

Diverse Paraphrasing and its Effectiveness in Data Augmentation

Ashutosh Kumar*, Satwik Bhattamishra*, Manik Bhandari, Partha Talukdar

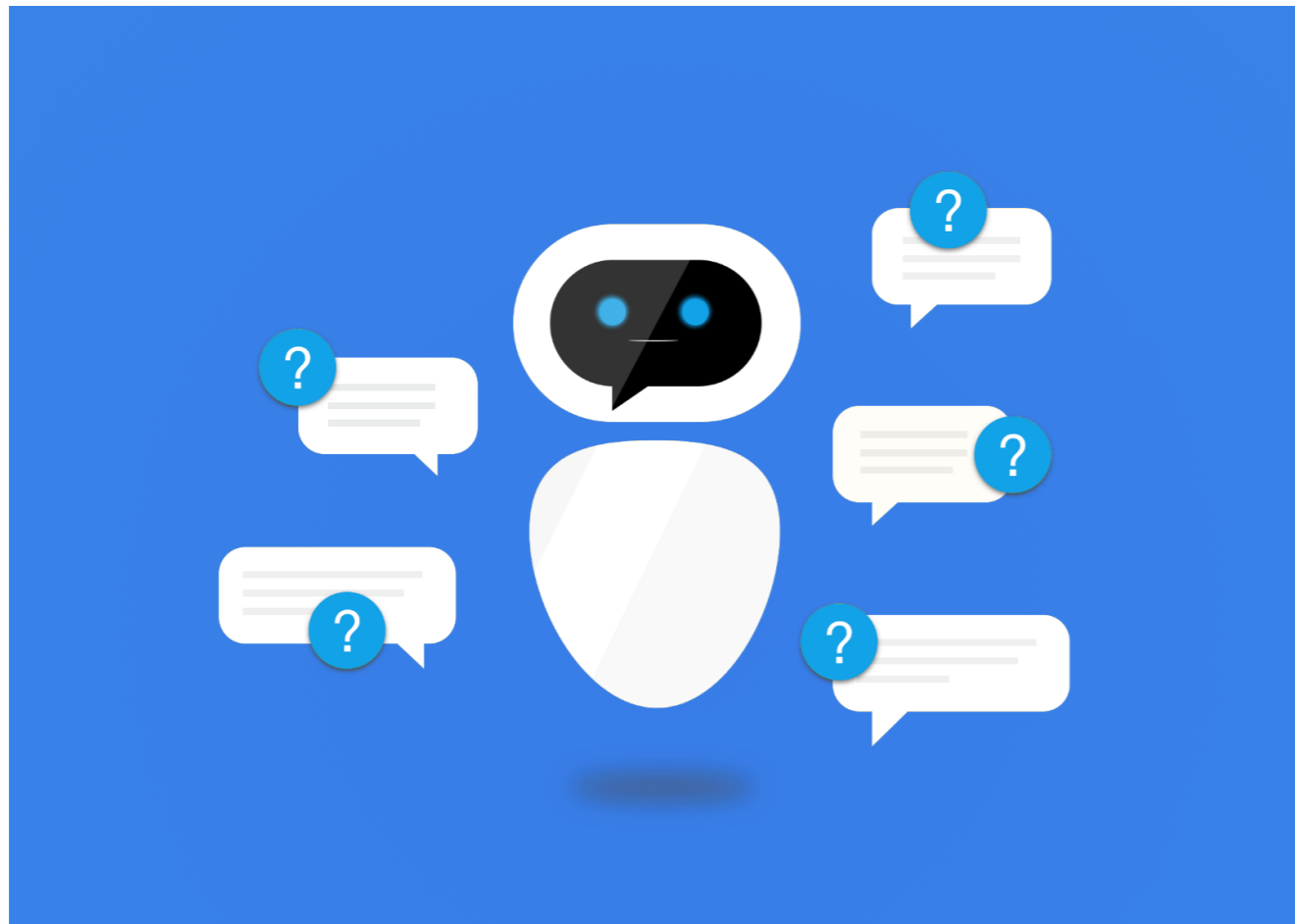
Machine and Language Learning Lab (MALL)
Indian Institute of Science, Bangalore

*Equal Contributions

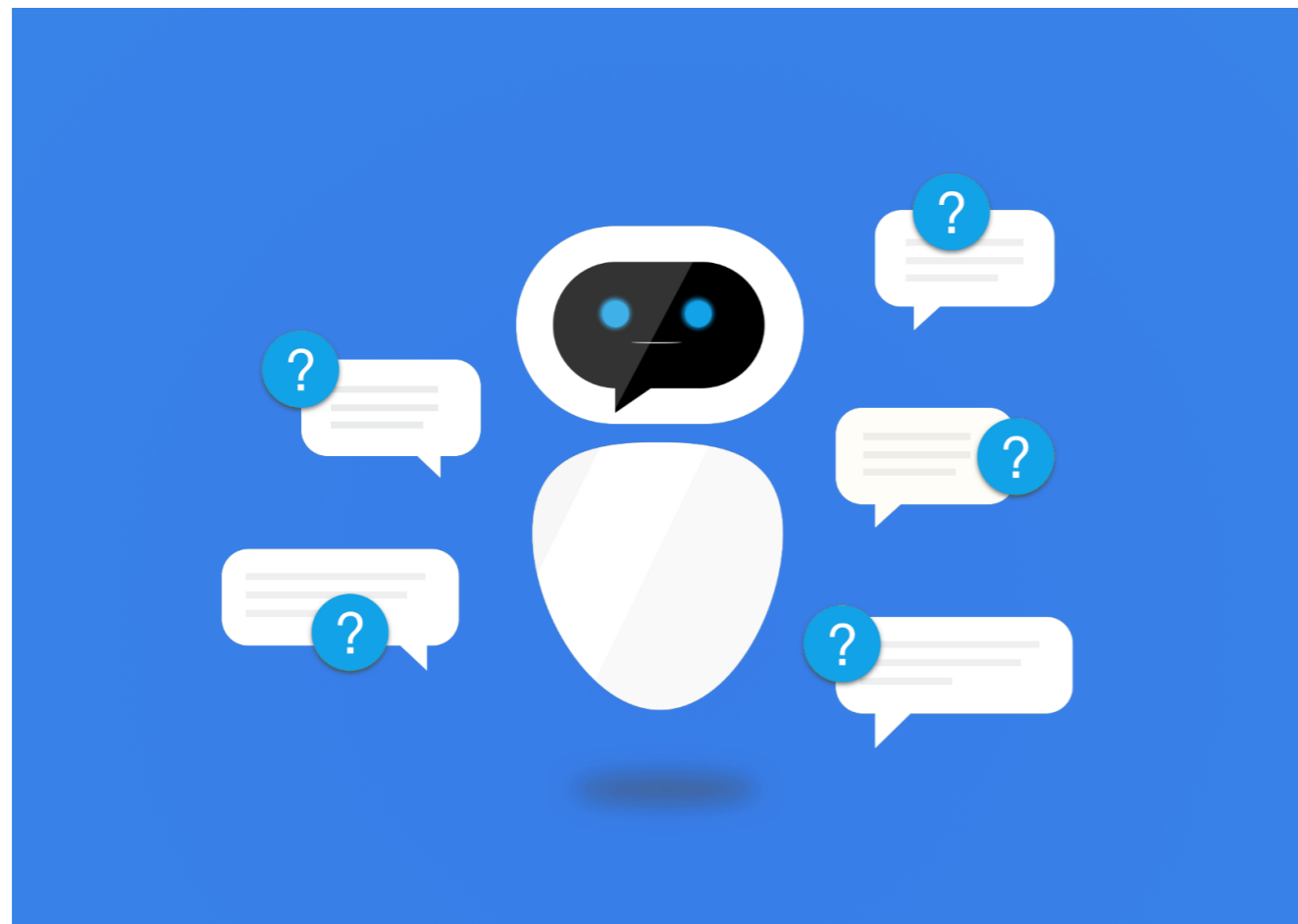


Conversational Agents

Conversational Agents



Conversational Agents



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

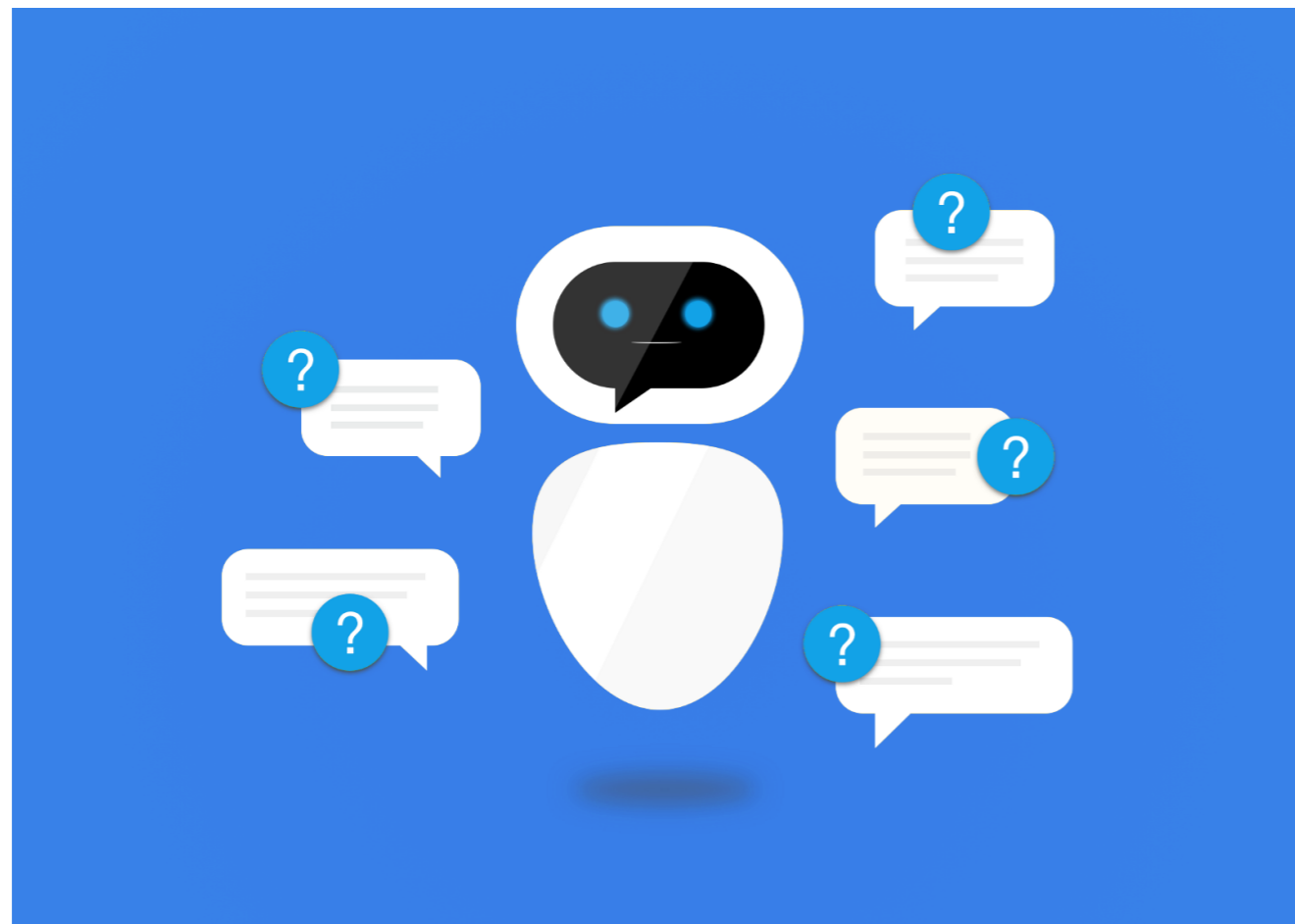
Conversational Agents



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

Bot: Sure. When are you planning to travel ?

Conversational Agents



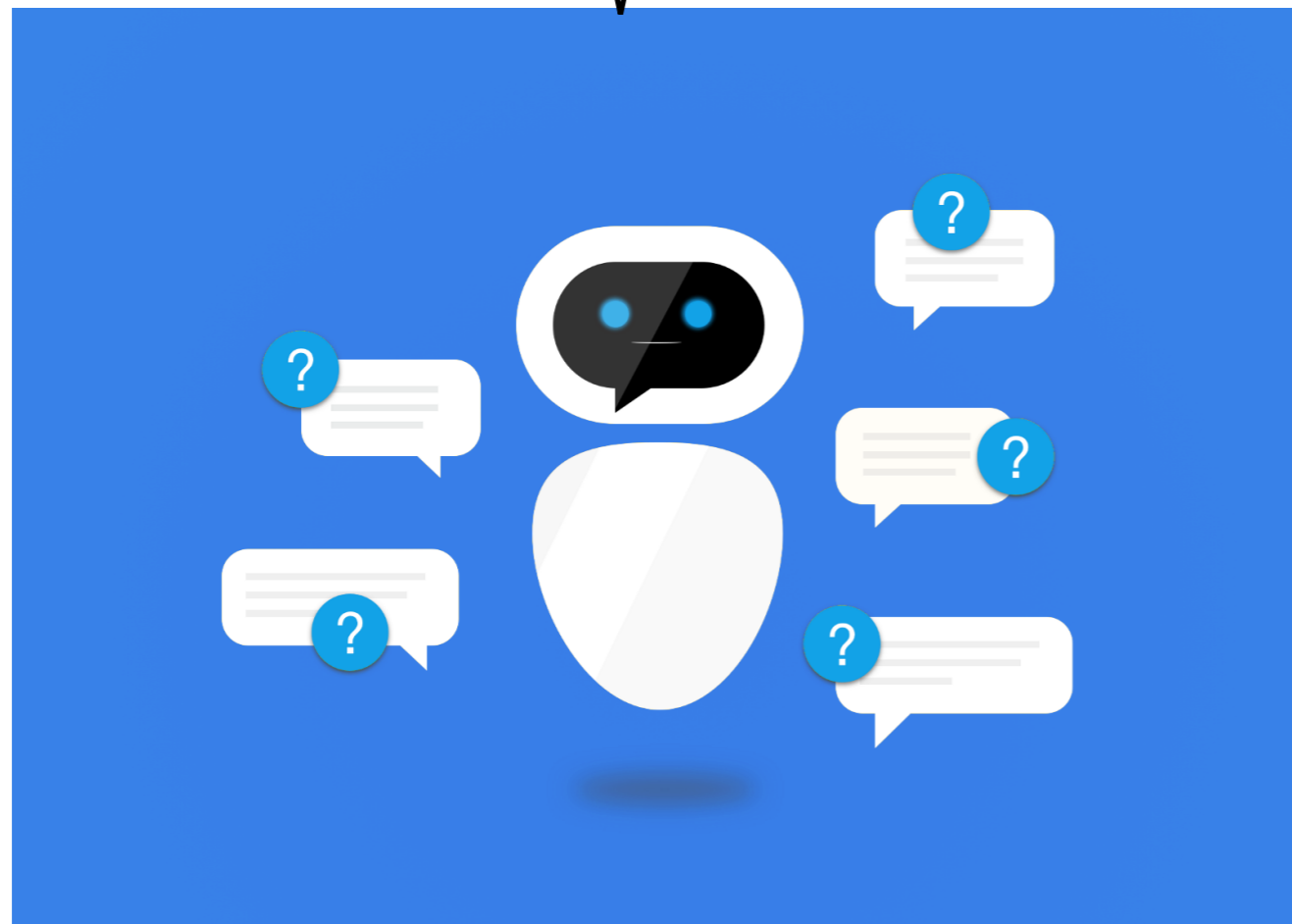
User: I want to book a flight from Minneapolis to New York

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 ?

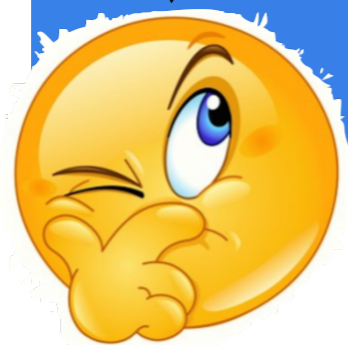
User: Can you book plane to New York from Minneapolis

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	<ul style="list-style-type: none">• 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	<ul style="list-style-type: none">• 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	<ul style="list-style-type: none">• 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)

Current State

Current State

Synonym or phrase replacement

Current State

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

Current State

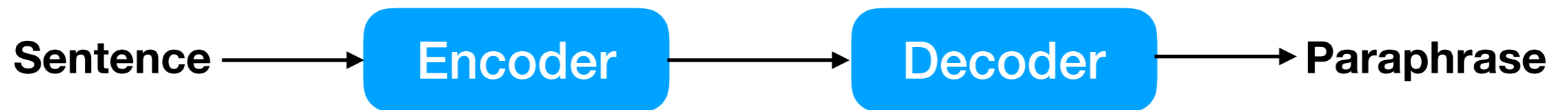
Synonym or phrase replacement	
Source	how do i increase body height ?
Synonym	how do i <u>grow</u> body height ?

