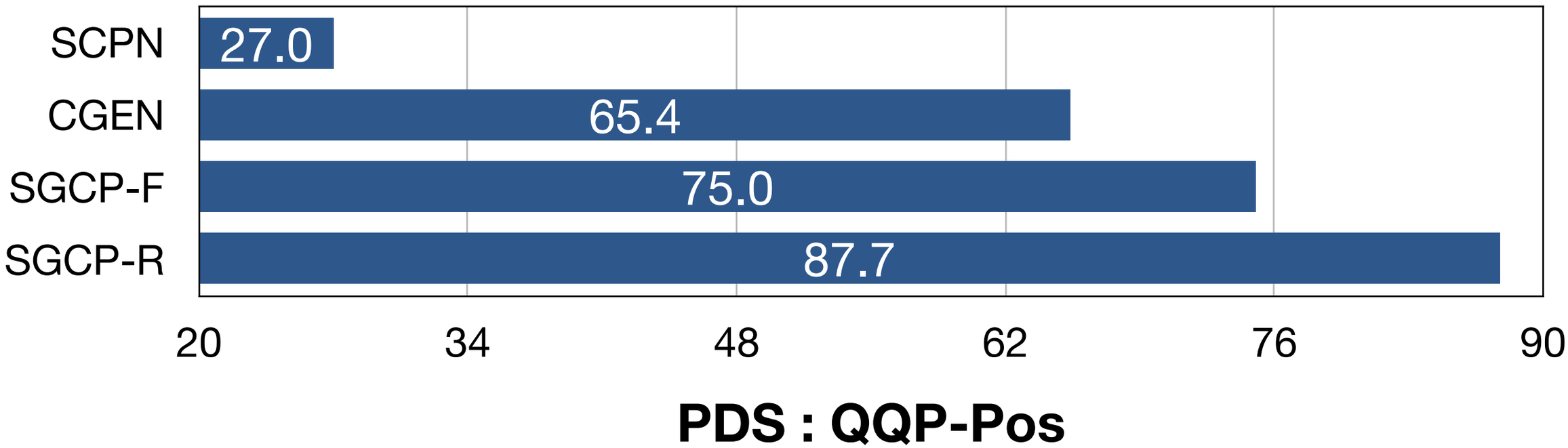
Fidelity
ParaNMT-small Dataset



High Lexical Overlap with Reference Sentence

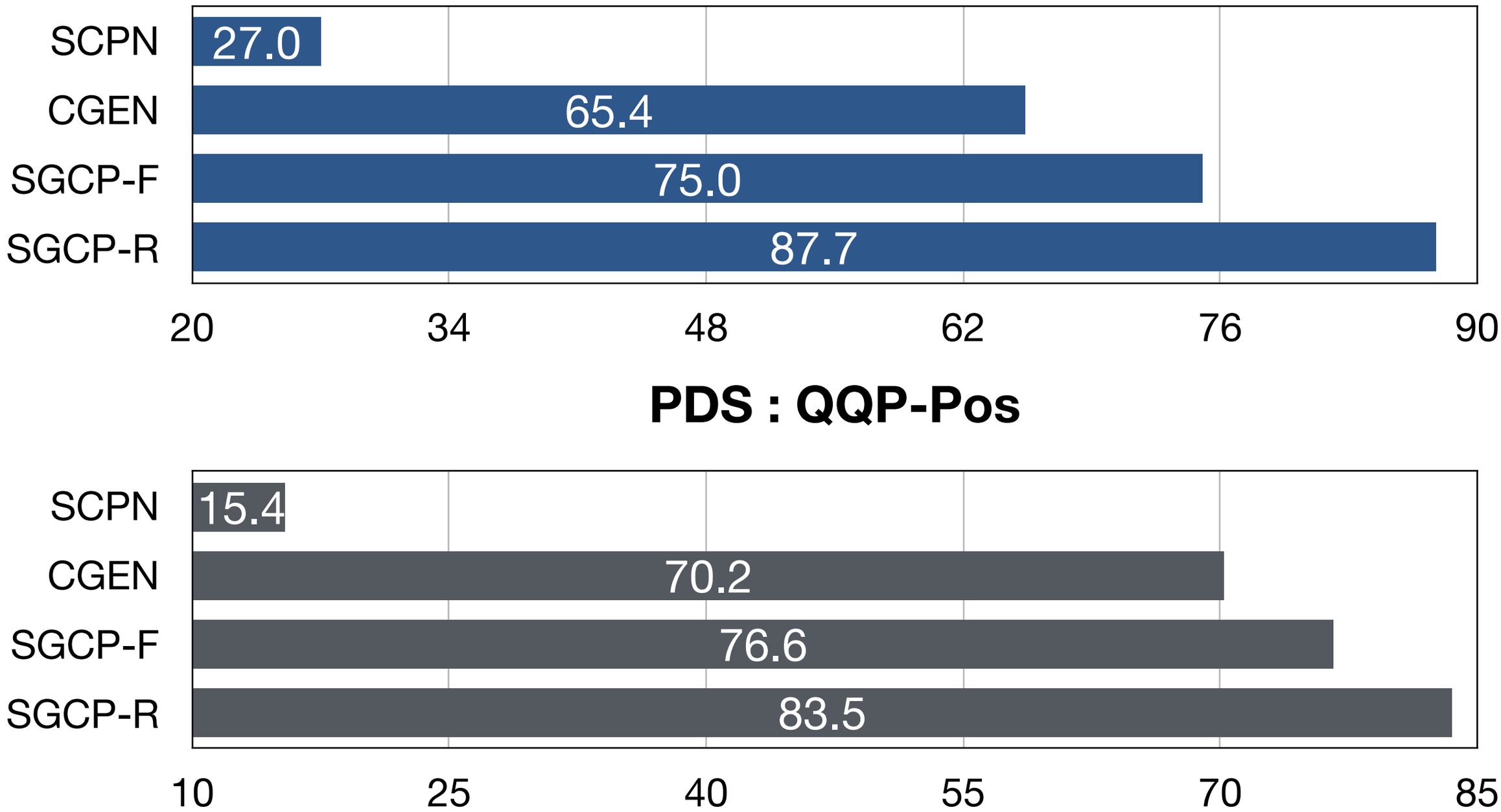
Fidelity : Paraphrase Detection Score (PDS)

Fidelity : Paraphrase Detection Score (PDS)



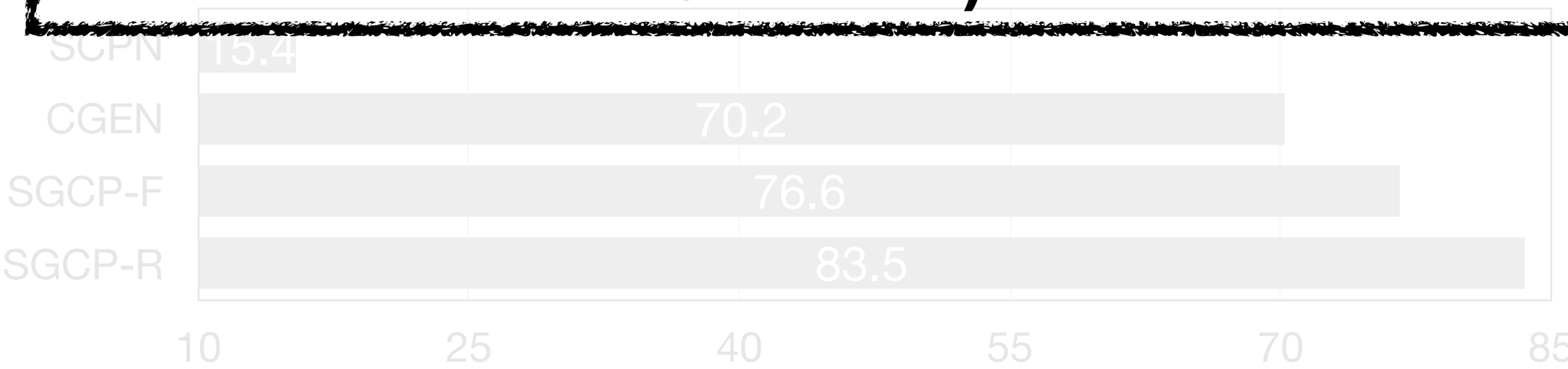
PDS : QQP-Pos

Fidelity : Paraphrase Detection Score (PDS)



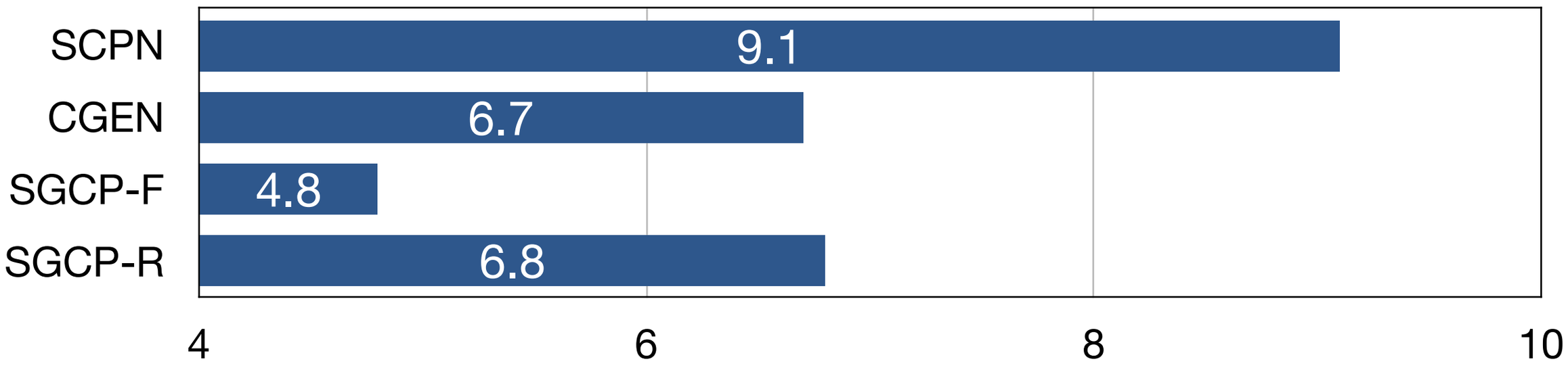
Fidelity : Paraphrase Detection Score (PDS)

High Model-based Semantic Scores (wrt Source Sentence)



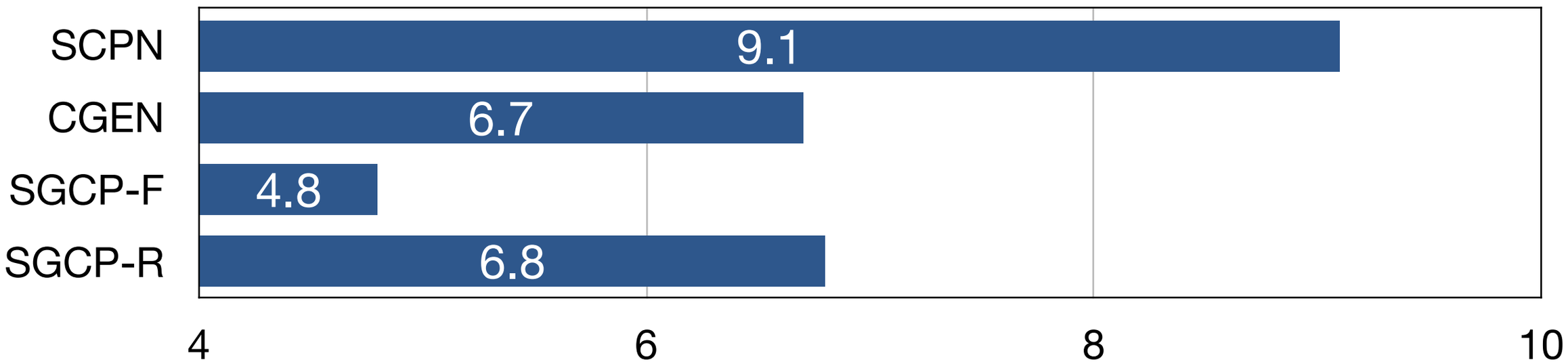
Syntax Conformation QQP-Pos Dataset

Syntax Conformation
QQP-Pos Dataset

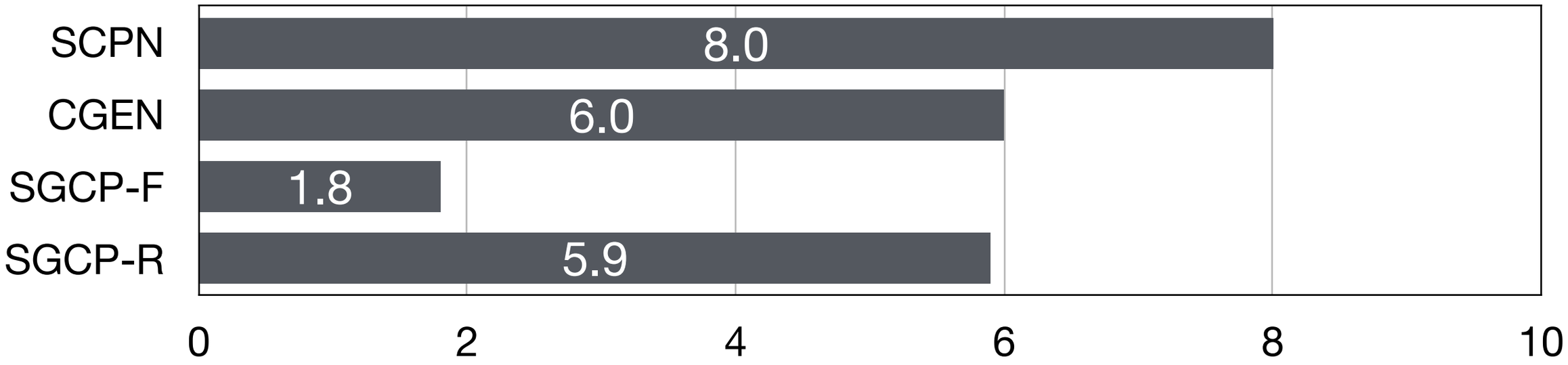


Tree Edit Distance with Reference (Lower is better)

Syntax Conformation
QQP-Pos Dataset

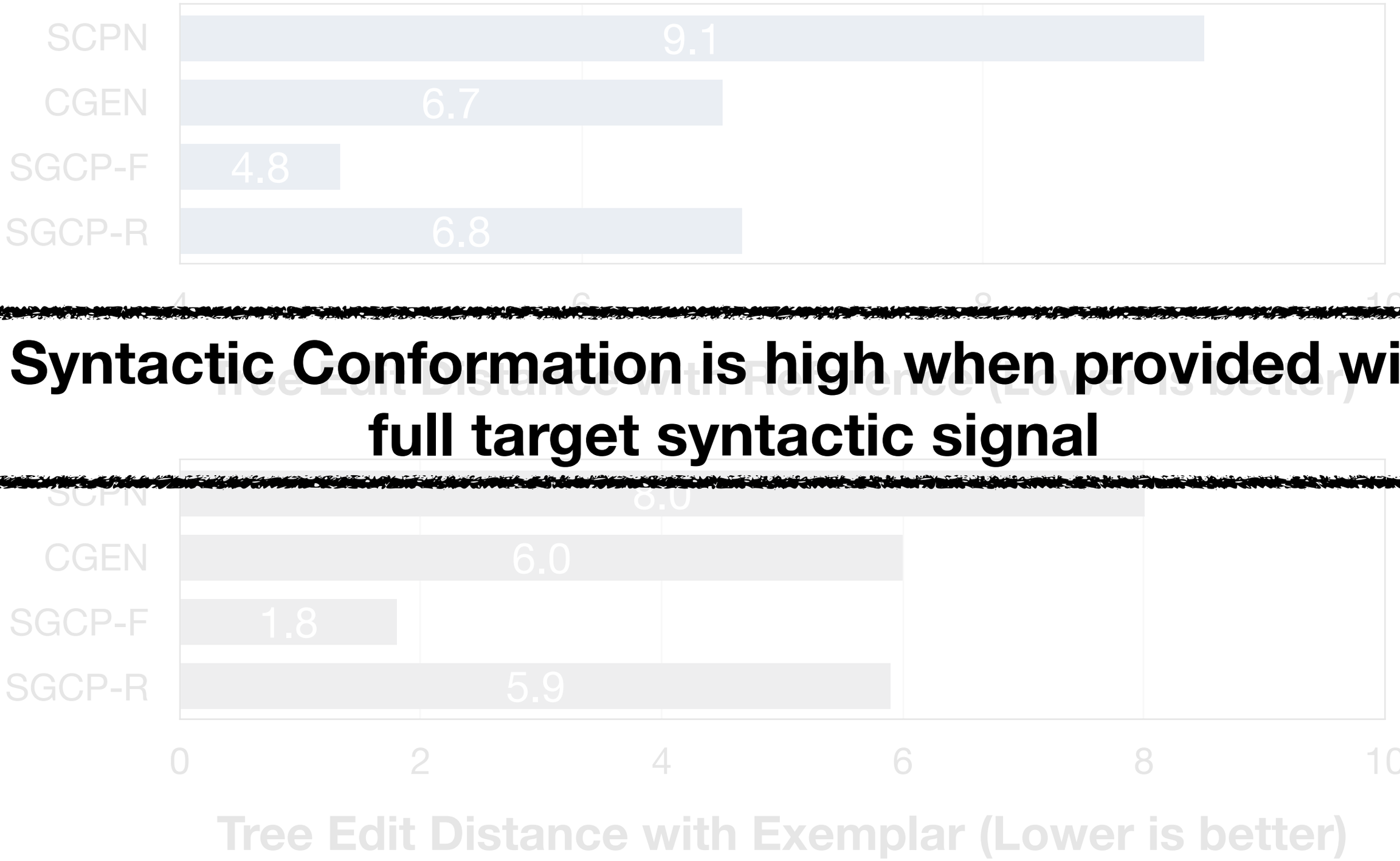


Tree Edit Distance with Reference (Lower is better)



Tree Edit Distance with Exemplar (Lower is better)

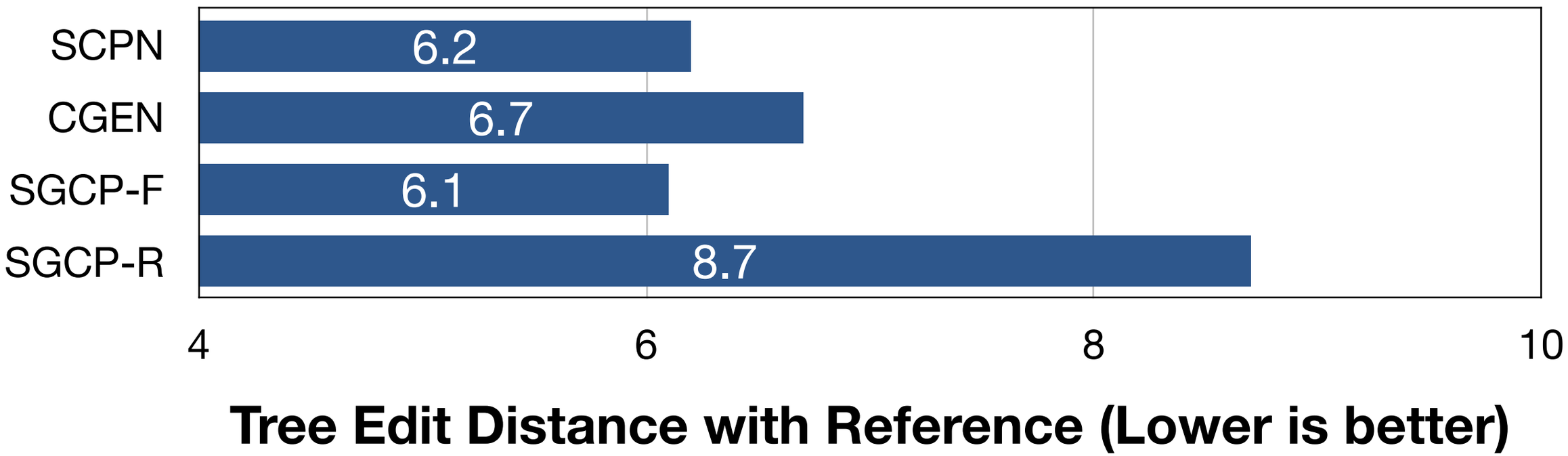
Syntax Conformation QQP-Pos Dataset



Syntactic Conformation is high when provided with full target syntactic signal

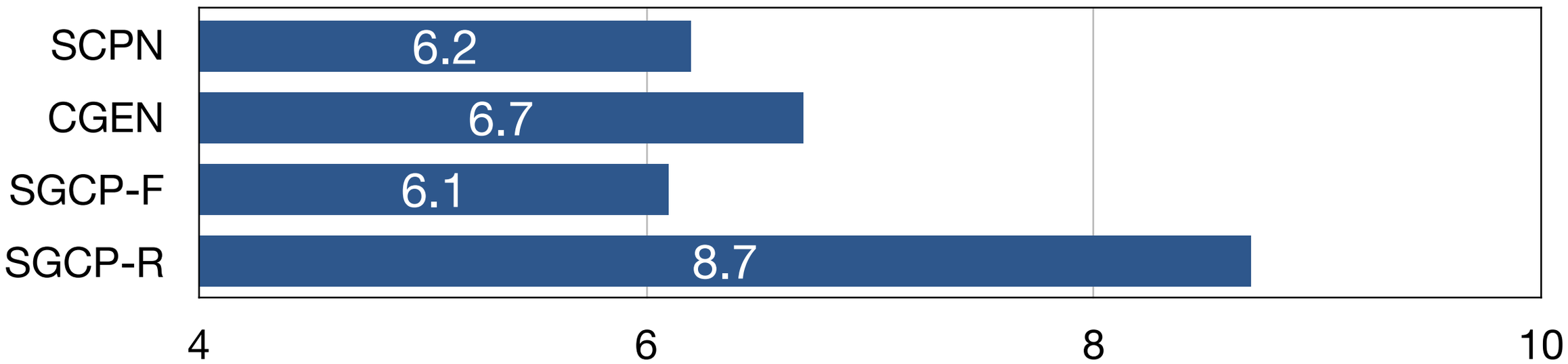
Syntax Conformation ParaNMT-small Dataset