Current State

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

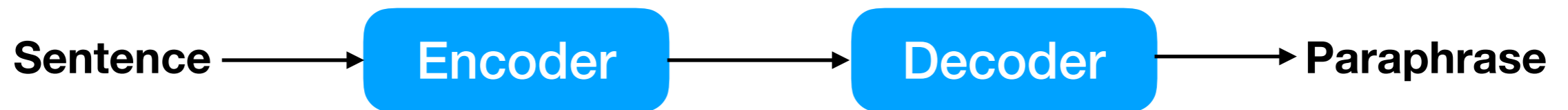
Current State

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



Current State

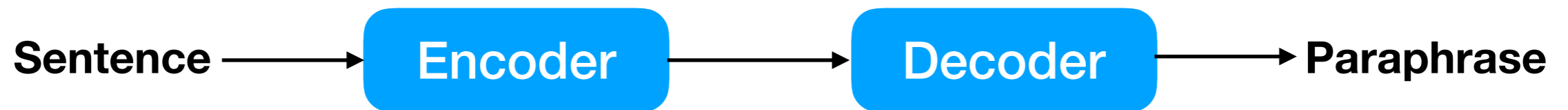
Synonym or phrase replacement	
Source	how do i increase body height ?
Synonym	how do i <u>grow</u> body height ?
Phrase	how do i increase <u>the body measurement vertically</u> ?



Subsequence Selection - Beam Search (Top-k)

Current State

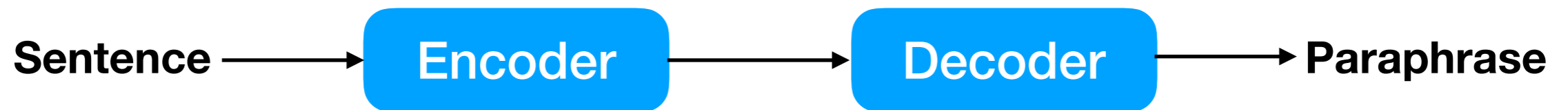
Synonym or phrase replacement	
Source	how do i increase body height ?
Synonym	how do i <u>grow</u> body height ?
Phrase	how do i increase <u>the body measurement vertically</u> ?



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

Current State

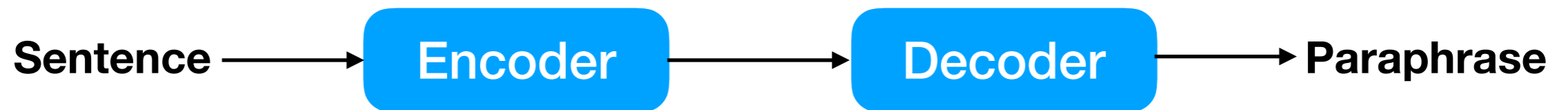
Synonym or phrase replacement	
Source	how do i increase body height ?
Synonym	how do i <u>grow</u> body height ?
Phrase	how do i increase <u>the body measurement vertically</u> ?



Subsequence Selection - Beam Search (Top-k)	
Source	how do i increase body height ?
Beam	<ul style="list-style-type: none">• 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 ?

Current State

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

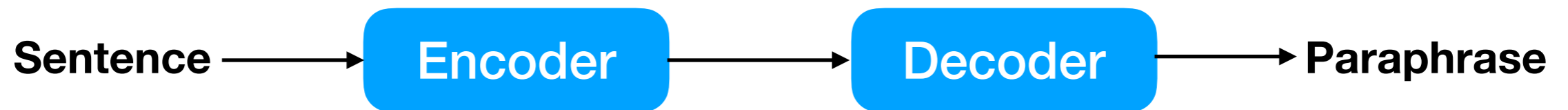


Subsequence Selection - Beam Search (Top-k)	
Source	how do i increase body height ?
Beam	<ul style="list-style-type: none">• how do i increase my height ?• how do i increase my body height ?• how do i increase the height ?• how would i increase my body height ?

Fidelity ✓

Current State

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



Subsequence Selection - Beam Search (Top-k)	
Source	how do i increase body height ?
Beam	<ul style="list-style-type: none">• 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 ✗

What can we do ?

What can we do ?

Subsequence Selection - Beam Search (Diverse selection)

What can we do ?

Subsequence Selection - Beam Search (Diverse selection)

Source

how do i increase body height ?

What can we do ?

Subsequence Selection - Beam Search (Diverse selection)	
Source	how do i increase body height ?
Beam	<ul style="list-style-type: none">• 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 ?

What can we do ?

Subsequence Selection - Beam Search (Diverse selection)	
Source	how do i increase body height ?
Beam	<ul style="list-style-type: none">• 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 ✓

What can we do ?

Subsequence Selection - Beam Search (Diverse selection)	
Source	how do i increase body height ?
Beam	<ul style="list-style-type: none">• 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

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 ...

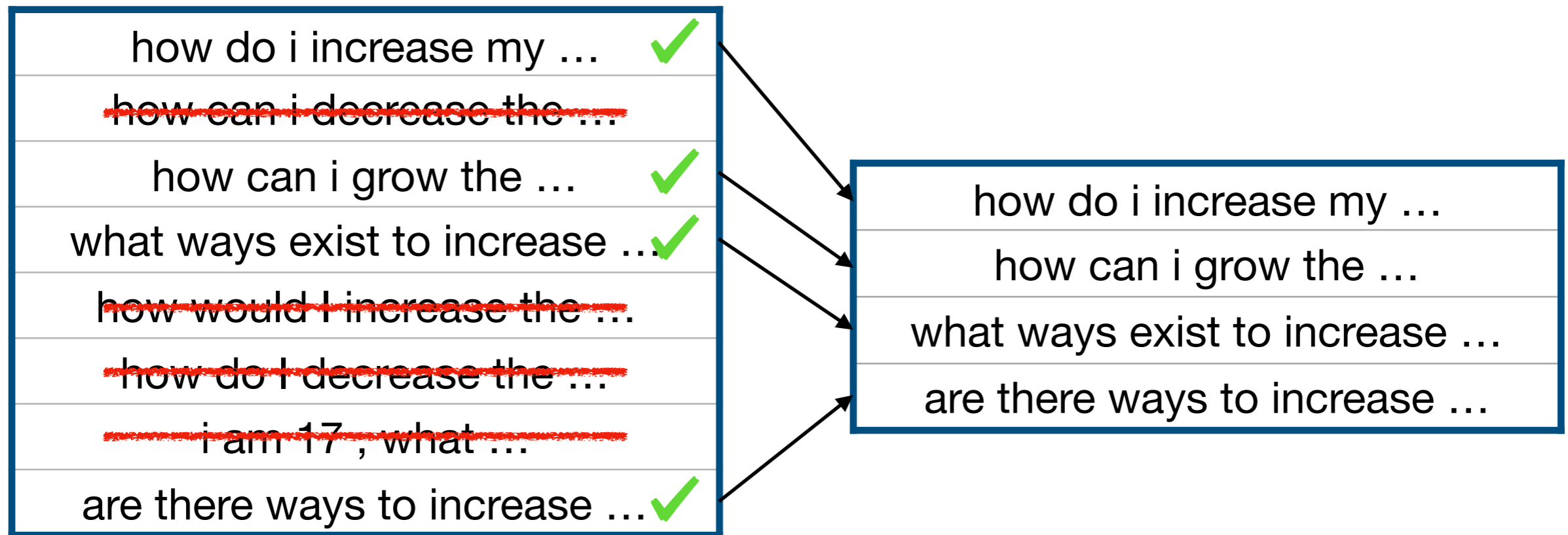
how do I decrease the ...

i am 17 , what ...

are there ways to increase ...

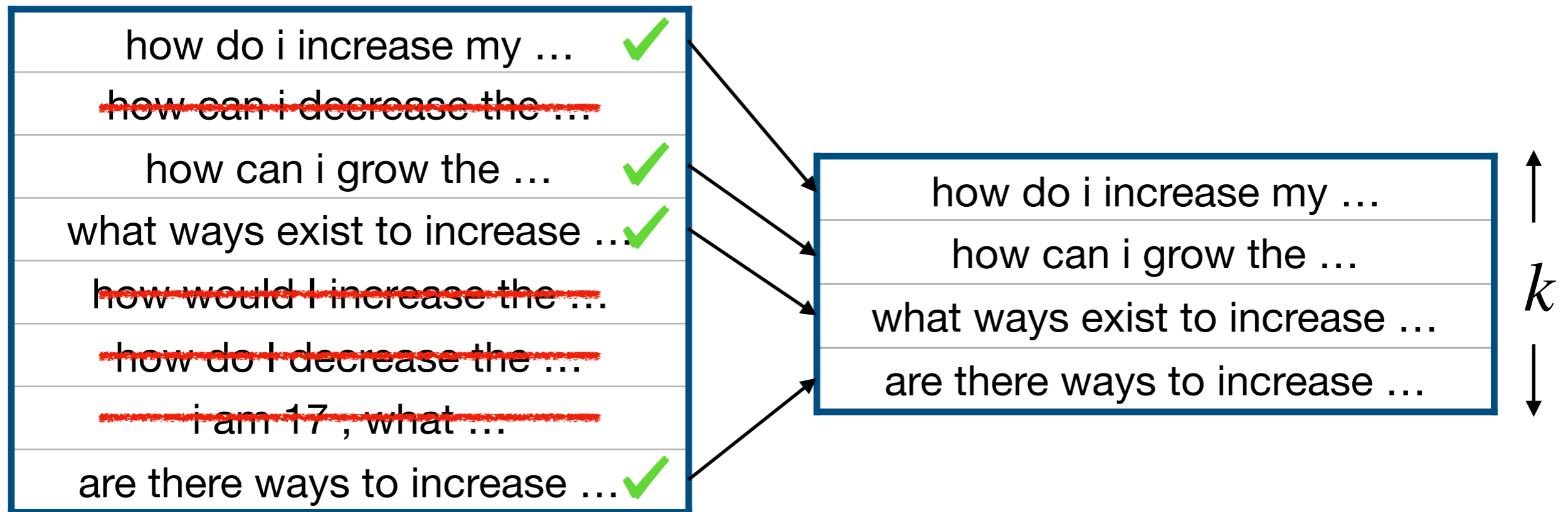
V^t

Subset Selection



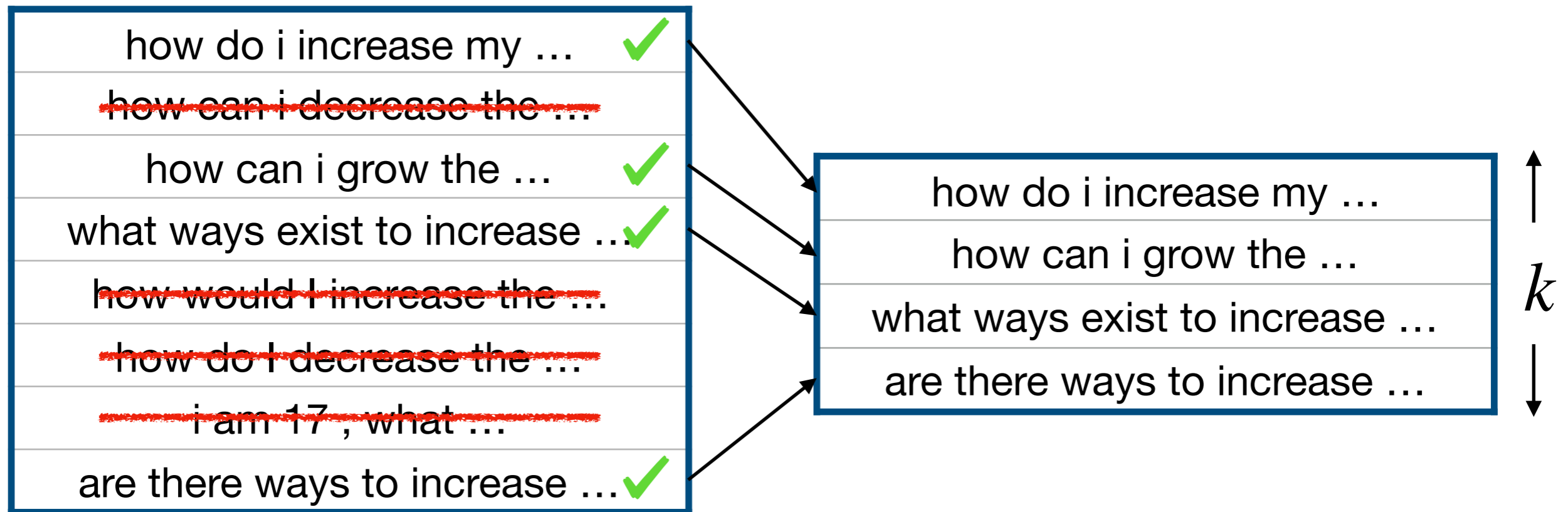
$$V^t \longrightarrow X$$

Subset Selection



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

Subset Selection



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

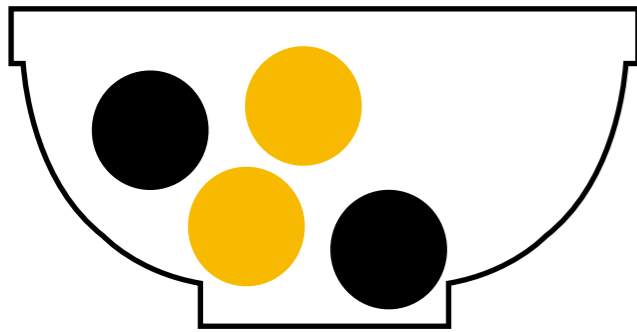
If F is sub modular + monotone = Greedy algo. with good bounds exists