Syntax Conformation ParaNMT-small Dataset

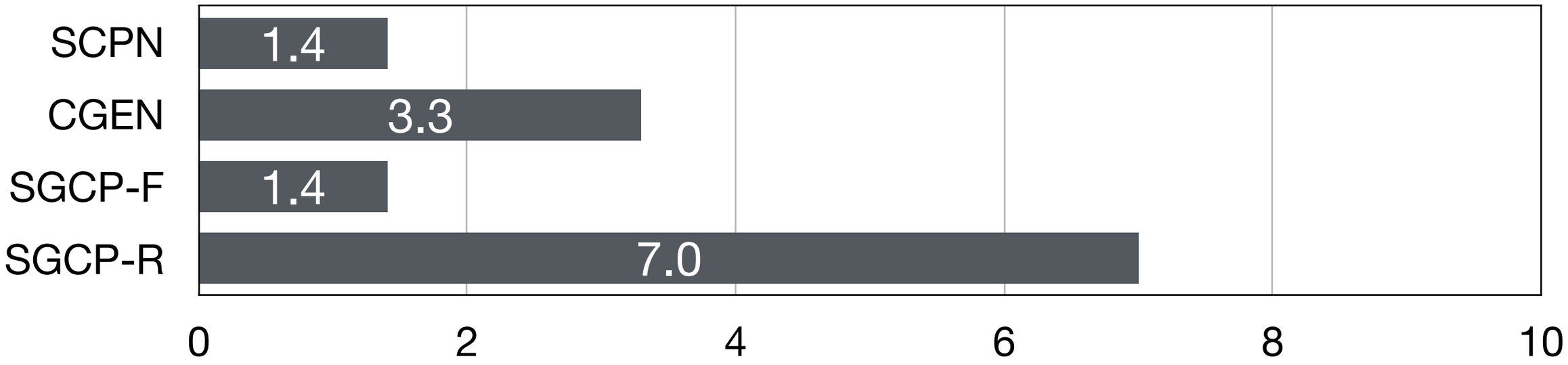


Tree Edit Distance with Reference (Lower is better)

Syntax Conformation ParaNMT-small Dataset

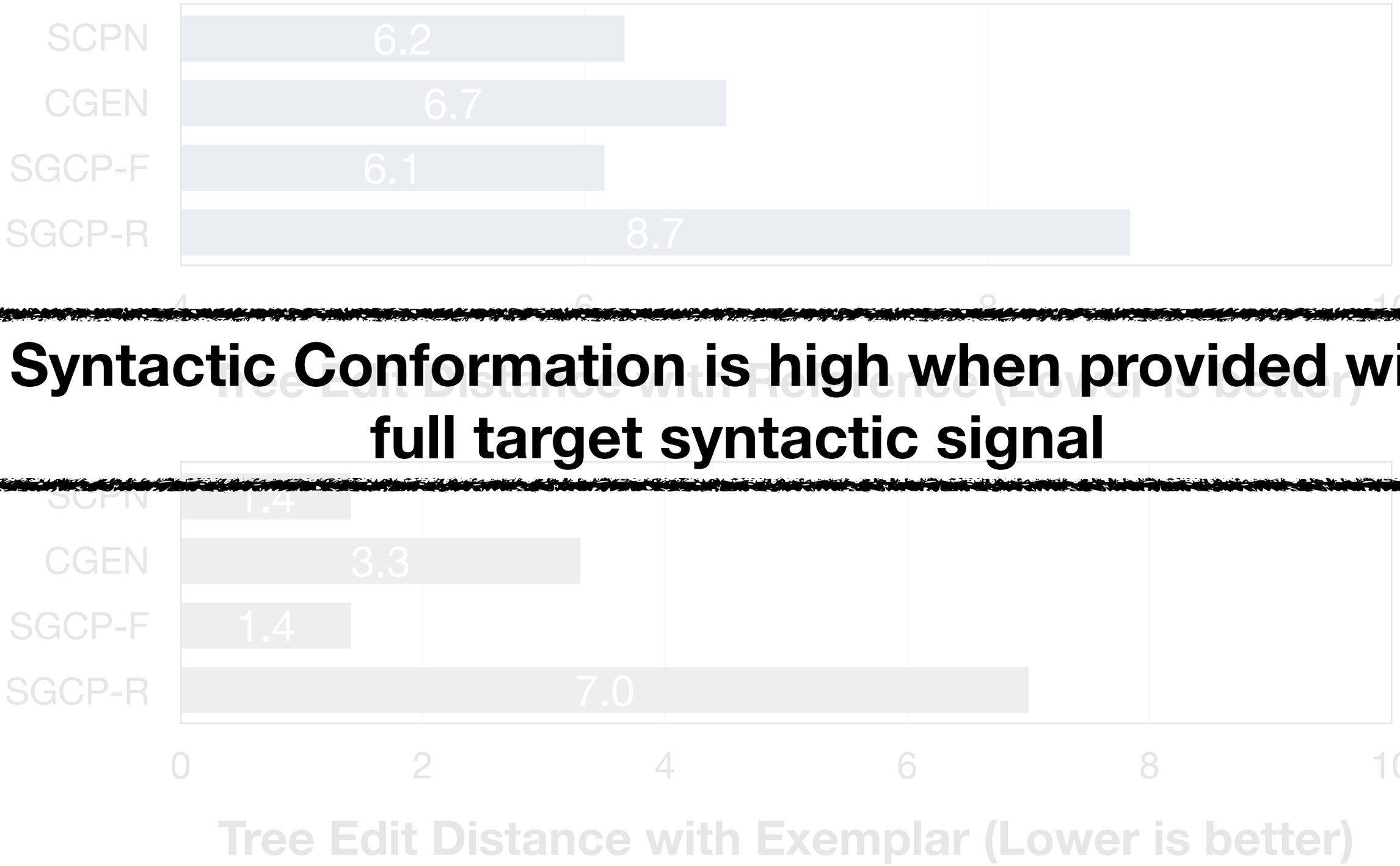


Tree Edit Distance with Reference (Lower is better)



Tree Edit Distance with Exemplar (Lower is better)

Syntax Conformation ParaNMT-small Dataset



Syntactic Conformation is high when provided with full target syntactic signal

Syntactically Diverse Exemplar Inputs

SOURCE : how do i develop my career in software ?	
SYNTACTIC EXEMPLAR	**SGCP-R GENERATIONS**
how can i get a domain for free ?	how can i develop a career in software ?
what is the best way to register a company ?	what is the best way to develop career in software ?
what is chromosomal mutation ? what are some examples ?	what is a good career ? what are some of the ways to develop my career in software ?

Syntactically Diverse Exemplar Inputs

SOURCE : how do i develop my career in software ?	
SYNTACTIC EXEMPLAR	**SGCP-R GENERATIONS**
how can i get a domain for free ?	how can i develop a career in software ?
what is the best way to register a company ?	what is the best way to develop career in software ?
what is chromosomal mutation ? what are some examples ?	what is a good career ? what are some of the ways to develop my career in software ?

Syntacticality In Paraphrase Generation: Summary

Syntacticality In Paraphrase Generation: Summary

Problem

Syntactically Controlled Generation

While preserving semantics

Syntacticality In Paraphrase Generation: Summary

Problem

Syntactically Controlled Generation

While preserving semantics

Method

SGCP

Guiding Decoder Using Syntactic Signals

Syntacticality In Paraphrase Generation: Summary

Problem

Syntactically Controlled Generation

While preserving semantics

Method

SGCP

Guiding Decoder Using Syntactic Signals

Future Work

Target Syntax Compatibility

Data Augmentation Using Syntactic Paraphrasing

Part 2: Consistency in Paraphrase Detection

Part 2: Consistency in Paraphrase Detection

Addresses the gap in reducing inconsistency in left-to-right and right-to-left Paraphrase detection research

Part 2: Consistency in Paraphrase Detection

Addresses the gap in reducing inconsistency in left-to-right and right-to-left Paraphrase detection research

Findings of ACL 2022

Findings ACL 2022: Striking a Balance: Alleviating Inconsistency in Pre-trained Models for Symmetric Classification Tasks
Kumar and Joshi 2022

Symmetric Task: Paraphrase Detection

Symmetric Task: Paraphrase Detection

X	A **provisional government or a revolutionary** government has been declared several times
Y	A **revolutionary government or a provisional** government has been declared several times

Symmetric Task: Paraphrase Detection

X	A **provisional government or a revolutionary** government has been declared several times
Y	A **revolutionary government or a provisional** government has been declared several times

Model / Input Sequence		
		

Symmetric Task: Paraphrase Detection

X	A **provisional government or a revolutionary** government has been declared several times
Y	A **revolutionary government or a provisional** government has been declared several times

Input Sequence \ Model		
X Y		

Symmetric Task: Paraphrase Detection

X	A **provisional government or a revolutionary** government has been declared several times	
Y	A **revolutionary government or a provisional** government has been declared several times	
Input Sequence / **Model**		
X Y	 (98.6)	

Symmetric Task: Paraphrase Detection

X	A **provisional government or a revolutionary** government has been declared several times
Y	A **revolutionary government or a provisional** government has been declared several times

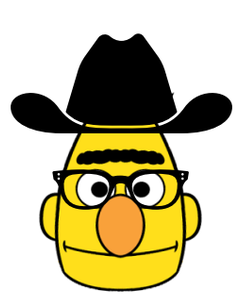
Input Sequence \ Model			
X	Y		
Y	X	 (98.6)	

Symmetric Task: Paraphrase Detection

X	A **provisional government or a revolutionary** government has been declared several times	
Y	A **revolutionary government or a provisional** government has been declared several times	
Input Sequence \ **Model**		
X Y	 (98.6)	
Y X	 (92.2)	

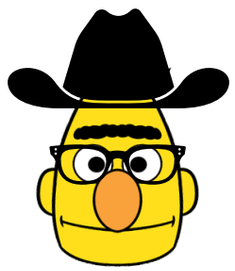
Symmetric Task: Paraphrase Detection

X	A **provisional government or a revolutionary** government has been declared several times
Y	A **revolutionary government or a provisional** government has been declared several times

Input Sequence \ Model		
X Y	 (98.6)	
Y X	 (92.2)	

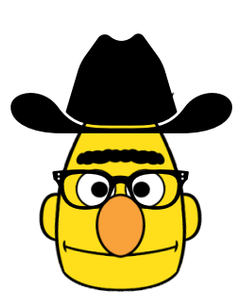
Symmetric Task: Paraphrase Detection

X	A **provisional government or a revolutionary** government has been declared several times
Y	A **revolutionary government or a provisional** government has been declared several times

Input Sequence \ Model		
X Y	 (98.6)	 (88.3)
Y X	 (92.2)	

Symmetric Task: Paraphrase Detection

X	A **provisional government or a revolutionary** government has been declared several times
Y	A **revolutionary government or a provisional** government has been declared several times

Input Sequence \ Model		
X Y	 (98.6)	 (88.3)
Y X	 (92.2)	 (87.9)

Symmetric Task: Paraphrase Detection

X	A **provisional government or a revolutionary** government has been declared several times
Y	A **revolutionary government or a provisional** government has been declared several times

Input Sequence \ Model			
X	Y	(98.6)	(88.3)
Y	X	(92.2)	(87.9)

Erroneous Label and Confidence scores

Symmetric Task: Paraphrase Detection

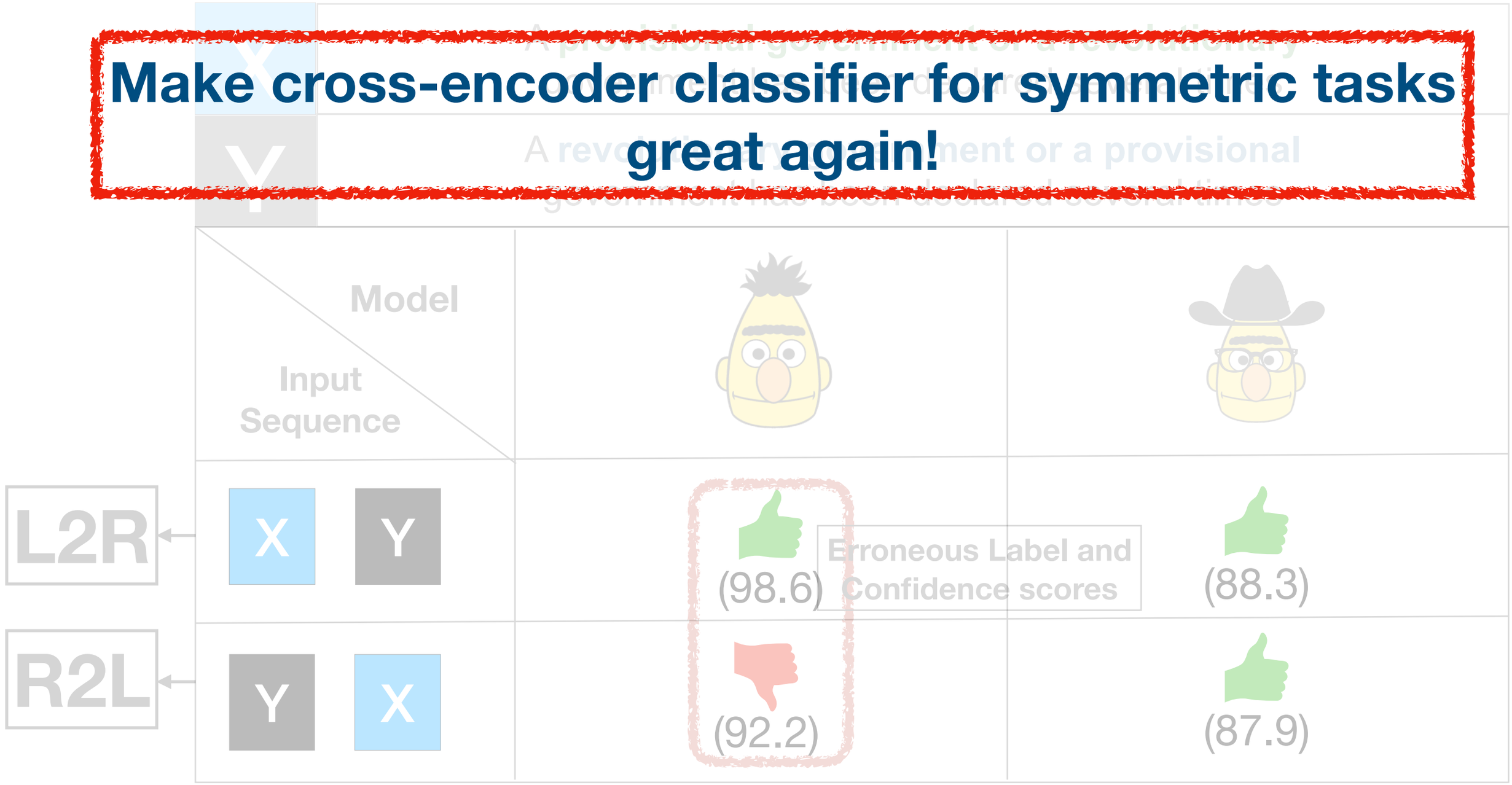
X	A **provisional government or a revolutionary** government has been declared several times
Y	A **revolutionary government or a provisional** government has been declared several times

Model / Input Sequence				
L2R ←	X Y	 (98.6)	(88.3)
R2L ←	Y X	 (92.2)	(87.9)