Sub-modularity

F = # Unique Coloured items

Sub-modularity

F = # Unique Coloured items

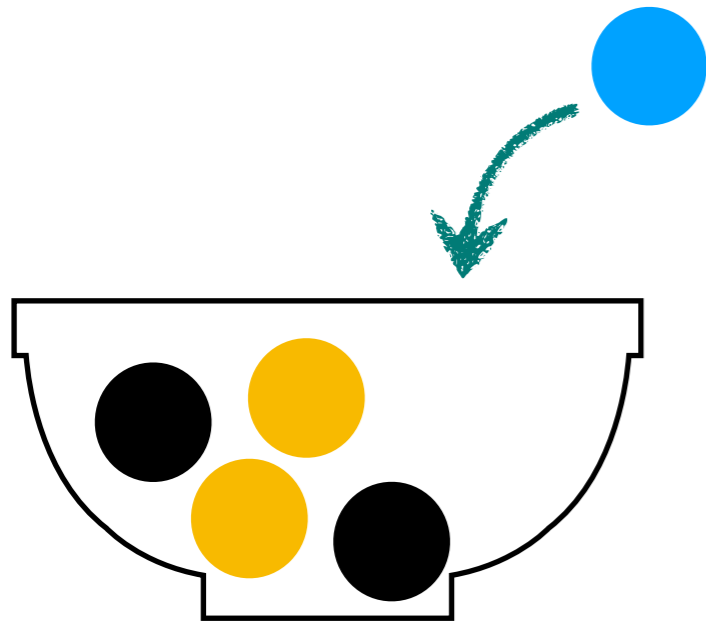


Items = 4

F = 2

Sub-modularity

F = # Unique Coloured items

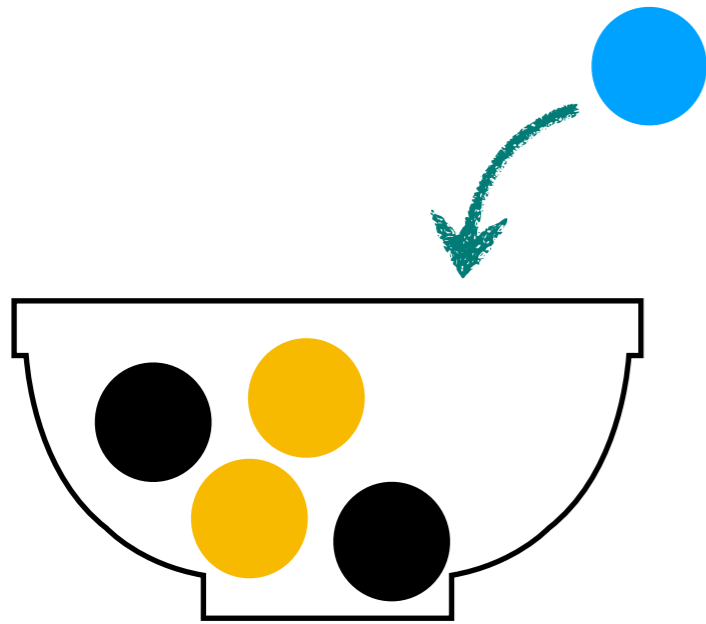


Items = 4

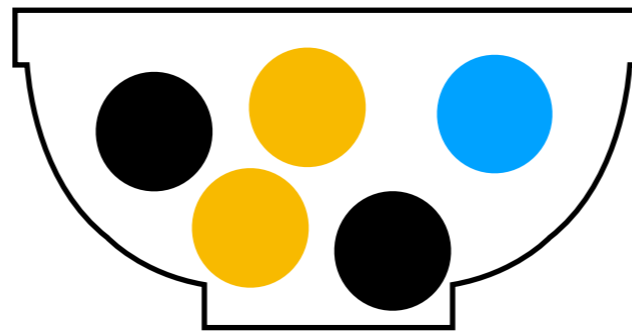
F = 2

Sub-modularity

$F = \#$ Unique Coloured items



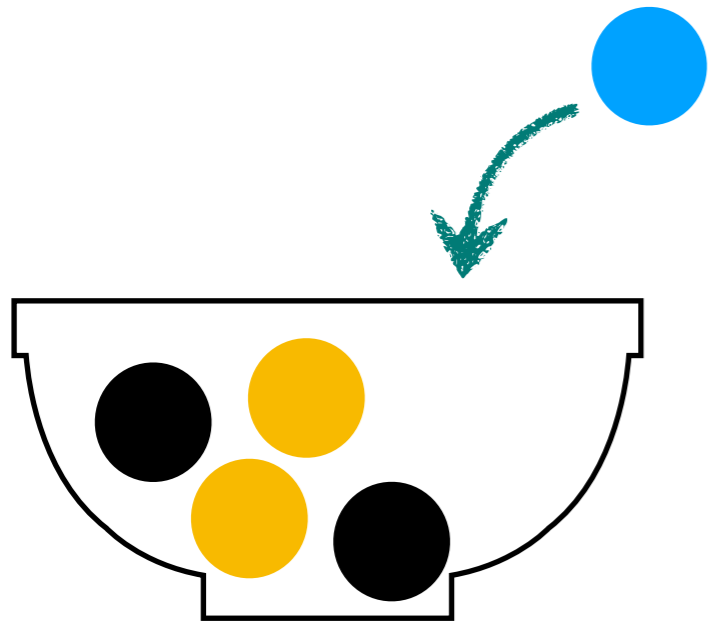
Items = 4
 $F = 2$



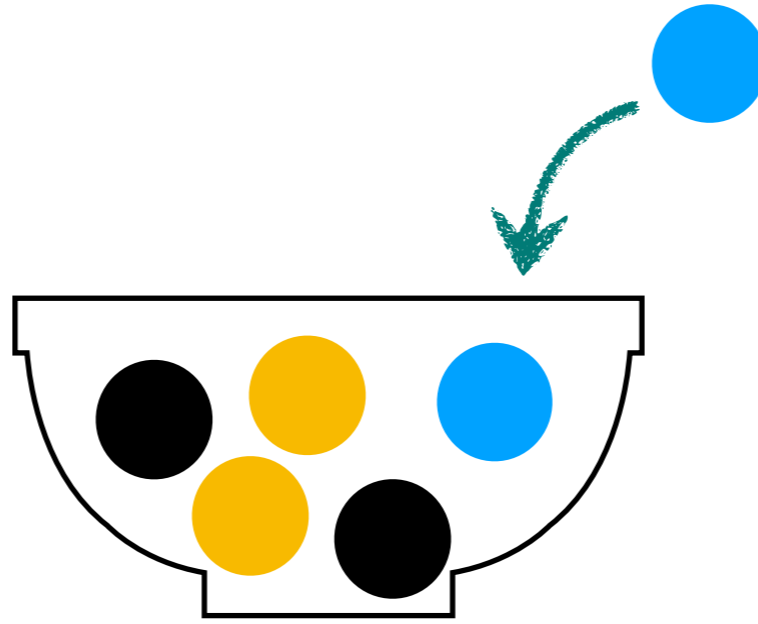
Items = 4 + 1
 $F = 2 + 1$

Sub-modularity

$F = \#$ Unique Coloured items



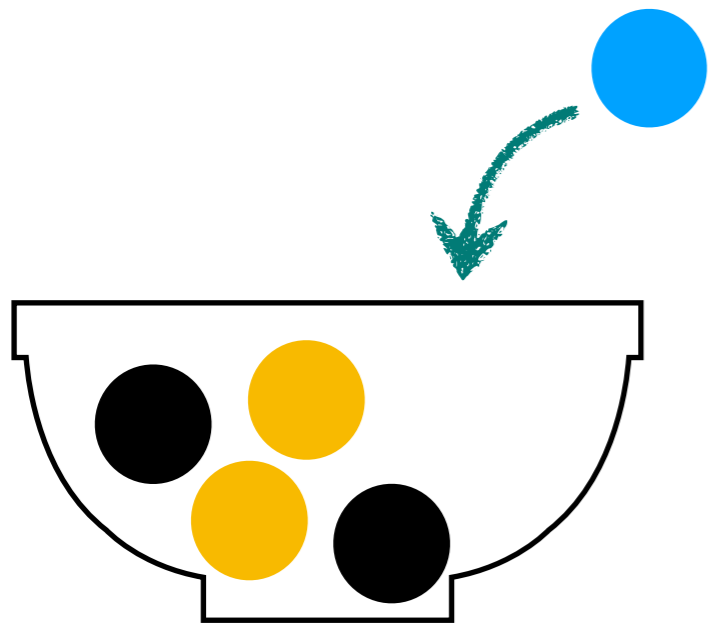
Items = 4
 $F = 2$



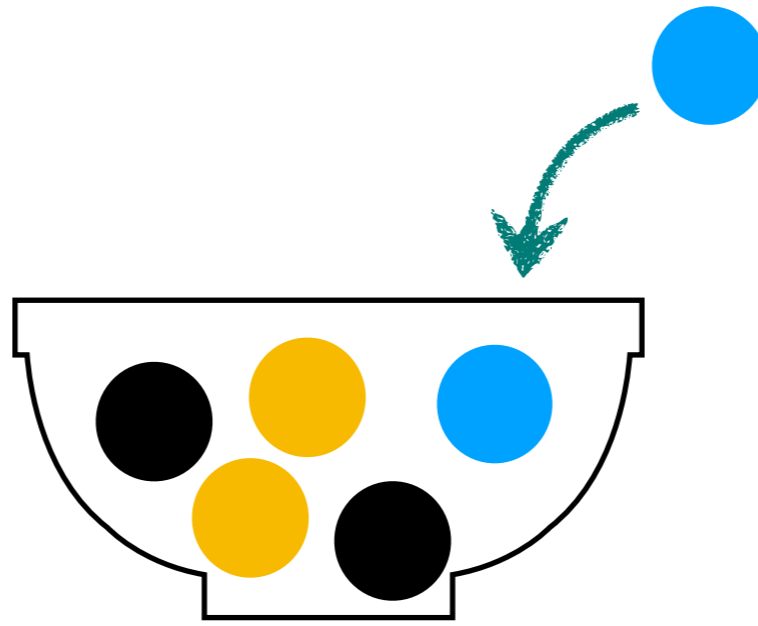
Items = 4 + 1
 $F = 2 + 1$

Sub-modularity

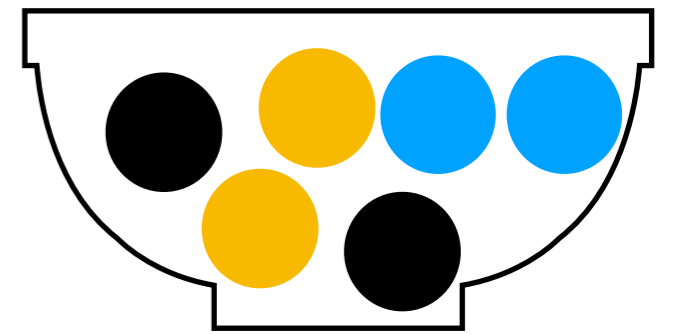
$F = \#$ Unique Coloured items



Items = 4
 $F = 2$



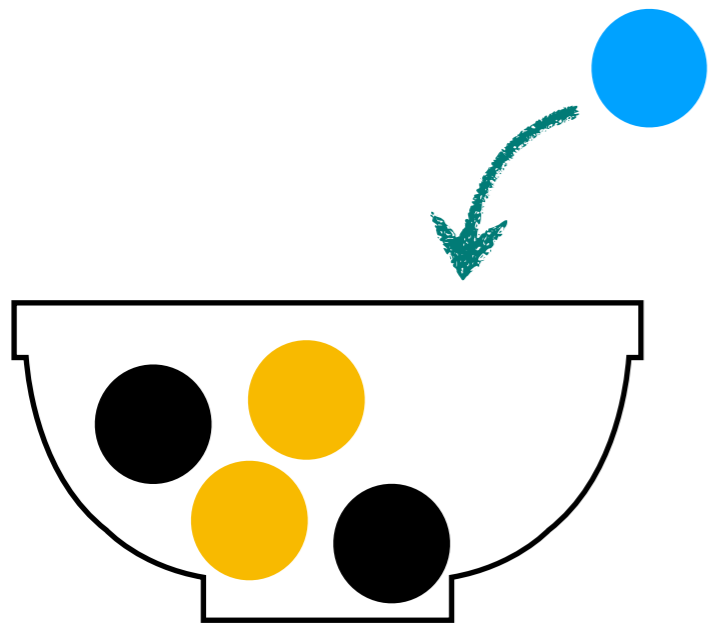
Items = 4 + 1
 $F = 2 + 1$



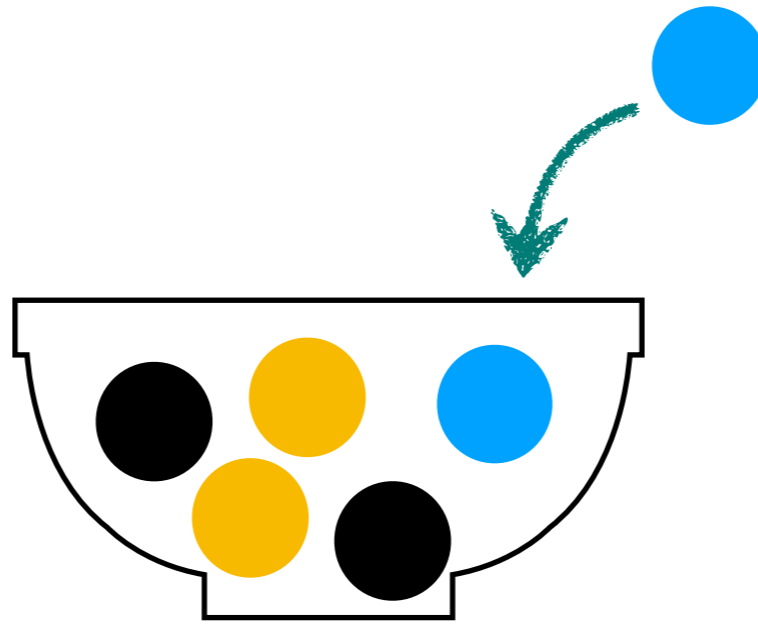
Items = 5 + 1
 $F = 3 + 0$

Sub-modularity

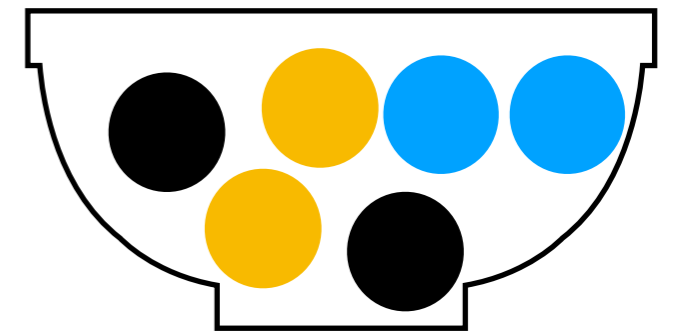
$F = \# \text{ Unique Coloured items}$



Items = 4
 $F = 2$



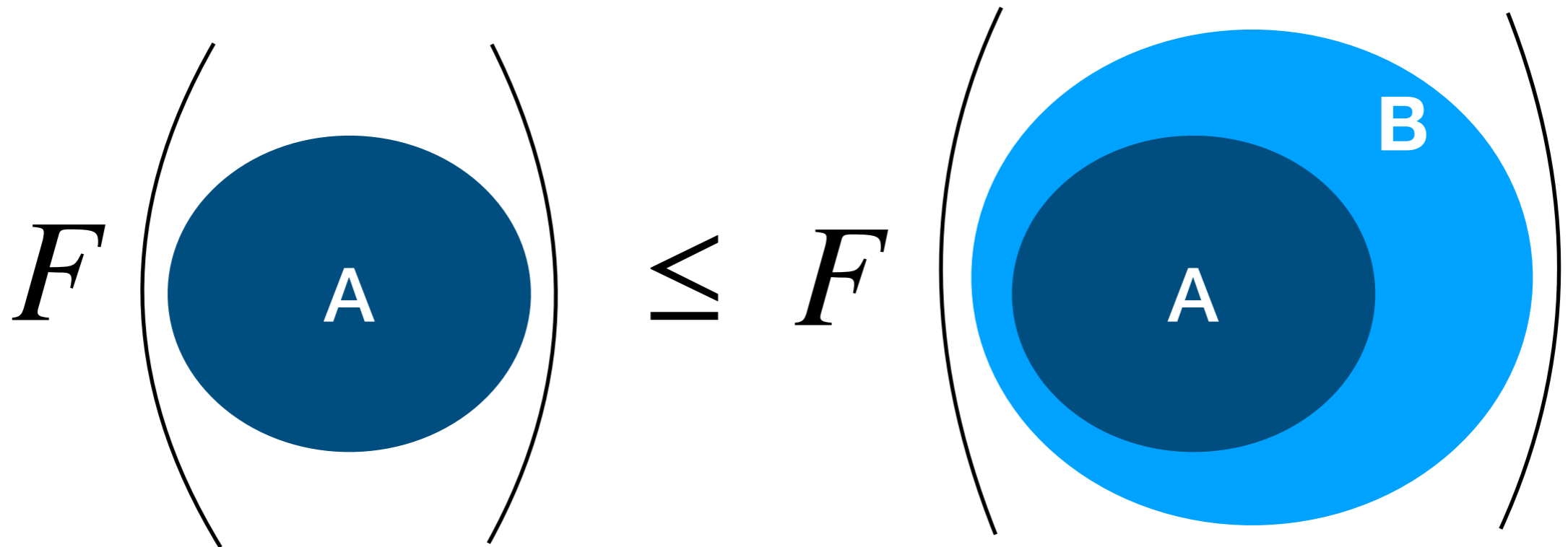
Items = 4 + 1
 $F = 2 + 1$



Items = 5 + 1
 $F = 3 + 0$

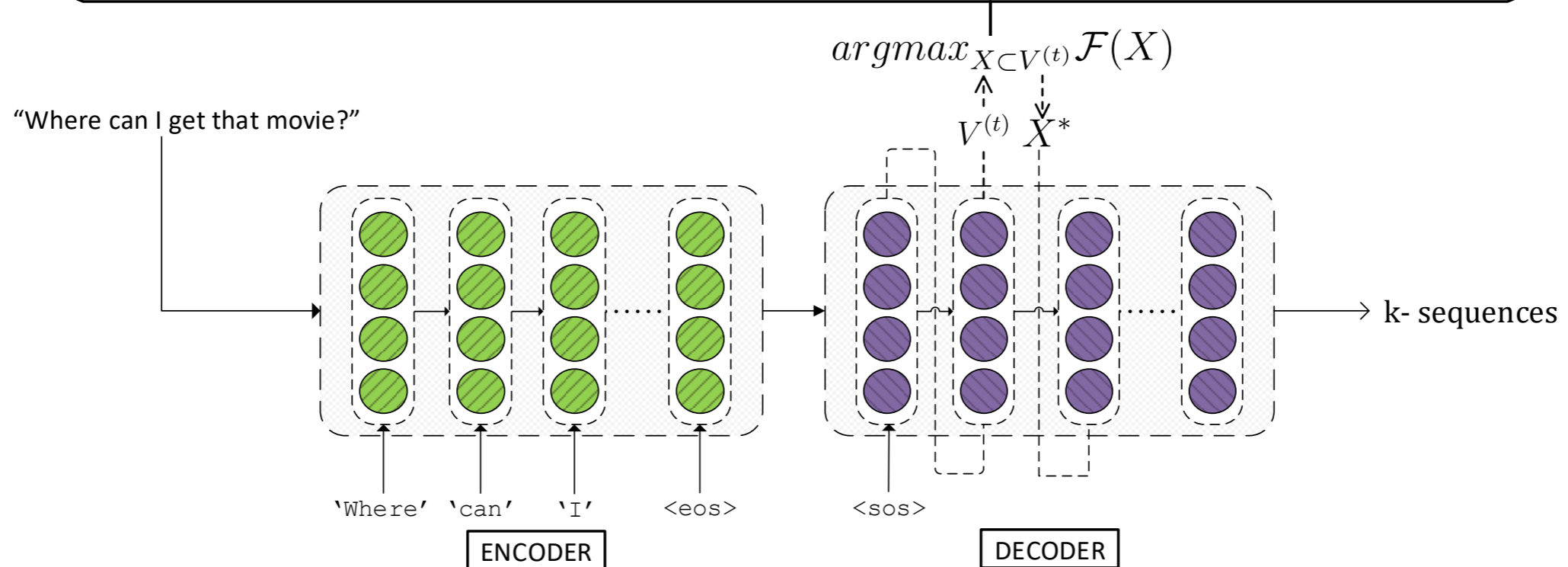
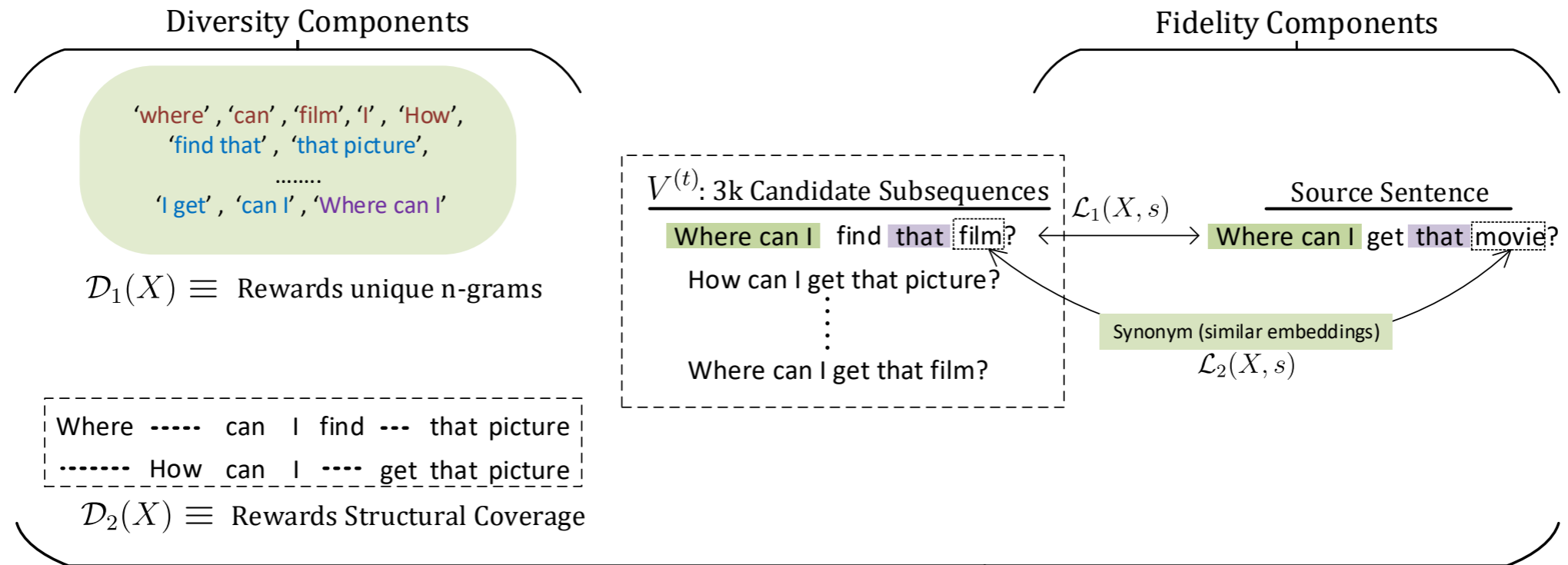
Diminishing Returns

Monotonicity



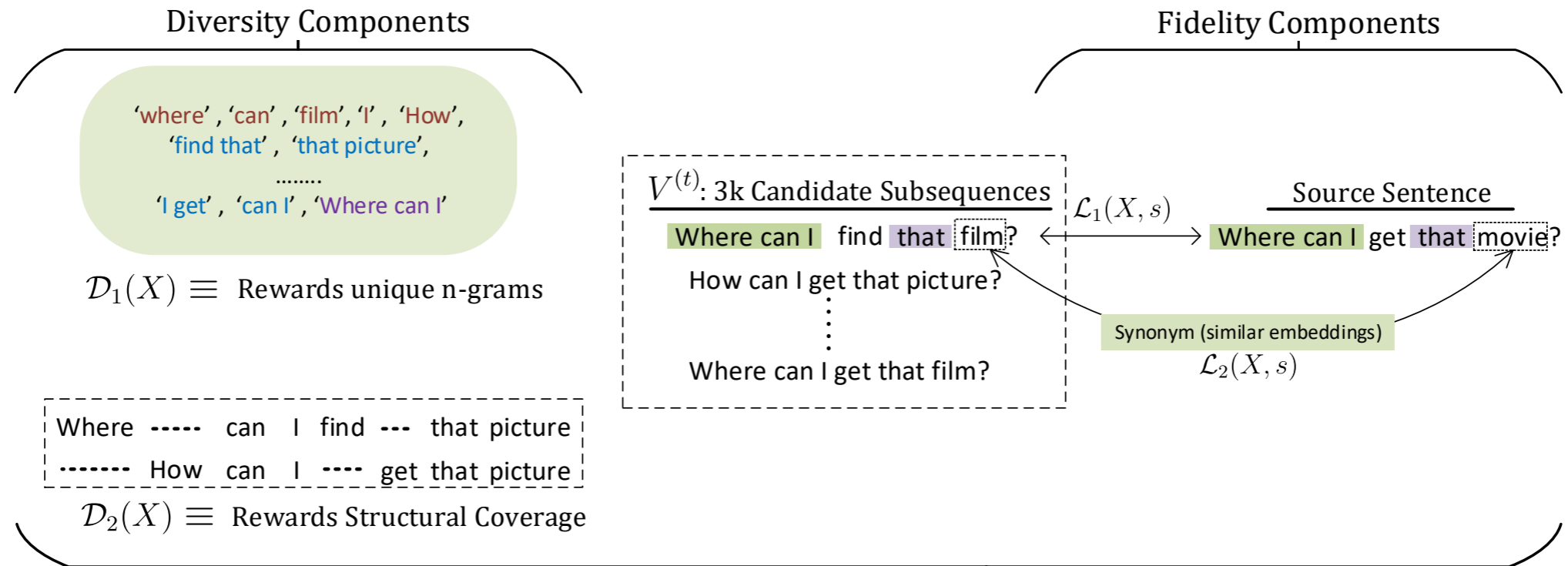
DiPS

Induce Diversity while not compromising on Fidelity

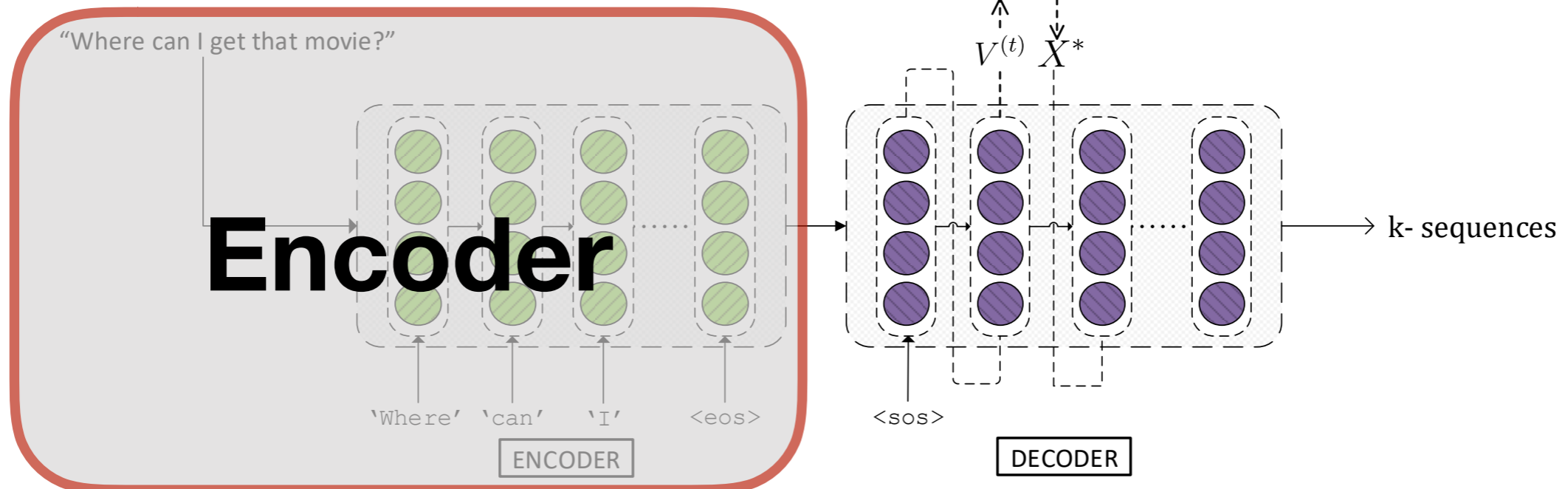


DiPS

Induce Diversity while not compromising on Fidelity

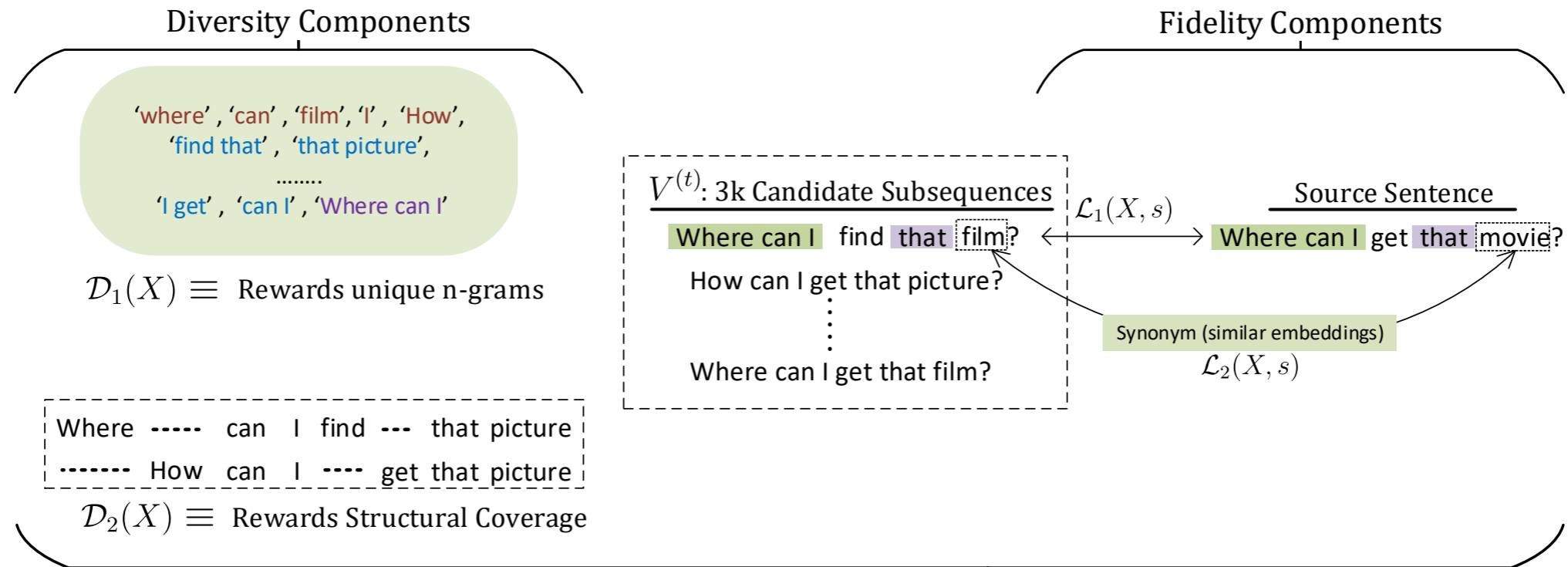


$$\operatorname{argmax}_{X \subset V^{(t)}} \mathcal{F}(X)$$

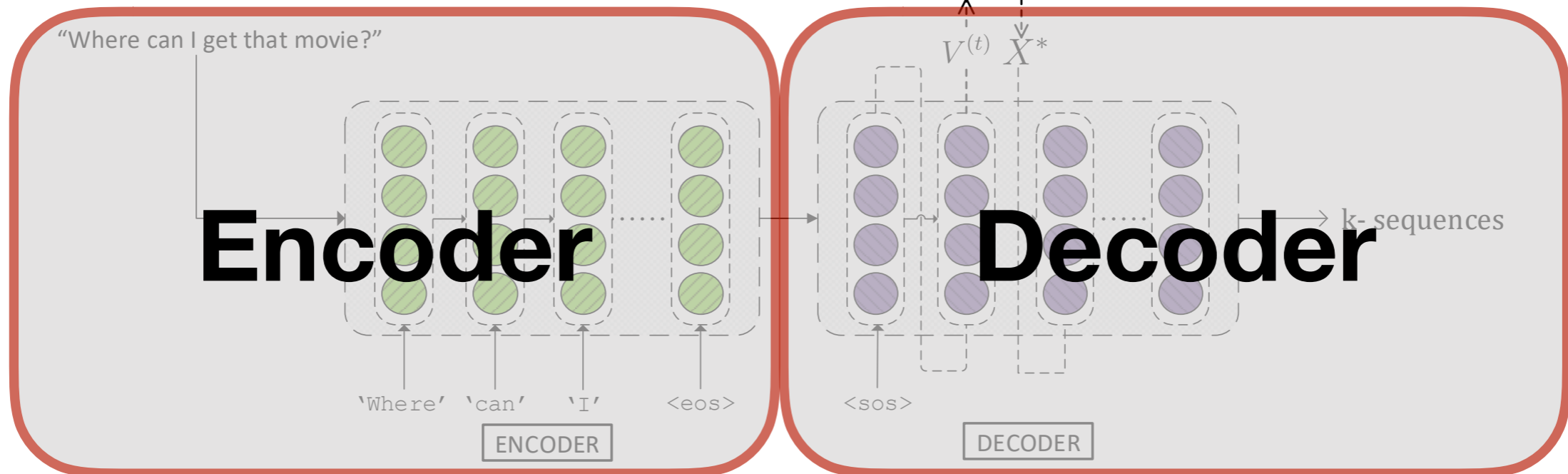


DiPS

Induce Diversity while not compromising on Fidelity



$$\operatorname{argmax}_{X \in V^{(t)}} \mathcal{F}(X)$$



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'

Diversity

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

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?

Source Sentence

Where can I get that movie?

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

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

$$\operatorname{argmax}_{X \in V^{(t)}} \mathcal{F}(X)$$

"Where can I get that movie?"

Encoder

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

ENCODER

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

Decoder

<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'

Diversity

$\mathcal{D}_2(X) \equiv$ Rewards Unique Segments

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

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

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

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

Fidelity Components

Source Sentence

Where can I get that movie?

Fidelity

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

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

$$\operatorname{argmax}_{X \in V^{(t)}} \mathcal{F}(X)$$

"Where can I get that movie?"