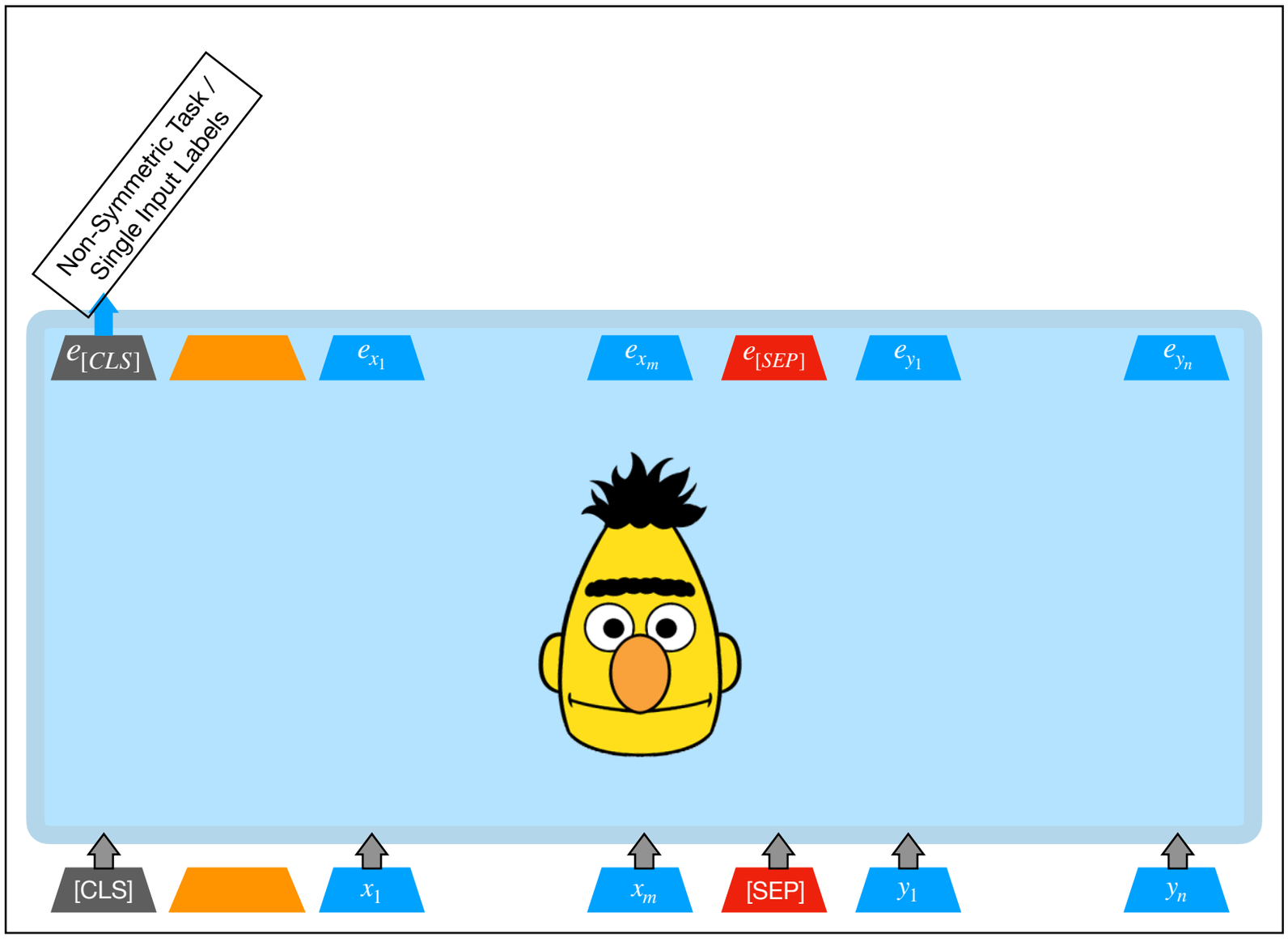
Erroneous Label and Confidence scores

Symmetric Task: Paraphrase Detection

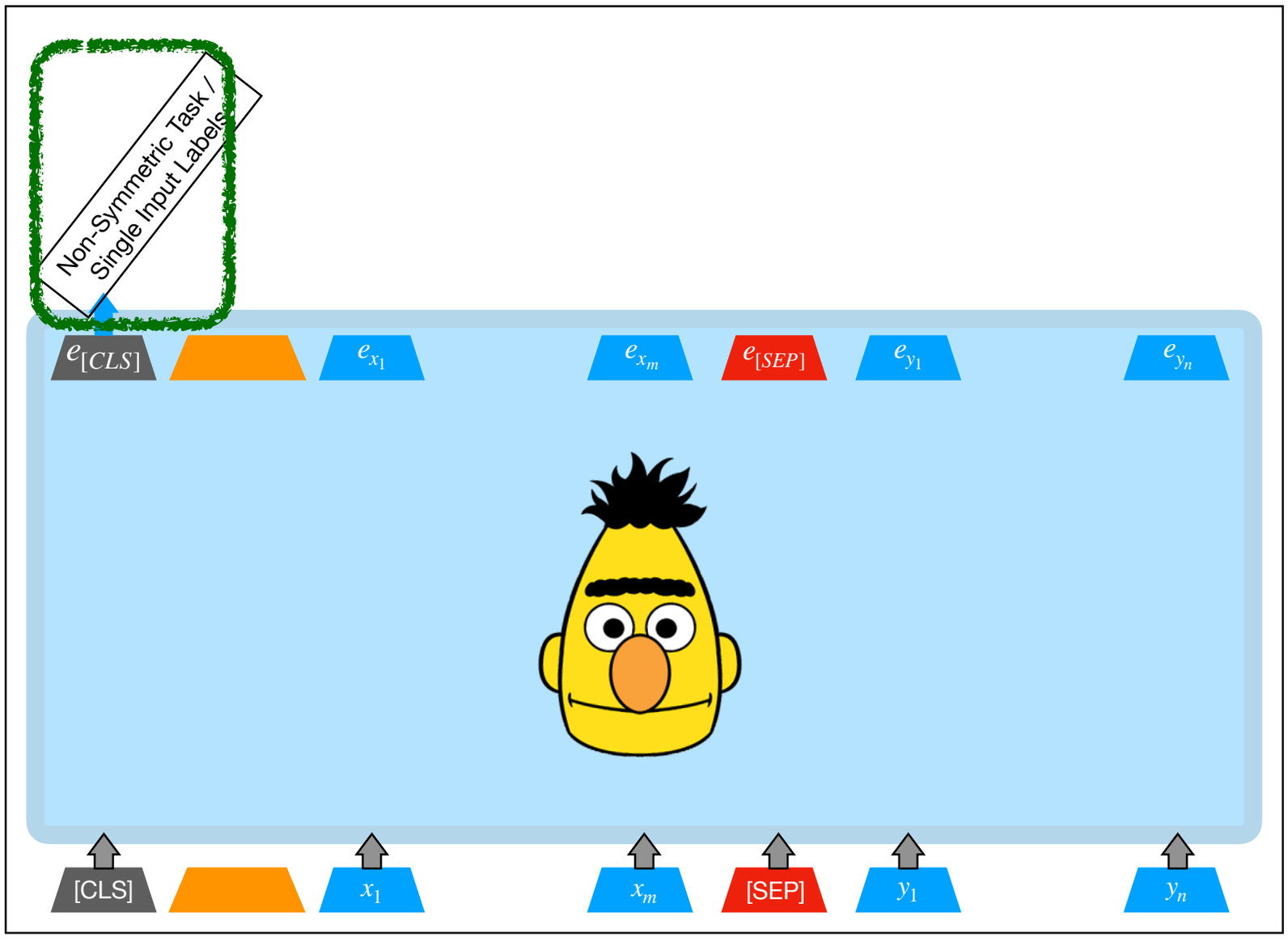
Make cross-encoder classifier for symmetric tasks great again!



Technique
[Cross-encoder Models Only]



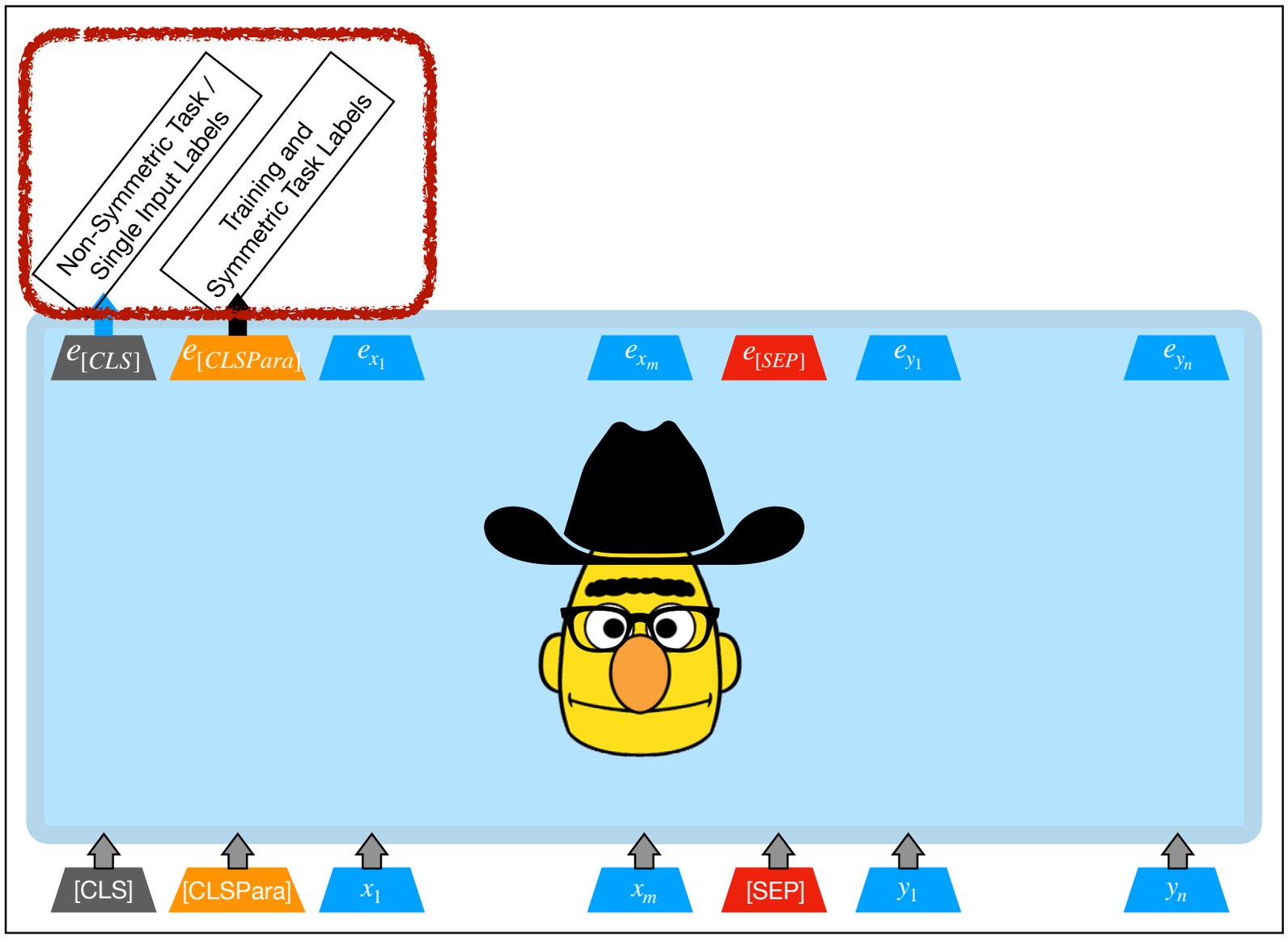
Technique
[Cross-encoder Models Only]



Classic Technique

$$\mathscr{L}_{ce}(y, \hat{y}) = -\sum_i y_i \log \hat{y}_i$$

Technique
[Cross-encoder Models Only]



Classic Technique

$$\mathscr{L}_{ce}(y, \hat{y}) = -\sum_i y_i \log \hat{y}_i$$

Proposal

$$\mathscr{L} = \mathscr{L}_{ce}(y, \hat{y}_{L2R}) + \mathscr{L}_{ce}(y, \hat{y}_{R2L}) + \lambda * \mathscr{D}(p_{L2R} || p_{R2L})$$

$$KL(p||q) = \sum_{x \in X} p(x) \log \frac{p(x)}{q(x)}$$

$$JS(p||q) = \frac{1}{2} KL(p||m) + \frac{1}{2} KL(q||m)$$

Experimental Setup

Experimental Setup

Datasets

Experimental Setup

Datasets

Symmetric Datasets

1. Quora Question Pairs
2. PAWS
3. MRPC

Experimental Setup

Datasets

Symmetric Datasets

1. Quora Question Pairs
2. PAWS
3. MRPC

Non-Symmetric Datasets *(2 stage fine-tune)*

- **Single Sentence**
    1. SST2
- **Pairwise Non-Symmetric**
    2. QNLI
    3. RTE

Experimental Setup

Datasets

Symmetric Datasets

1. Quora Question Pairs
2. PAWS
3. MRPC

Non-Symmetric Datasets *(2 stage fine-tune)*

- **Single Sentence**
    1. SST2
- **Pairwise Non-Symmetric**
    2. QNLI
    3. RTE

Evaluation

Experimental Setup

Datasets

Symmetric Datasets

1. Quora Question Pairs
2. PAWS
3. MRPC

Non-Symmetric Datasets *(2 stage fine-tune)*

- **Single Sentence**
    1. SST2
- **Pairwise Non-Symmetric**
    2. QNLI
    3. RTE

Evaluation

Prediction Consistency

$$\text{Score} = \frac{\mathbf{1}_{(l_{L2R}=l_{R2L})}}{(\#\text{ of } L2R \text{ Samples})} * 100$$

Experimental Setup

Datasets

Symmetric Datasets

1. Quora Question Pairs
2. PAWS
3. MRPC

Non-Symmetric Datasets *(2 stage fine-tune)*

- **Single Sentence**
    1. SST2
- **Pairwise Non-Symmetric**
    2. QNLI
    3. RTE

Evaluation

Prediction Consistency

$$\text{Score} = \frac{\mathbf{1}_{(l_{L2R}=l_{R2L})}}{(\#\text{ of } L2R \text{ Samples})} * 100$$

Confidence Consistency

Mean Squared Error and Pearson Correlation

Experimental Setup

Datasets

Symmetric Datasets

1. Quora Question Pairs
2. PAWS
3. MRPC

Non-Symmetric Datasets *(2 stage fine-tune)*

- **Single Sentence**
    1. SST2
- **Pairwise Non-Symmetric**
    2. QNLI
    3. RTE

Evaluation

Prediction Consistency

$$\text{Score} = \frac{\mathbf{1}_{(l_{L2R}=l_{R2L})}}{(\#\text{ of } L2R \text{ Samples})} * 100$$