Encoder

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

ENCODER

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

Decoder

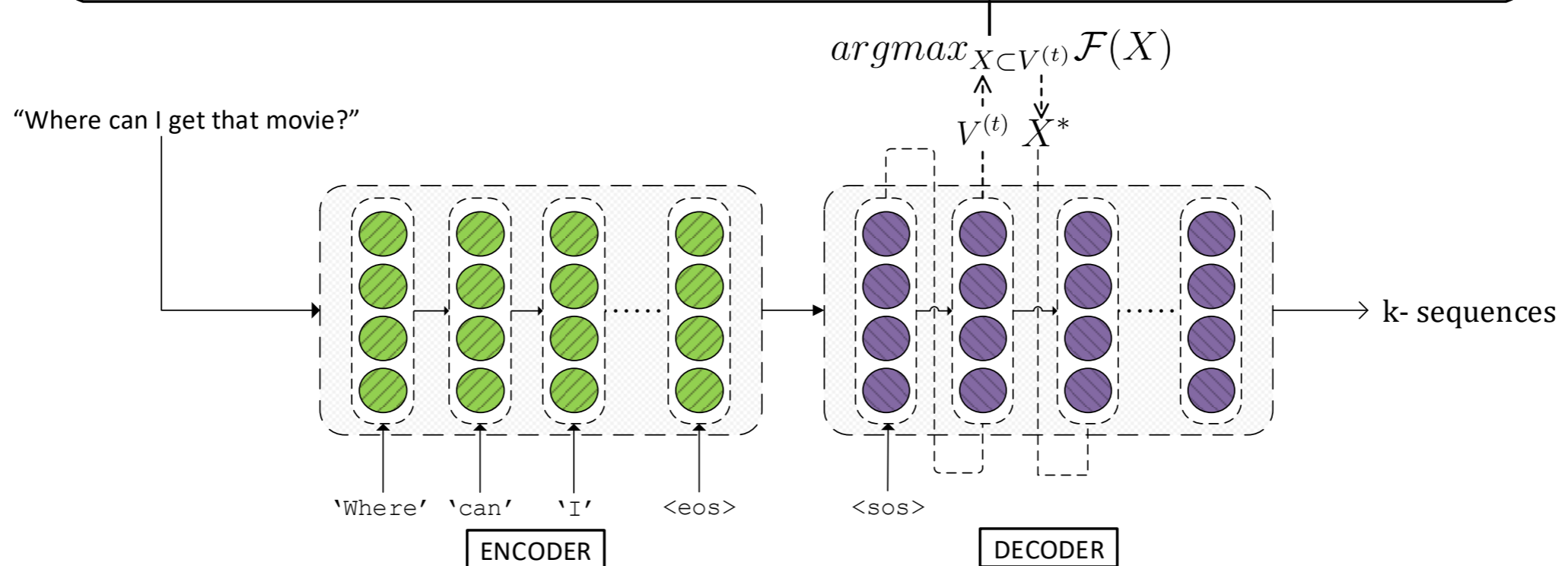
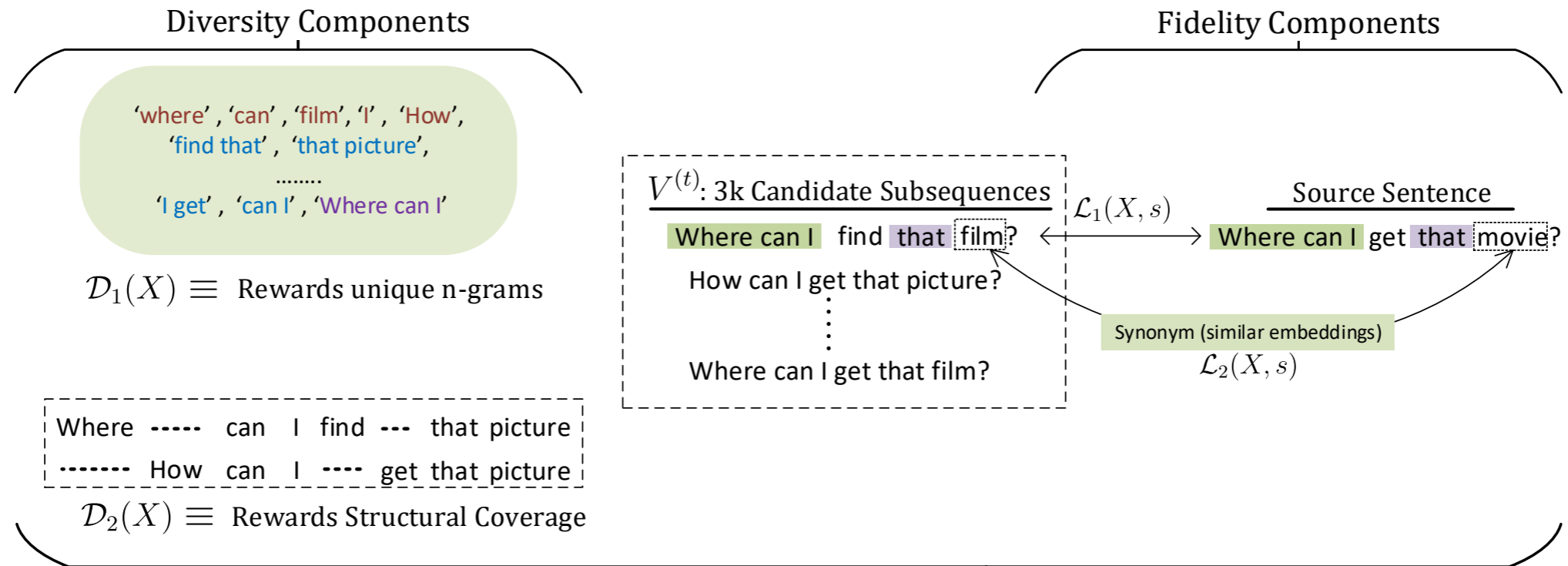
<sos>

DECODER

k-sequences

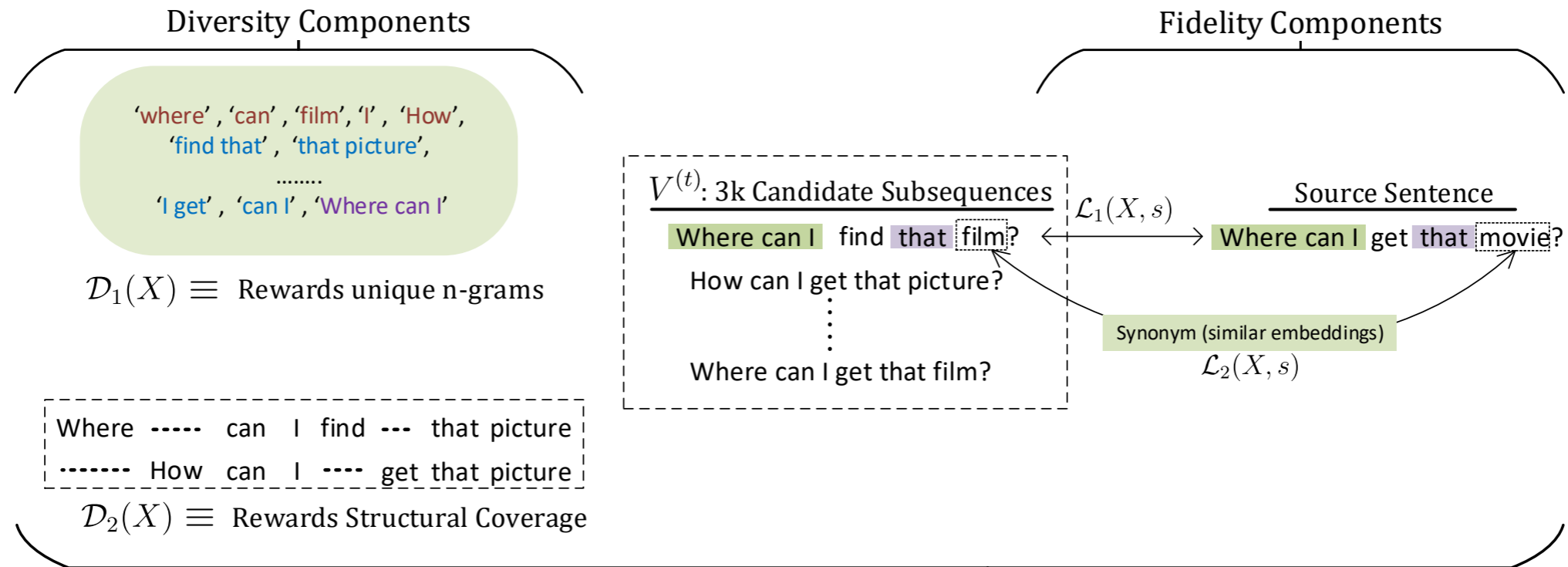
DiPS

Induce Diversity while not compromising on Fidelity

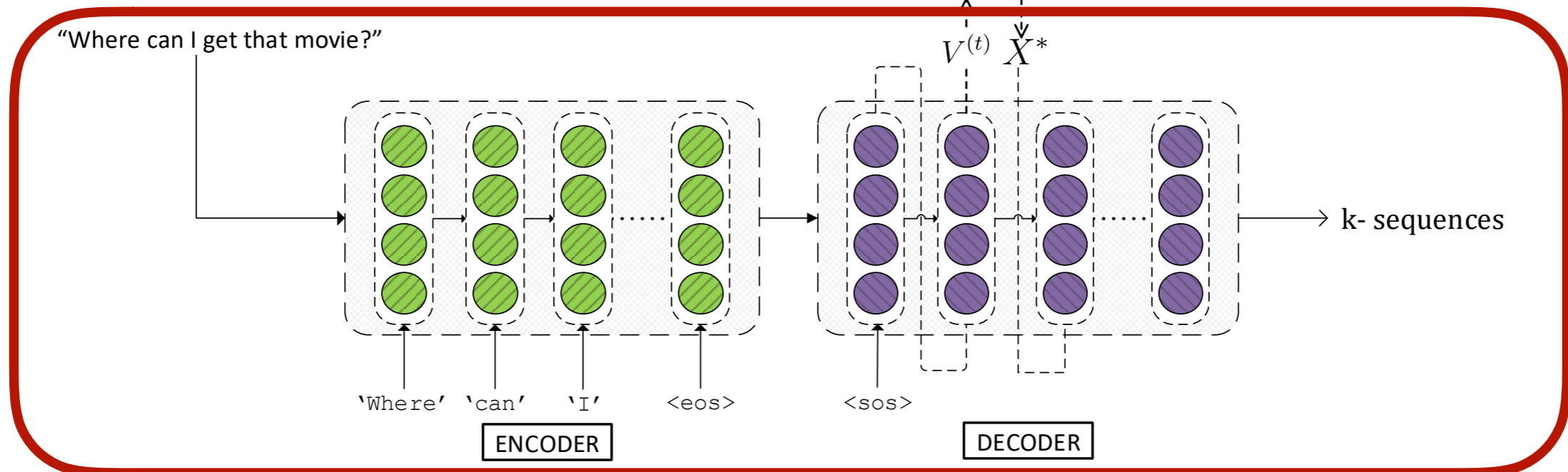


DiPS

Induce Diversity while not compromising on Fidelity

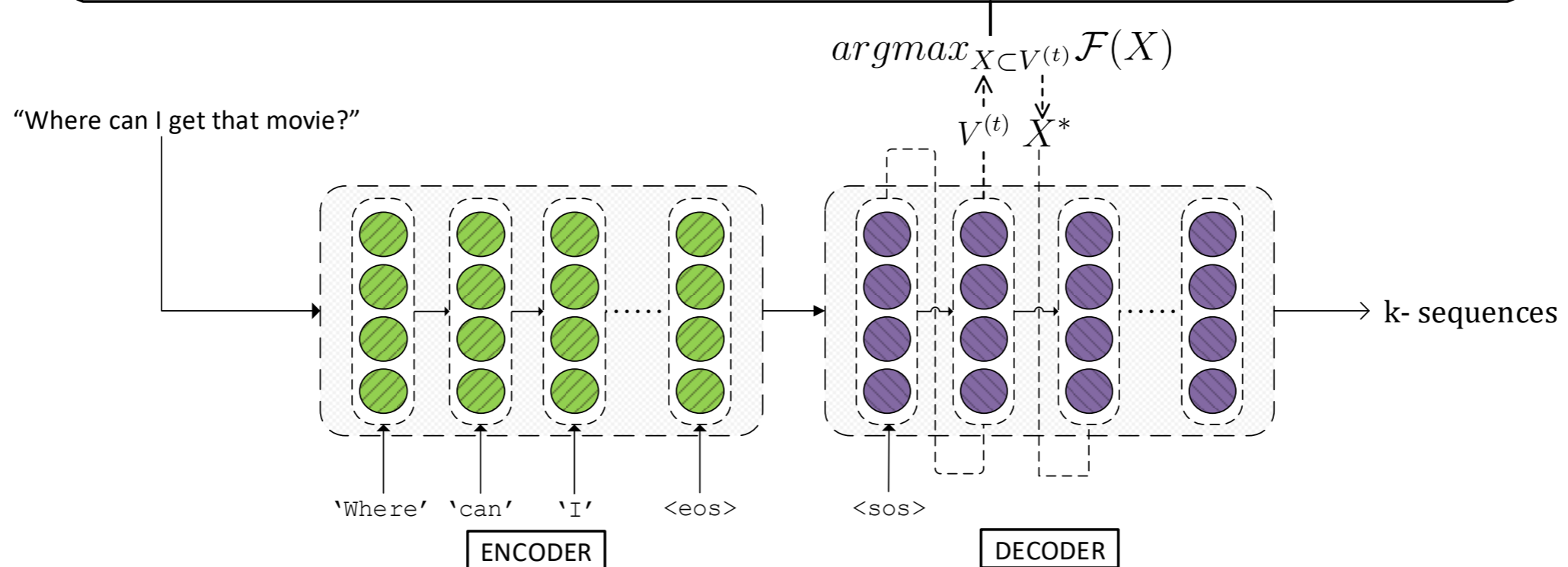
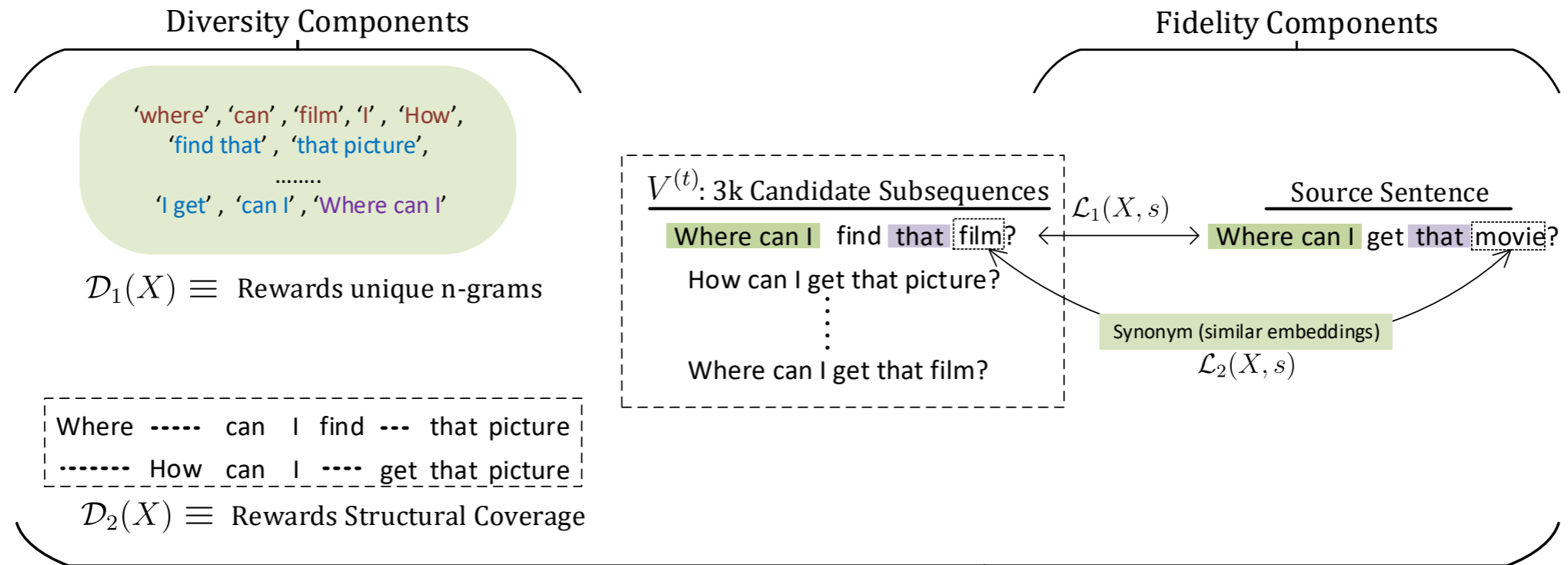


$$\operatorname{argmax}_{X \in V^{(t)}} \mathcal{F}(X)$$



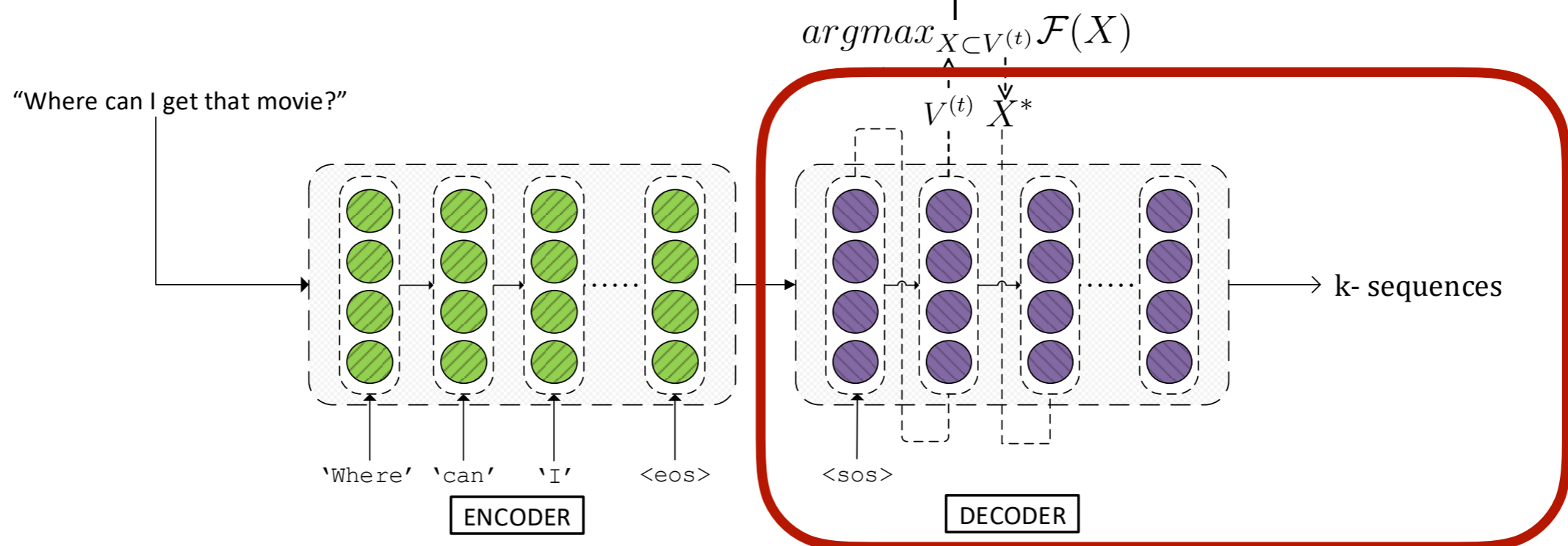
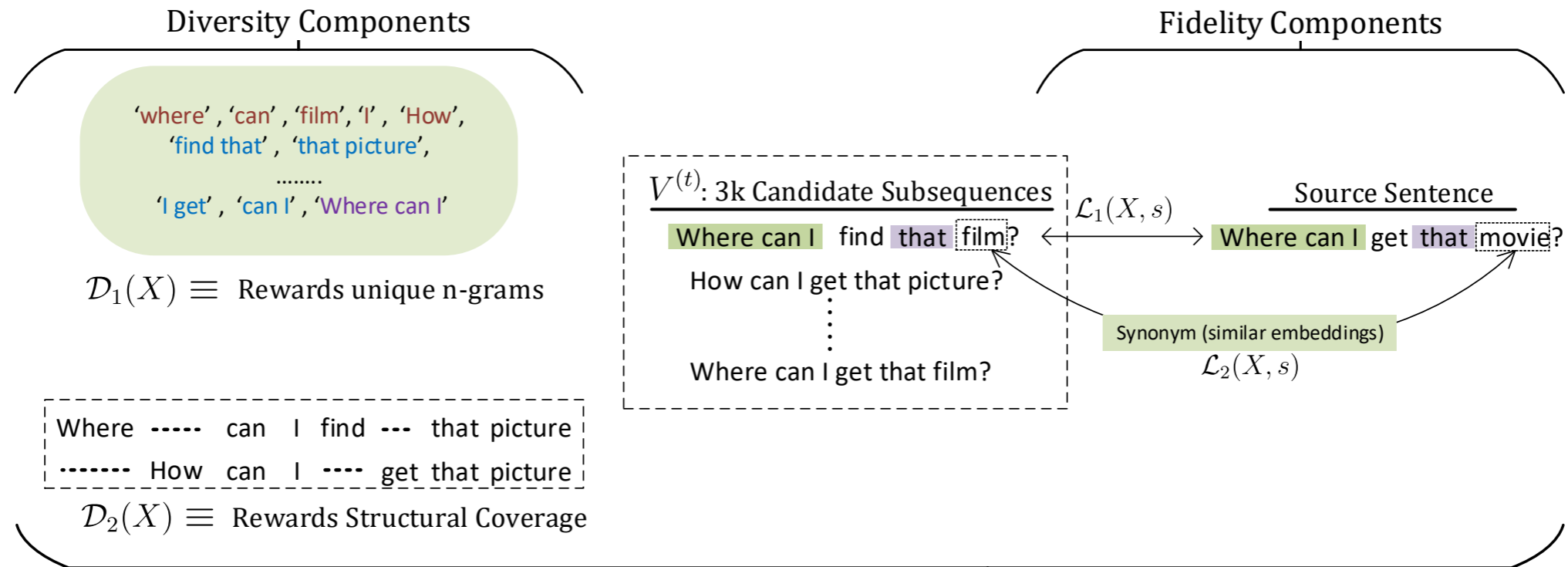
DiPS

Induce Diversity while not compromising on Fidelity



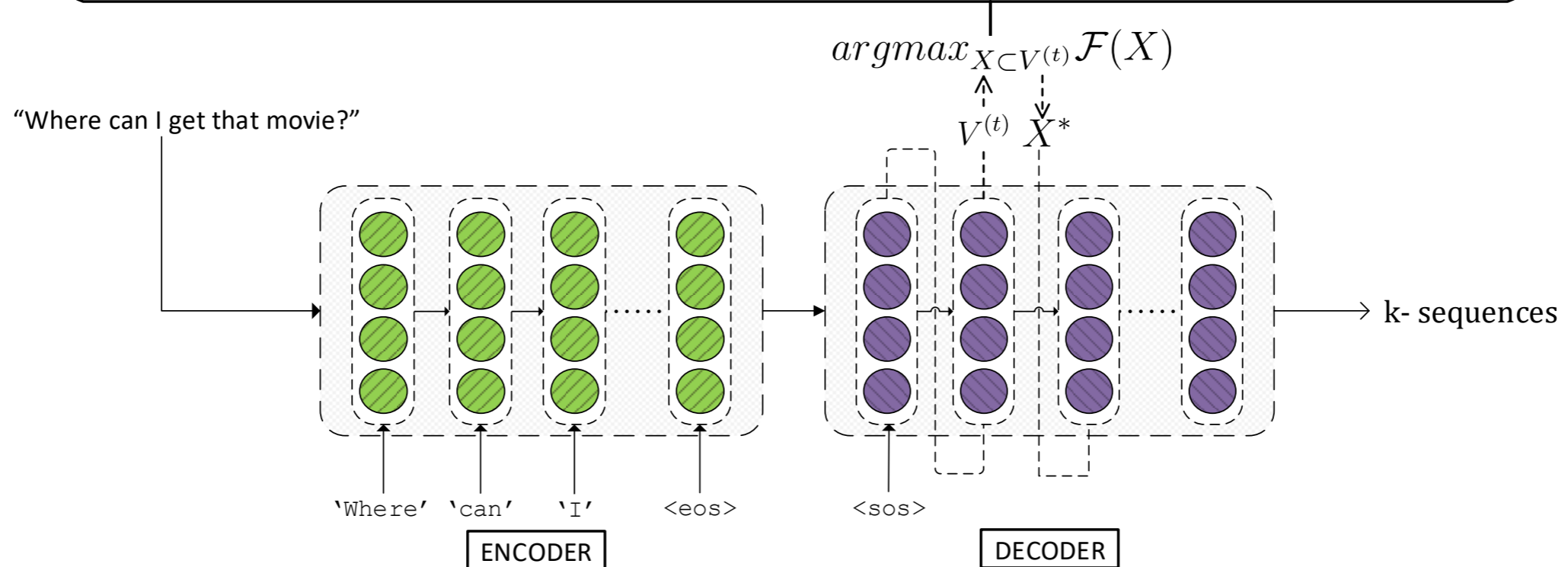
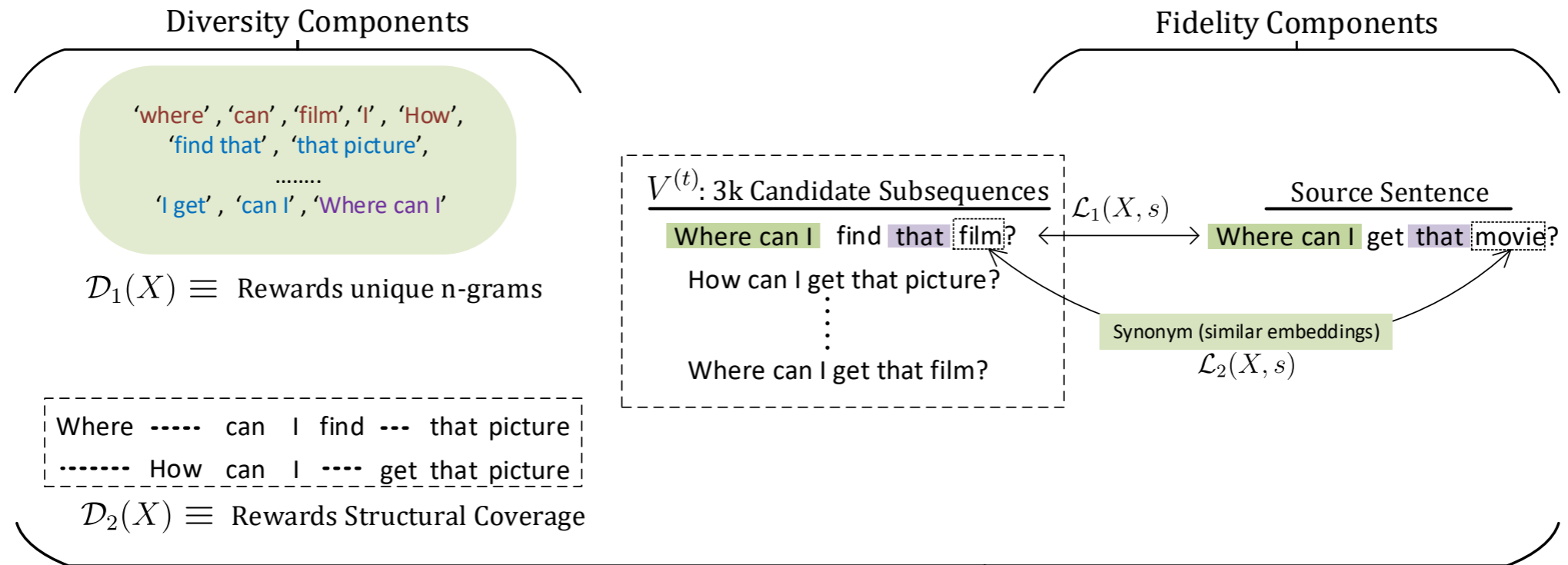
DiPS

Induce Diversity while not compromising on Fidelity



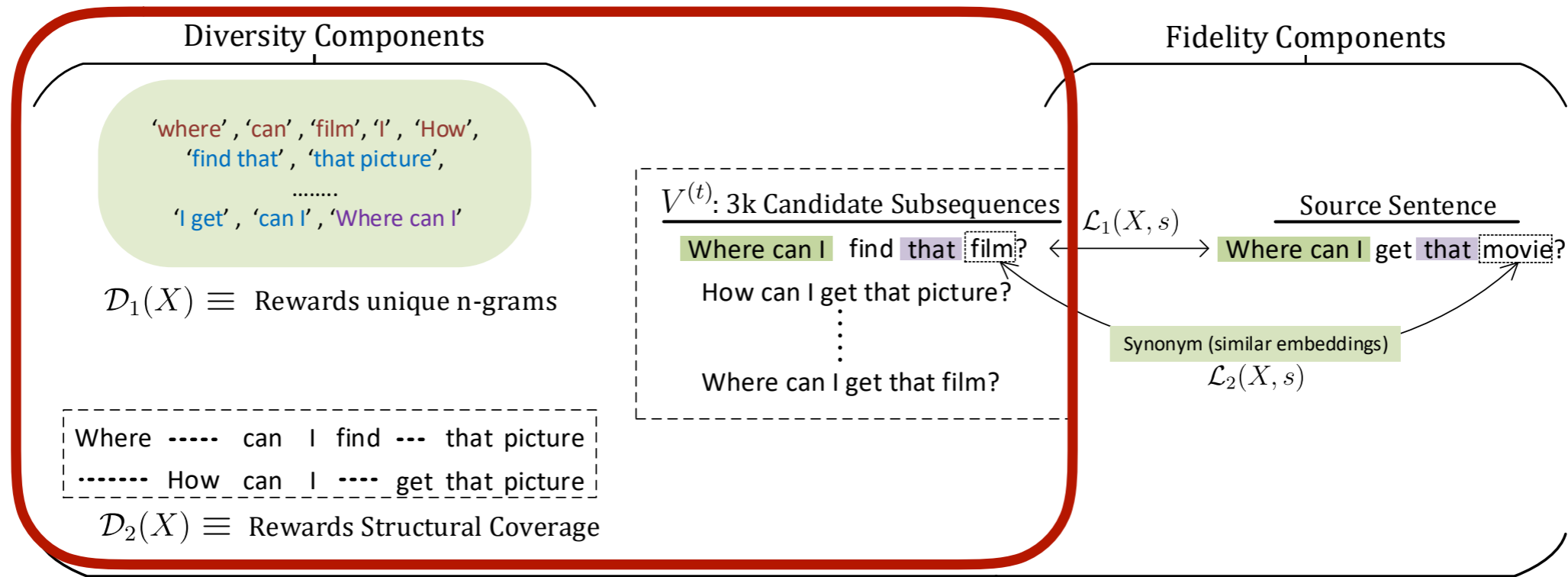
DiPS

Induce Diversity while not compromising on Fidelity



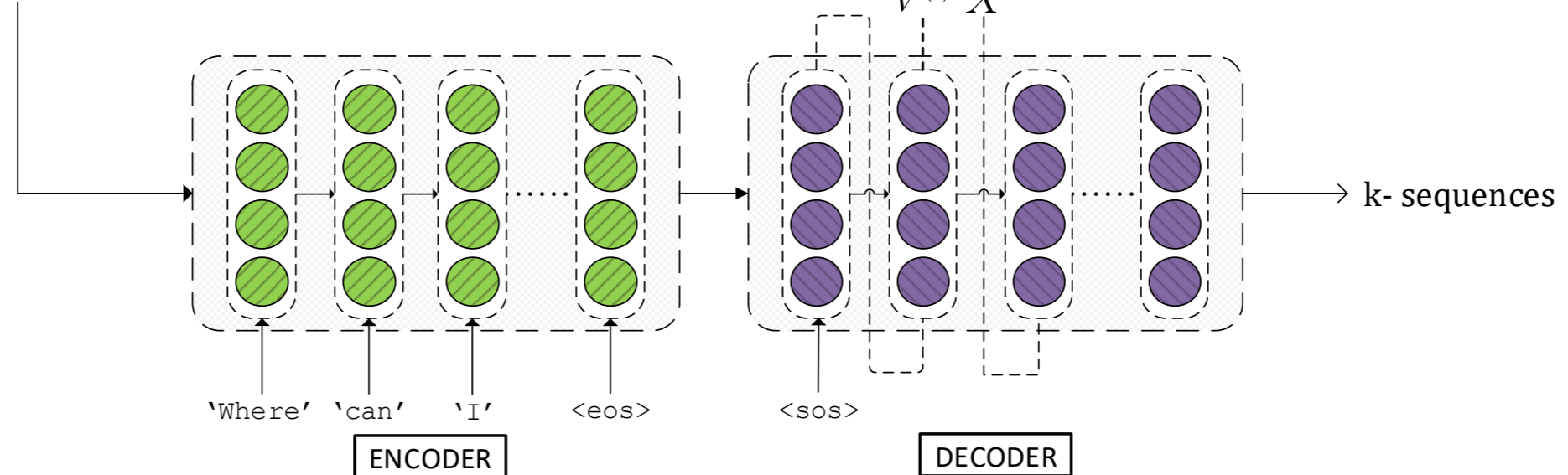
DiPS

Induce Diversity while not compromising on Fidelity



$$\operatorname{argmax}_{X \subset V^{(t)}} \mathcal{F}(X)$$

"Where can I get that movie?"