Confidence Consistency

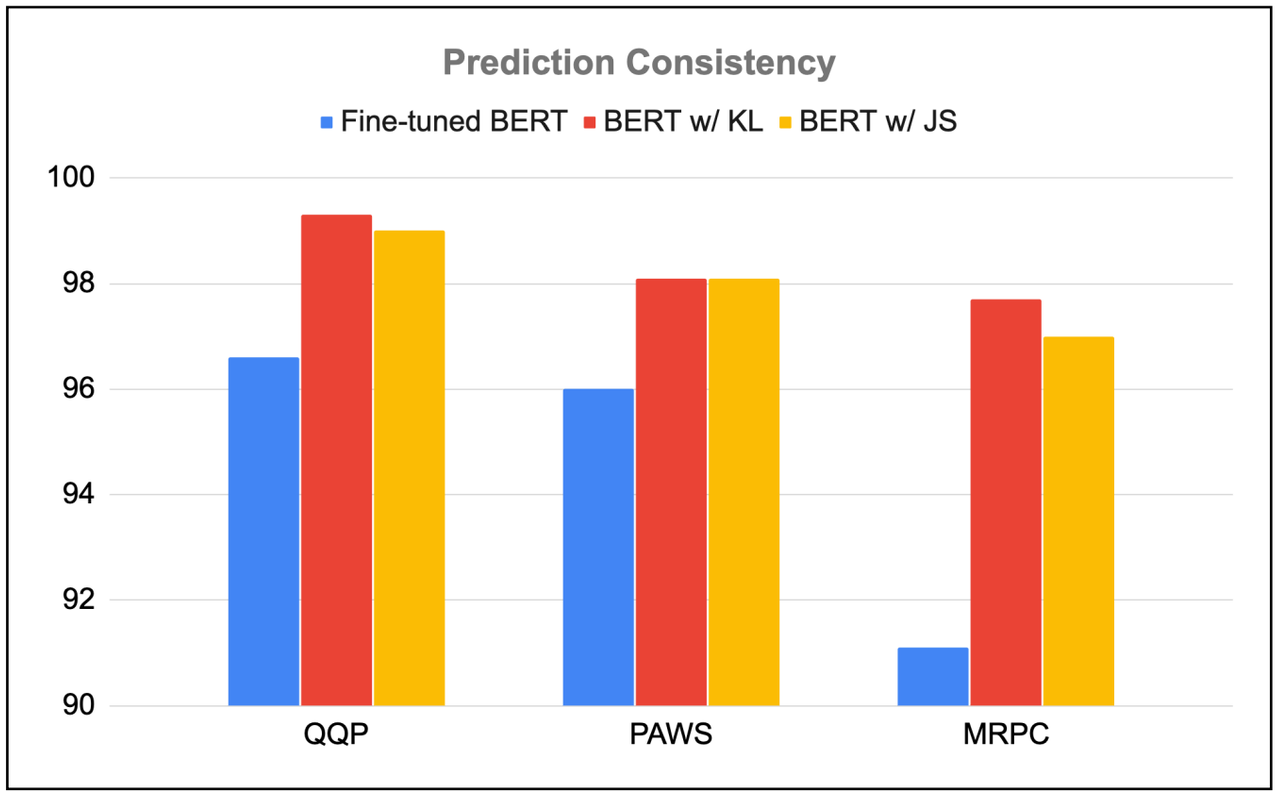
Mean Squared Error and Pearson Correlation