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

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

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

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

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

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

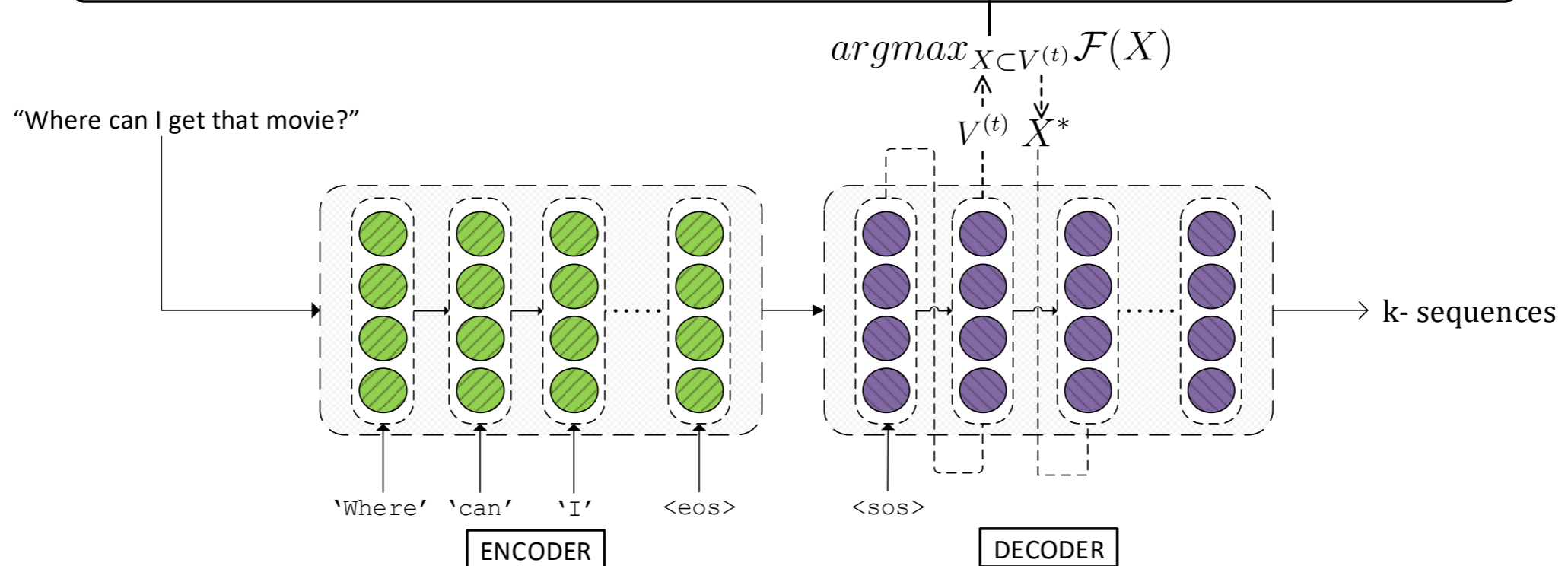
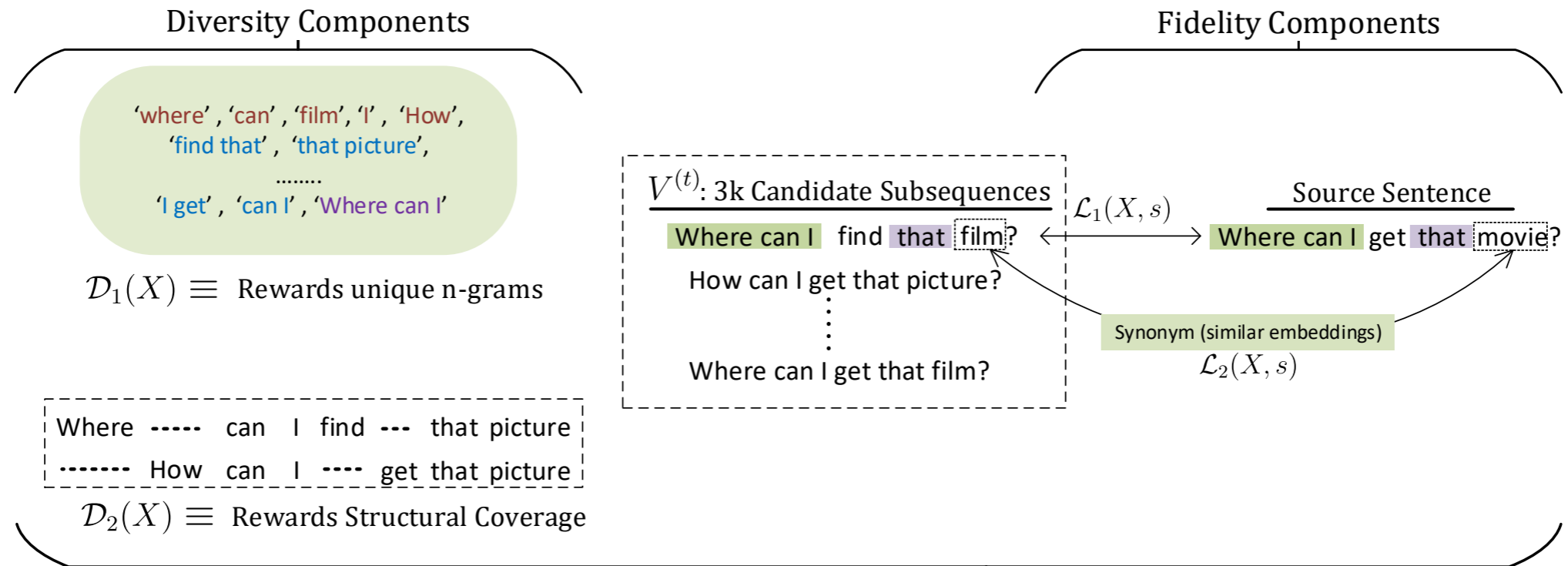
Structural Coverage

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

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

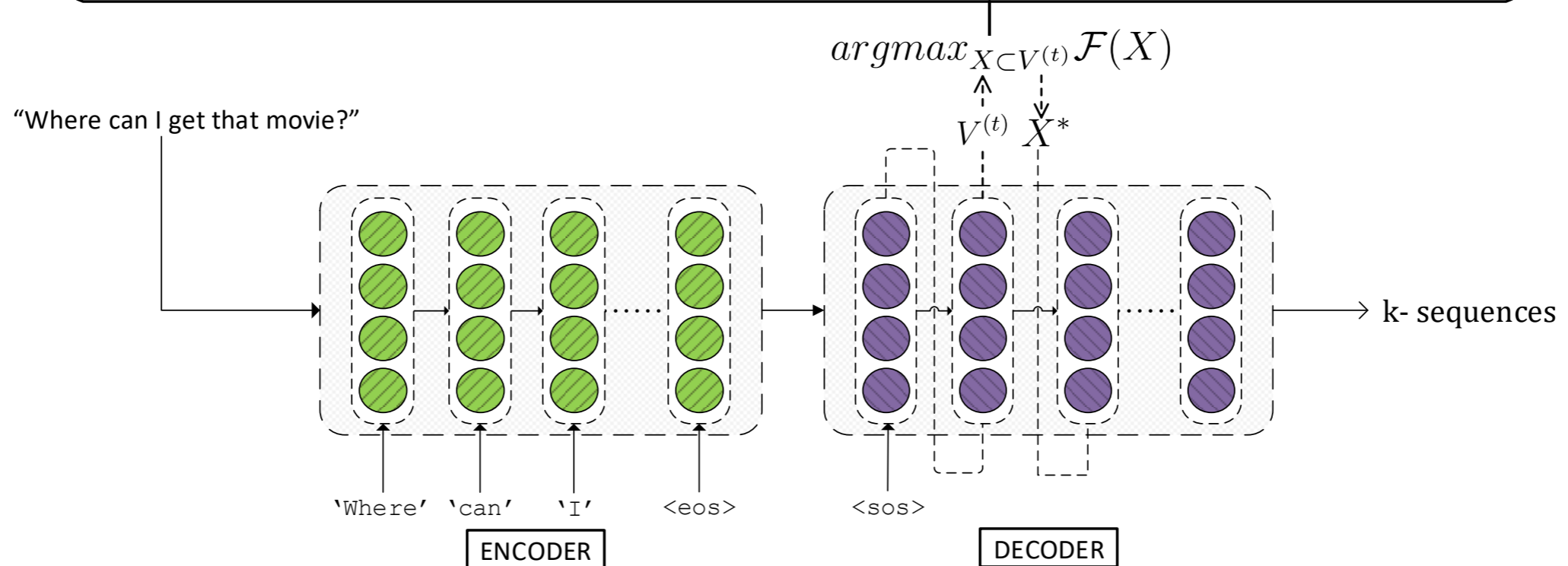
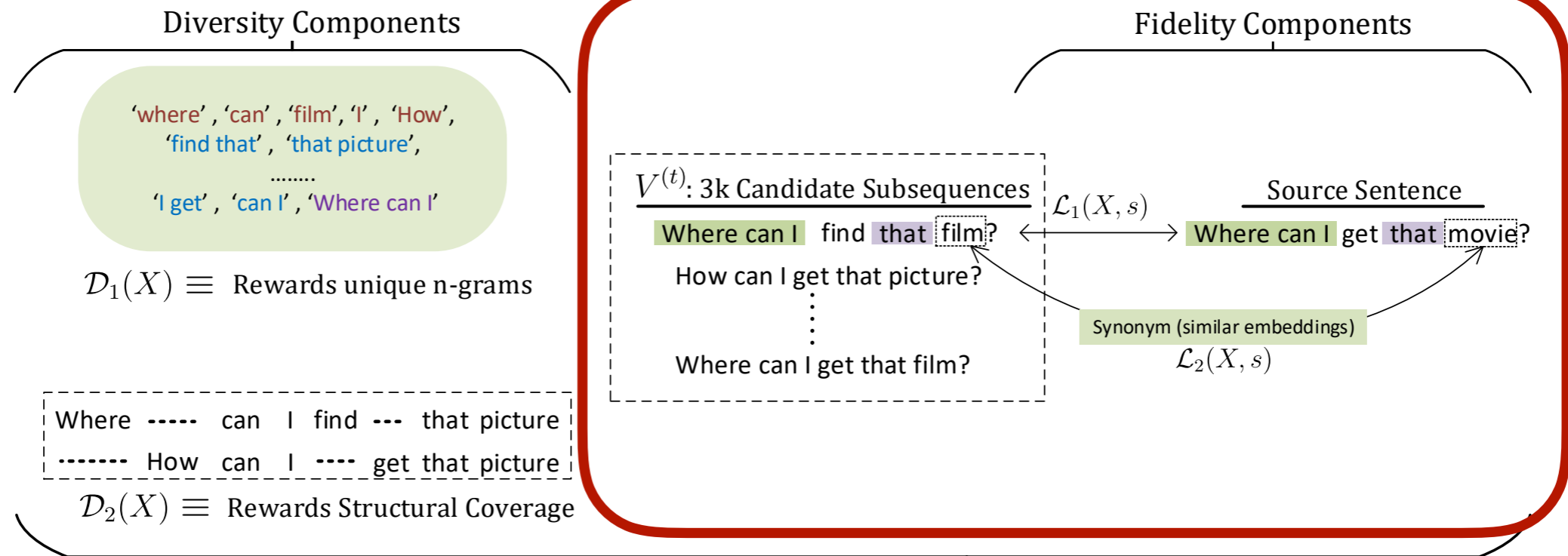
DiPS

Induce Diversity while not compromising on Fidelity

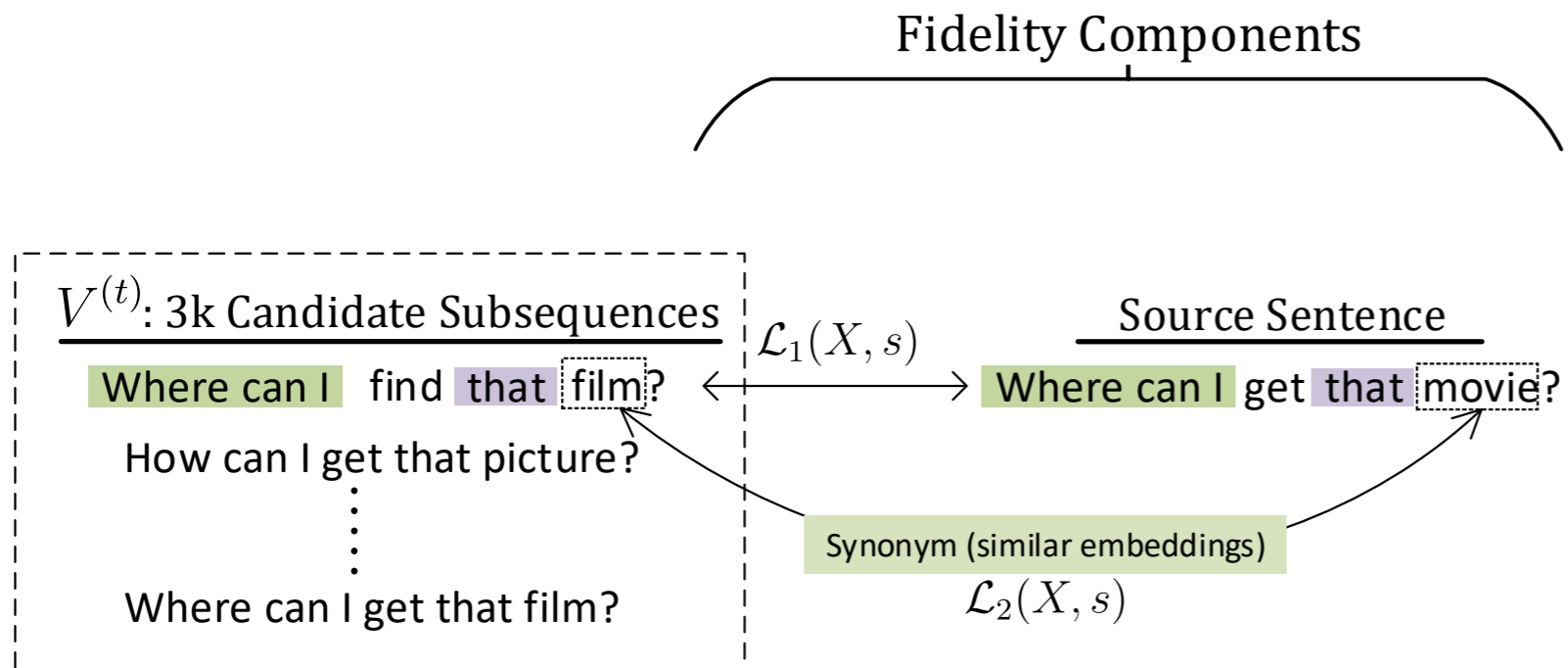


DiPS

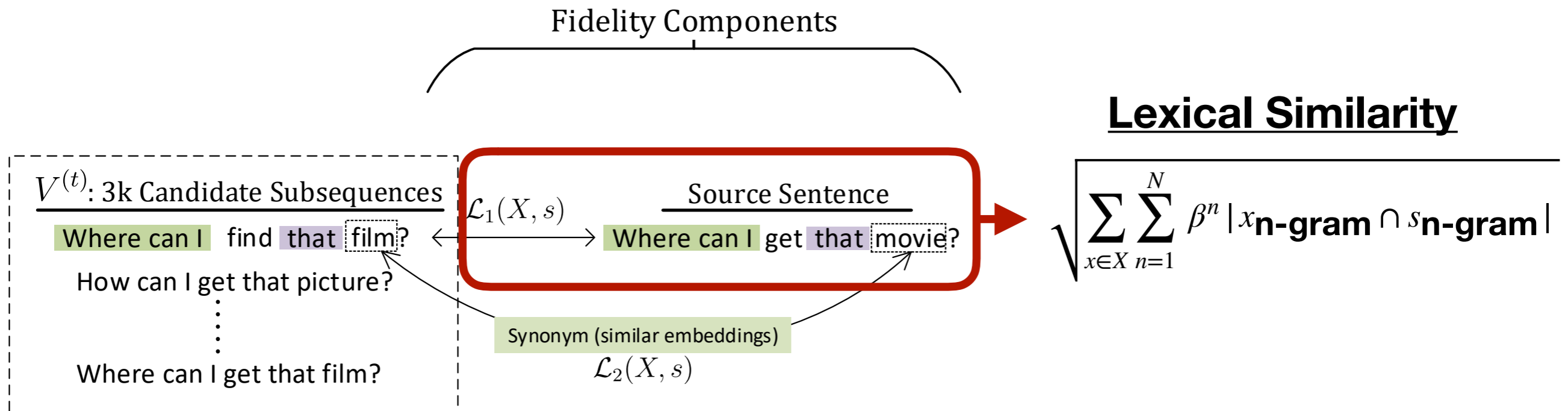
Induce Diversity while not compromising on Fidelity