Classification Performance

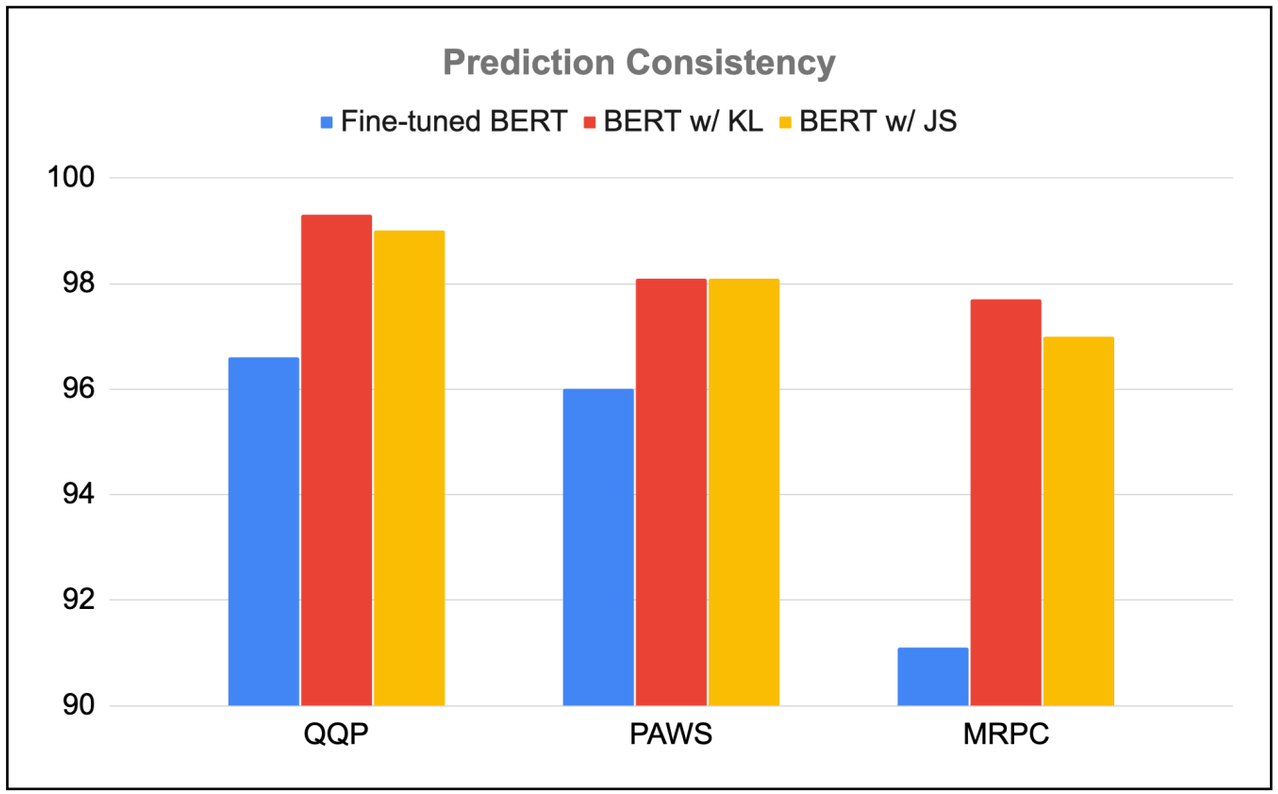
Accuracy, F1

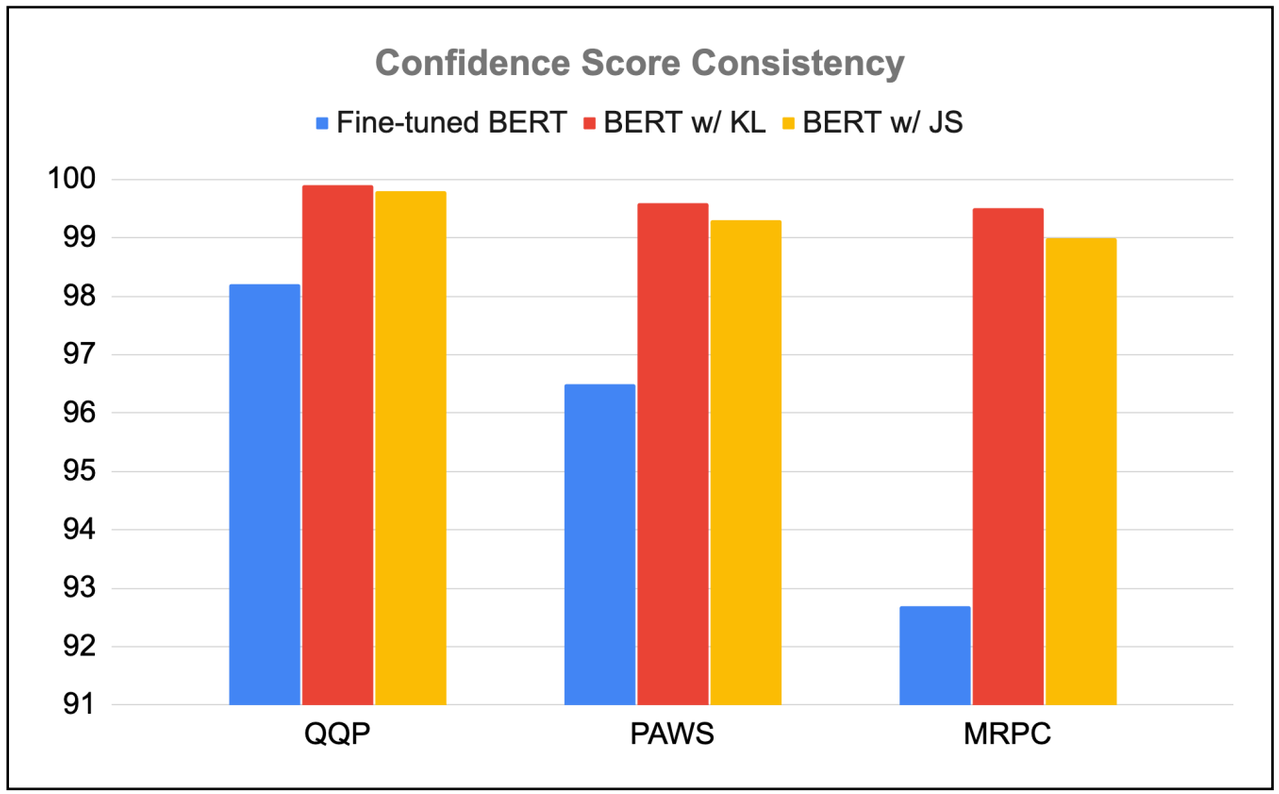
Primary Quantitative Results

Primary Quantitative Results

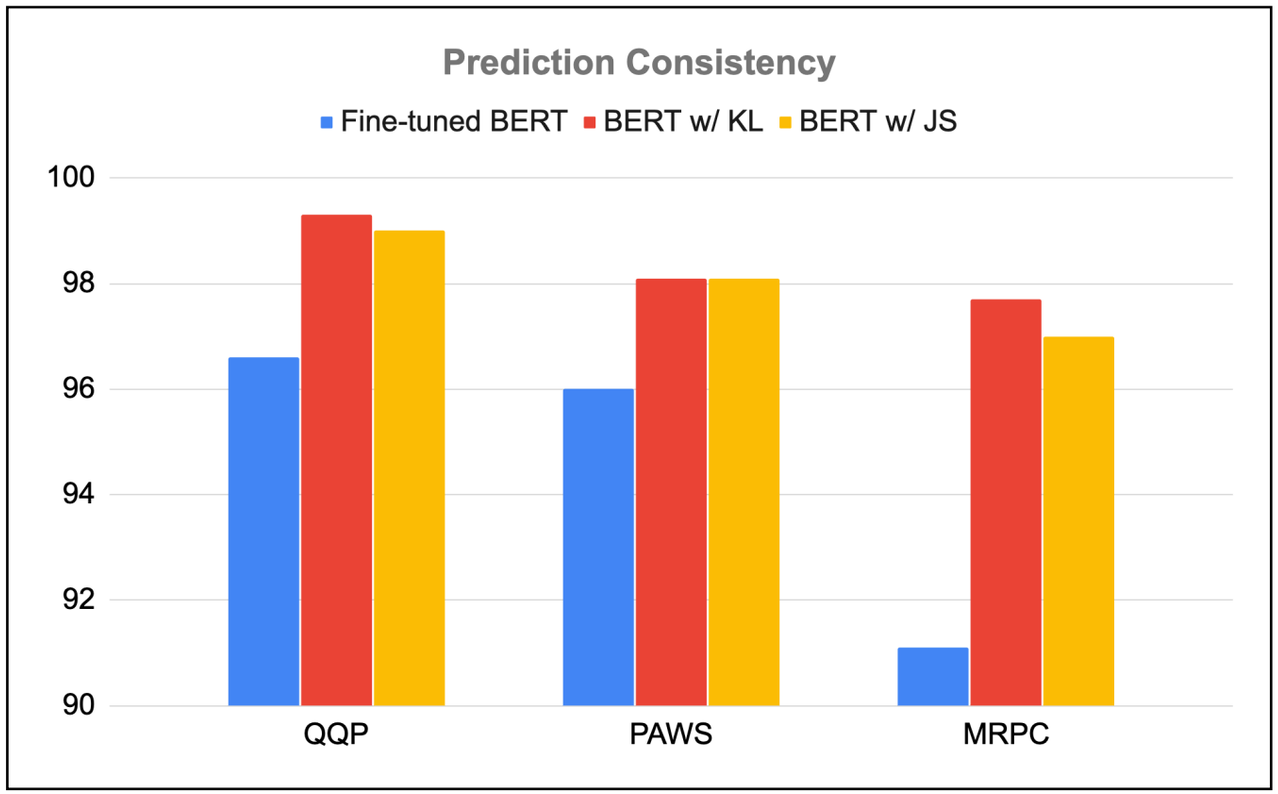


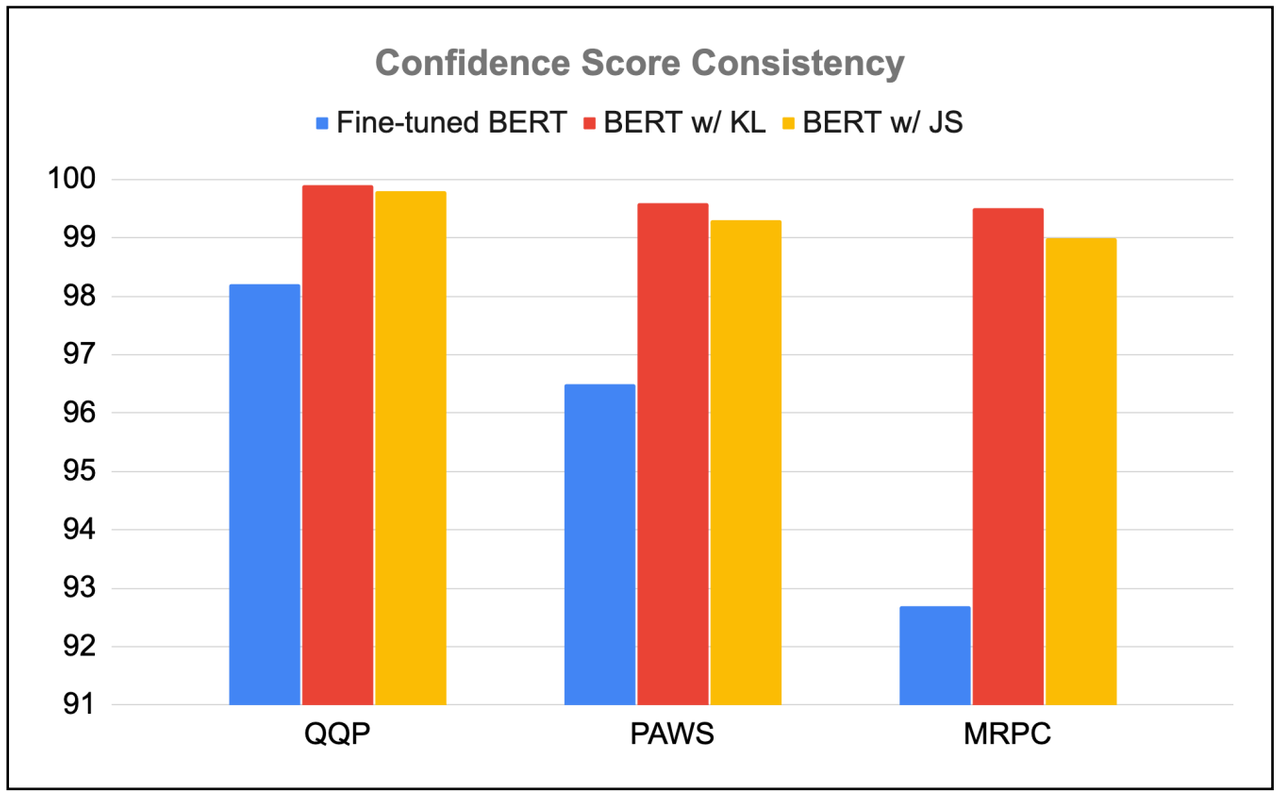
Primary Quantitative Results





Primary Quantitative Results





Results on other tasks can be found in the paper

Qualitative Error Analysis

Qualitative Error Analysis

Additional Details Missing	
X	‘The caretaker, identified by church officials as Jorge Manzon, was believed to be **among the nine missing - some of them children**’
Y	‘The caretaker, identified by church officials as Jorge Monzon, was believed to be among the missing, **who are presumed dead**’.

Qualitative Error Analysis

Additional Details Missing

X	‘The caretaker, identified by church officials as Jorge Manzon, was believed to be **among the nine missing - some of them children**’
Y	‘The caretaker, identified by church officials as Jorge Monzon, was believed to be among the missing, **who are presumed dead**’.

} **11/13** mispredictions fixed by

Qualitative Error Analysis

Additional Details Missing

X	‘The caretaker, identified by church officials as Jorge Manzon, was believed to be **among the nine missing - some of them children**’
Y	‘The caretaker, identified by church officials as Jorge Monzon, was believed to be among the missing, **who are presumed dead**’.

11/13 mispredictions fixed by

Different Answers Expected

X	Is consciousness possible without self-awareness?
Y	Is self-awareness possible without consciousness?

Qualitative Error Analysis

Additional Details Missing

X	‘The caretaker, identified by church officials as Jorge Manzon, was believed to be **among the nine missing - some of them children**’
Y	‘The caretaker, identified by church officials as Jorge Monzon, was believed to be among the missing, **who are presumed dead**’.

11/13 mispredictions fixed by

Different Answers Expected

X	Is consciousness possible without self-awareness?
Y	Is self-awareness possible without consciousness?

4/4 mispredictions fixed by

Qualitative Error Analysis

Additional Details Missing

X	‘The caretaker, identified by church officials as Jorge Manzon, was believed to be **among the nine missing - some of them children**’
Y	‘The caretaker, identified by church officials as Jorge Monzon, was believed to be among the missing, **who are presumed dead**’.

11/13 mispredictions fixed by

Different Answers Expected

X	Is consciousness possible without self-awareness?
Y	Is self-awareness possible without consciousness?

4/4 mispredictions fixed by

Comprehensive Error Analysis and criteria available in the main paper

Consistency in Paraphrase Detection: Summary

Consistency in Paraphrase Detection: Summary

Problem

Inconsistent performance of **cross-encoder models** on **Symmetric Classification tasks**

Order Dependent predictions for “symmetric tasks”

Inconsistent Prediction Labels
Inconsistent Confidence Scores

Consistency in Paraphrase Detection: Summary

Problem

Inconsistent performance of **cross-encoder models** on **Symmetric Classification tasks**

Order Dependent predictions for “symmetric tasks”

Inconsistent Prediction Labels
Inconsistent Confidence Scores

Method

Additional Consistency Objective

Minimize f-Divergence

Consistency in Paraphrase Detection: Summary

Problem

Inconsistent performance of **cross-encoder models** on **Symmetric Classification tasks**

Order Dependent predictions for “symmetric tasks”

Inconsistent Prediction Labels
Inconsistent Confidence Scores

Method

Additional Consistency Objective

Minimize f-Divergence

Take-Aways

Including task-specific information other than cross-entropy helps the ***cross-encoder model***

More consistent performance on symmetric tasks

No impact on single classification task

Consistency in Paraphrase Detection: Summary

Problem

Inconsistent performance of **cross-encoder models** on **Symmetric Classification tasks**

Order Dependent predictions for “symmetric tasks”

Inconsistent Prediction Labels
Inconsistent Confidence Scores

Method

Additional Consistency Objective

Minimize f-Divergence

Take-Aways

Including task-specific information other than cross-entropy helps the ***cross-encoder model***

More consistent performance on symmetric tasks

No impact on single classification task

Future Work:

1. Experiment on anti-symmetric tasks
2. Incorporating the objective into learned metrics like BLEURT

Summary of the Thesis

Problems in Paraphrasing	Examples	Technique	Key Takeaways
Diversity in paraphrase generation	**Input (X):** - how do i increase body height ? **Output (Y):** - how could I increase my height ? - what should I do to increase my height ? - what are the fastest ways to increase my height ? - is there any proven method to increase height ?	Monotone submodular function maximisation	DiPS model offers high diversity without compromising on fidelity Useful for data augmentation
Syntacticality in paraphrase generation	**Input** **(X):** What are pure substances ? What are some examples ? **Exemplar sentence** **(Z)**: What are the characteristics of the Elizabeth theatre ? **Output** **(Y):** What are the examples of a pure substance ?	TreeLSTM-based paraphrase generation	SGCP was the state-of-the-art syntax-guided paraphrase generation model [154]
Consistency in paraphrase detection	**Input** **X:** a provision government or a revolutionary government has been declared several times by insurgent groups in philippines . **Y:** a provision government or a revolutionary government has been declared several times in philippines by insurgent groups . **Output** **For (X, Y) as input:** 1 (88.3) **For (Y, X) as input:** 1 (87.9)	Minimise f-divergence between L2R and R2L label scores	Reduced inconsistency in confidence scores predicted by pre-trained models

Code Repositories

https://github.com/malllabiisc/DiPS

https://github.com/malllabiisc/SGCP

https://github.com/ashutoshml/alleviating-inconsistency

Code Repositories

https://github.com/malllabiisc/DiPS

https://github.com/malllabiisc/SGCP

https://github.com/ashutoshml/alleviating-inconsistency

Code Repositories

https://github.com/malllabiisc/DiPS

Transformer Version

https://github.com/GEM-benchmark/NL-Augmenter/

https://github.com/malllabiisc/SGCP

https://github.com/ashutoshml/alleviating-inconsistency

Code Repositories

https://github.com/malllabiisc/DiPS

Transformer Version

https://github.com/GEM-benchmark/NL-Augmenter/

https://github.com/malllabiisc/SGCP

Transformer Version

https://github.com/pluslabnlp/aesop

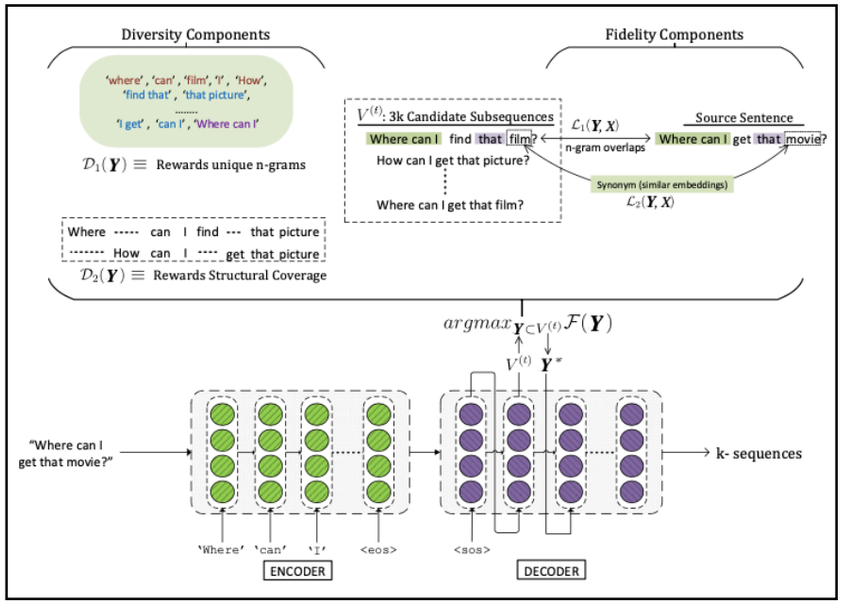
https://github.com/ashutoshml/alleviating-inconsistency

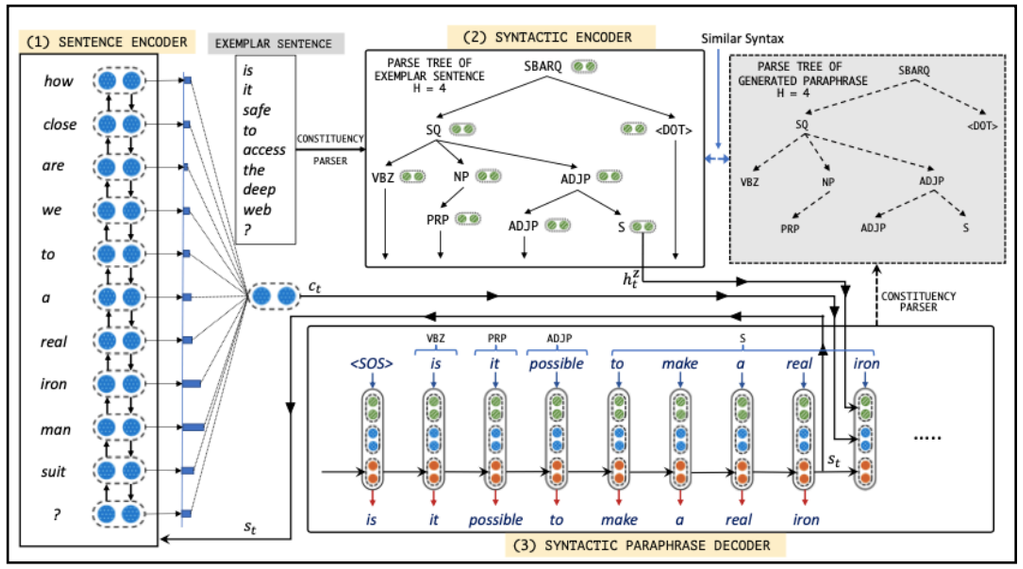
Future Work

Future Work

- Unsupervised Paraphrase Generation
- Text-Style Transfer
- Assessing strengths/weaknesses of models through paraphrases
- Domain Adaptation/Context Aware Paraphrasing
- Query reformulation
- Evaluation metrics for NLG models
- Towards better sentence representations
- Cross-lingual Applications

Inducing Constraints in Paraphrase Generation and Consistency in Paraphrase Detection





Diversity

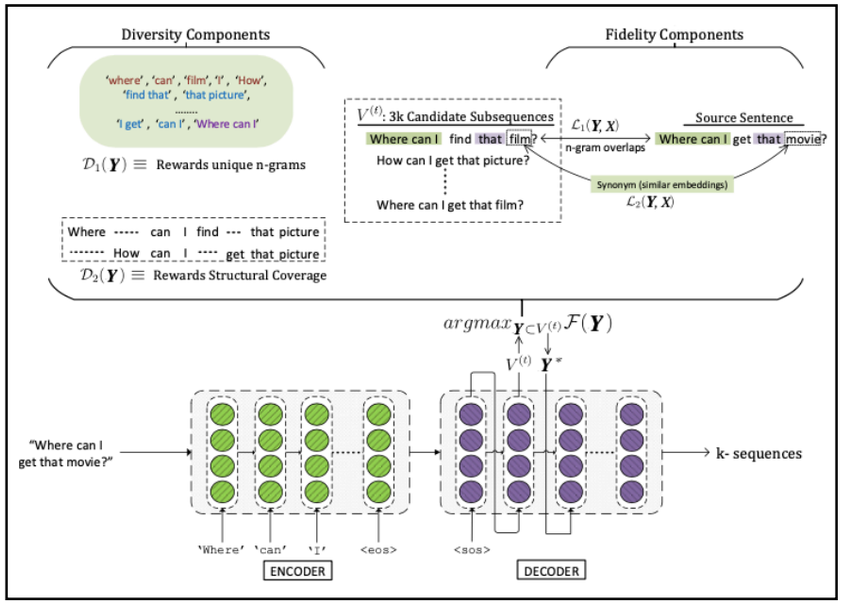
Syntacticality

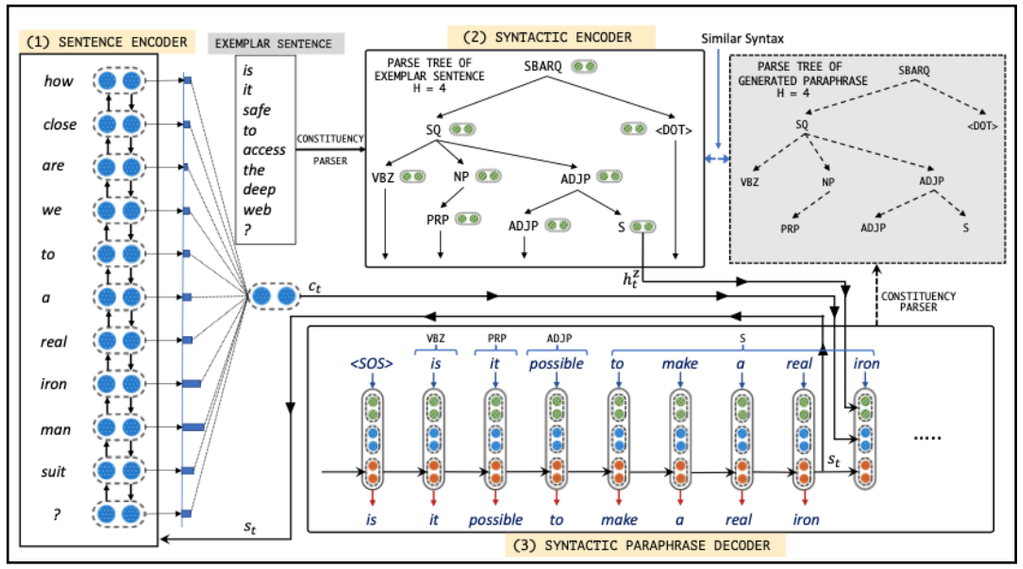
X	A provisional government or a revolutionary government has been declared several times by insurgent groups in the Philippines .
Y	A revolutionary government or a provisional government has been declared several times in the Philippines by insurgent groups .

Input Sequence \ Model		
X Y	(98.6)	(88.3)
Y X	(92.2)	(87.9)

Consistency

Inducing Constraints in Paraphrase Generation and Consistency in Paraphrase Detection





Diversity

Syntacticality

X	A provisional government or a revolutionary government has been declared several times by insurgent groups in the Philippines .
Y	A revolutionary government or a provisional government has been declared several times in the Philippines by insurgent groups .

Input Sequence \ Model		
X Y	(98.6)	(88.3)
Y X	(92.2)	(87.9)

Consistency

ashutosh@iisc.ac.in

References

1. Rahul Bhagat and Eduard Hovy. 2013. Squibs: What Is a Paraphrase?. *Computational Linguistics*, 39(3):463–472.

2. Mohit Iyyer, John Wieting, Kevin Gimpel, and Luke Zettlemoyer. 2018.
   Adversarial Example Generation with Syntactically Controlled Paraphrase Networks.
   In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), pages 1875–1885, New Orleans, Louisiana. Association for Computational Linguistics.

3. Mingda Chen, Qingming Tang, Sam Wiseman, and Kevin Gimpel. 2019.
   Controllable Paraphrase Generation with a Syntactic Exemplar.
   In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5972–5984, Florence, Italy. Association for Computational Linguistics.

4. Ashutosh Kumar*, Satwik Bhattamishra*, Manik Bhandari, and Partha Talukdar. 2019.
   Submodular Optimization-based Diverse Paraphrasing and its Effectiveness in Data Augmentation.
   In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 3609–3619, Minneapolis, Minnesota. Association for Computational Linguistics.

5. Ashutosh Kumar, Kabir Ahuja, Raghuram Vadapalli, and Partha Talukdar. 2020.
   Syntax-Guided Controlled Generation of Paraphrases.
   Transactions of the Association for Computational Linguistics, 8:329–345.

6. Ashutosh Kumar and Aditya Joshi. 2022.
   Striking a Balance: Alleviating Inconsistency in Pre-trained Models for Symmetric Classification Tasks.
   In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 1887–1895, Dublin, Ireland. Association for Computational Linguistics.