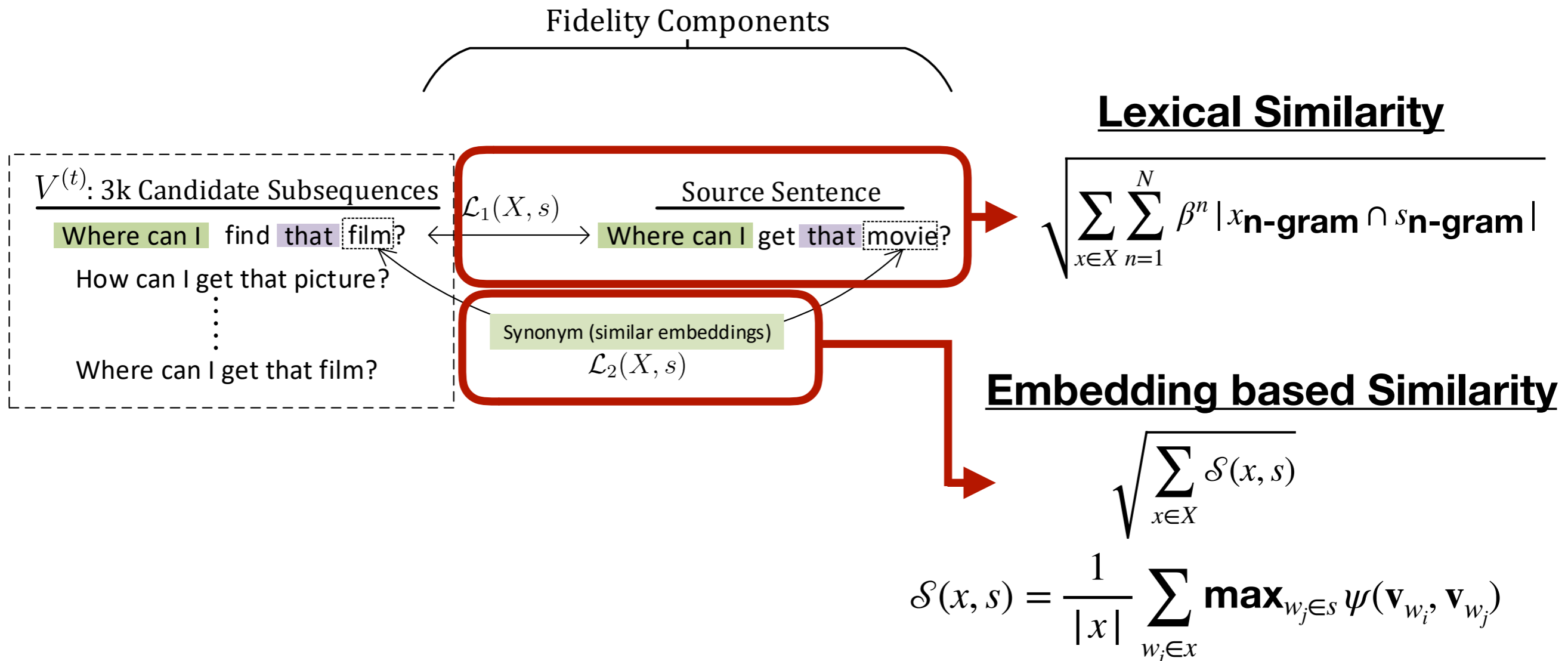
Fidelity Components



Fidelity Components



Fidelity Components



DiPS Objective

Diversity Components

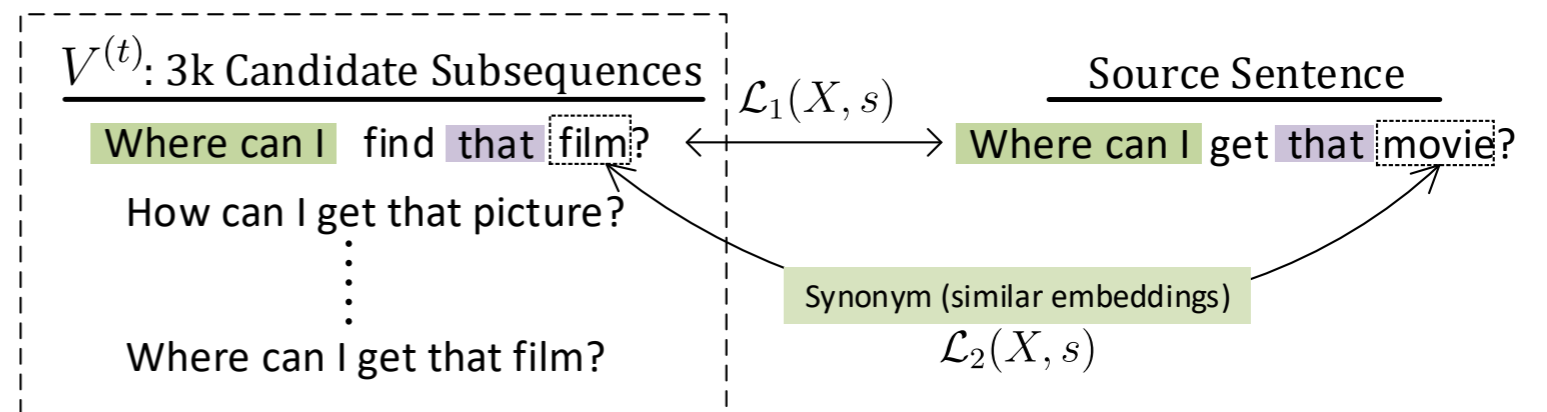
'where', 'can', 'film', 'I', 'How',
'find that', 'that picture',
.....
'I get', 'can I', 'Where can I'

$\mathcal{D}_1(X) \equiv$ 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

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

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

Diversity Components

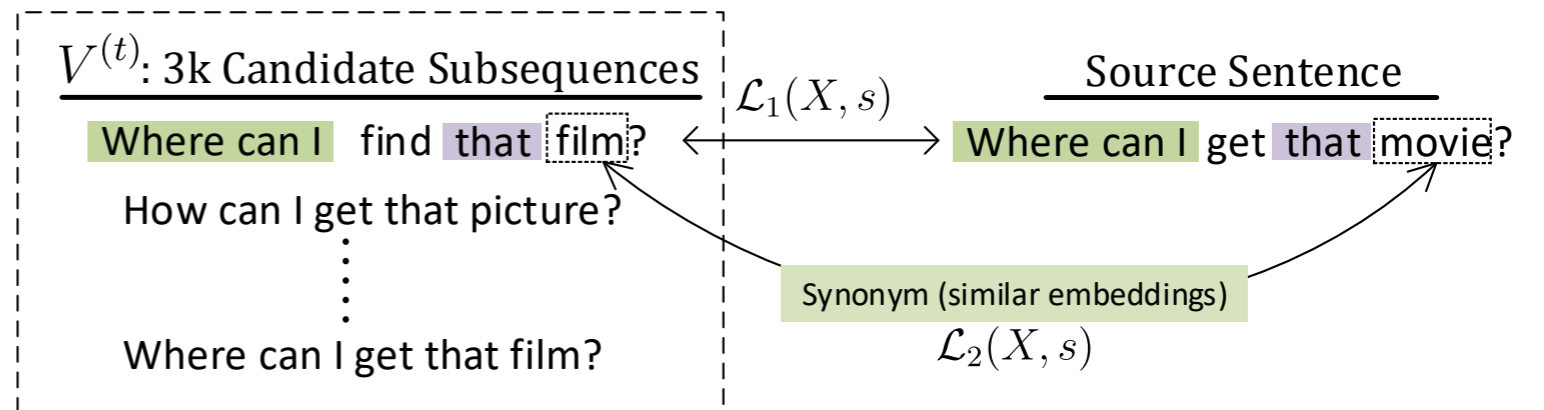
'where', 'can', 'film', 'I', 'How',
'find that', 'that picture',
.....
'I get', 'can I', 'Where can I'

$\mathcal{D}_1(X) \equiv$ 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



DiPS Objective

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

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

Diversity Components

'where', 'can', 'film', 'I', 'How',
 'find that', 'that picture',

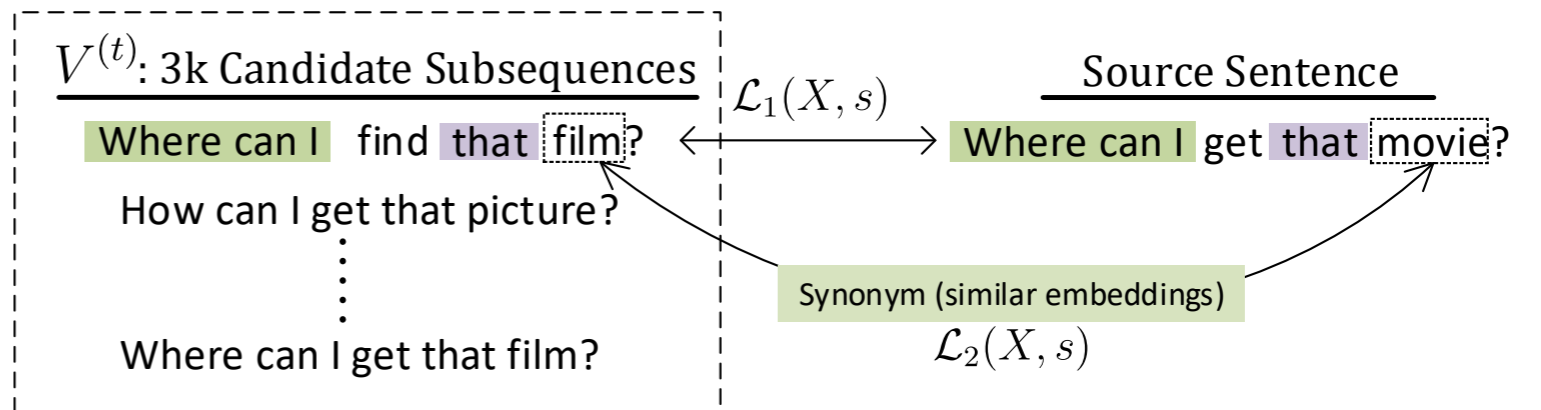
 'I get', 'can I', 'Where can I'

$\mathcal{D}_1(X) \equiv$ 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



DiPS Objective

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

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

Diversity Components

'where', 'can', 'film', 'I', 'How',
 'find that', 'that picture',

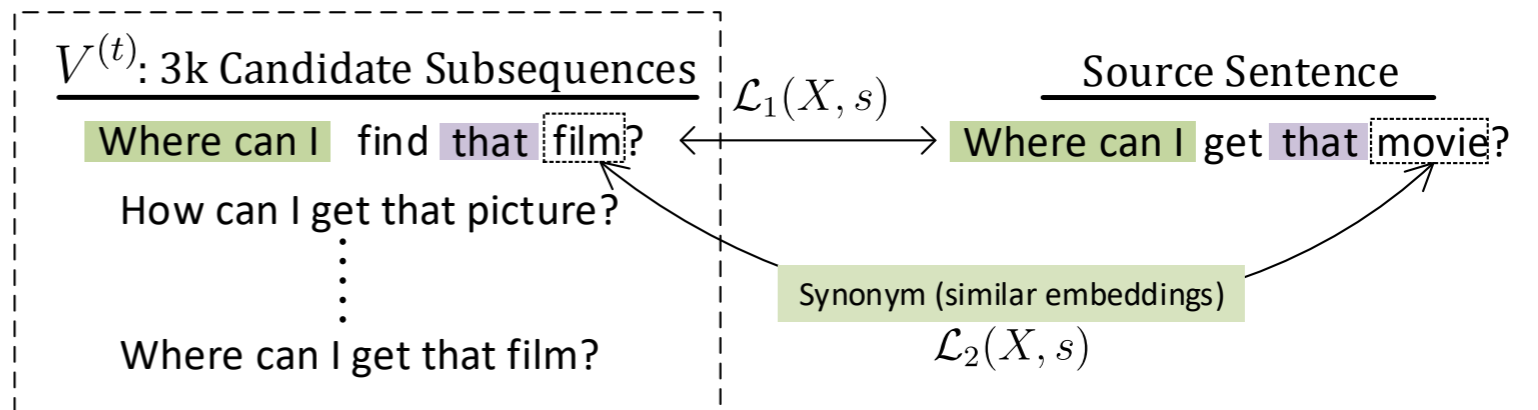
 'I get', 'can I', 'Where can I'

$\mathcal{D}_1(X) \equiv$ 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



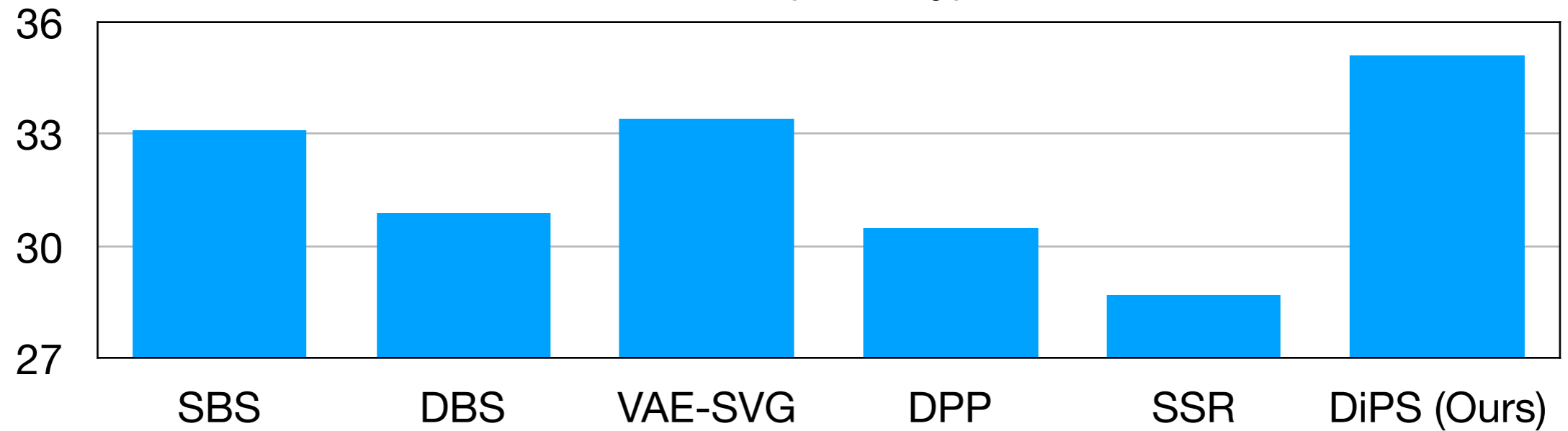
Fidelity & Diversity

(Quora Dataset)

Fidelity & Diversity

(Quora Dataset)

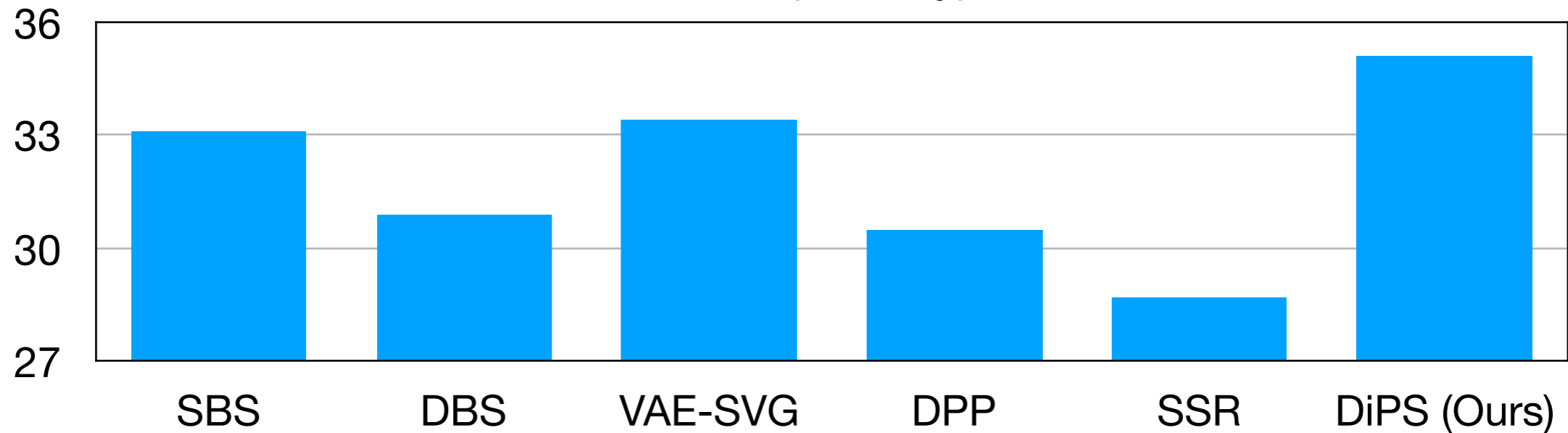
■ BLEU (Fidelity)



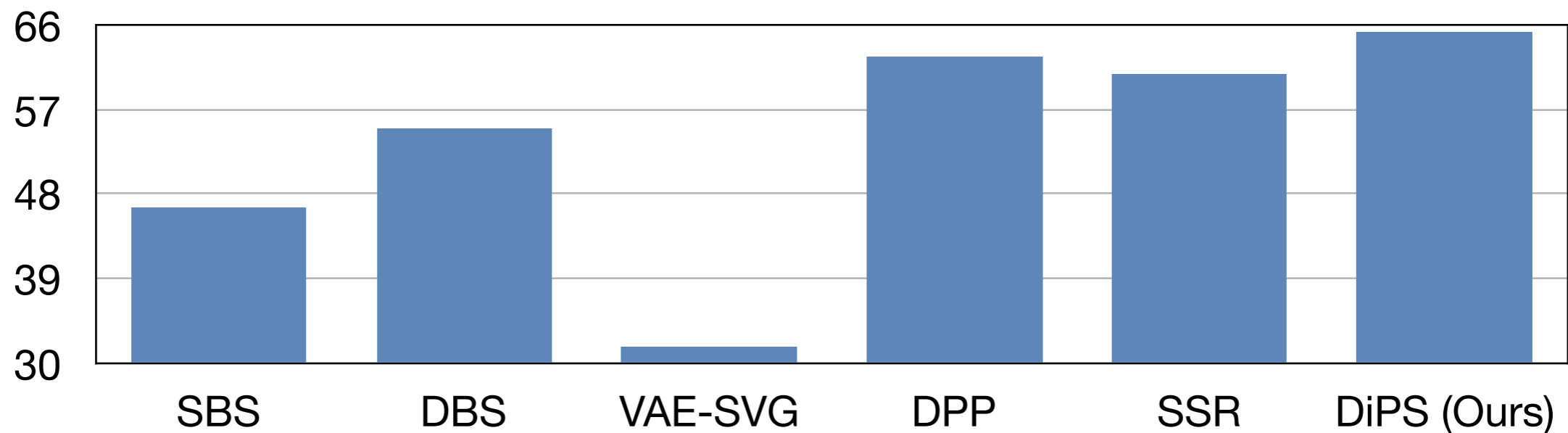
Fidelity & Diversity

(Quora Dataset)

■ BLEU (Fidelity)



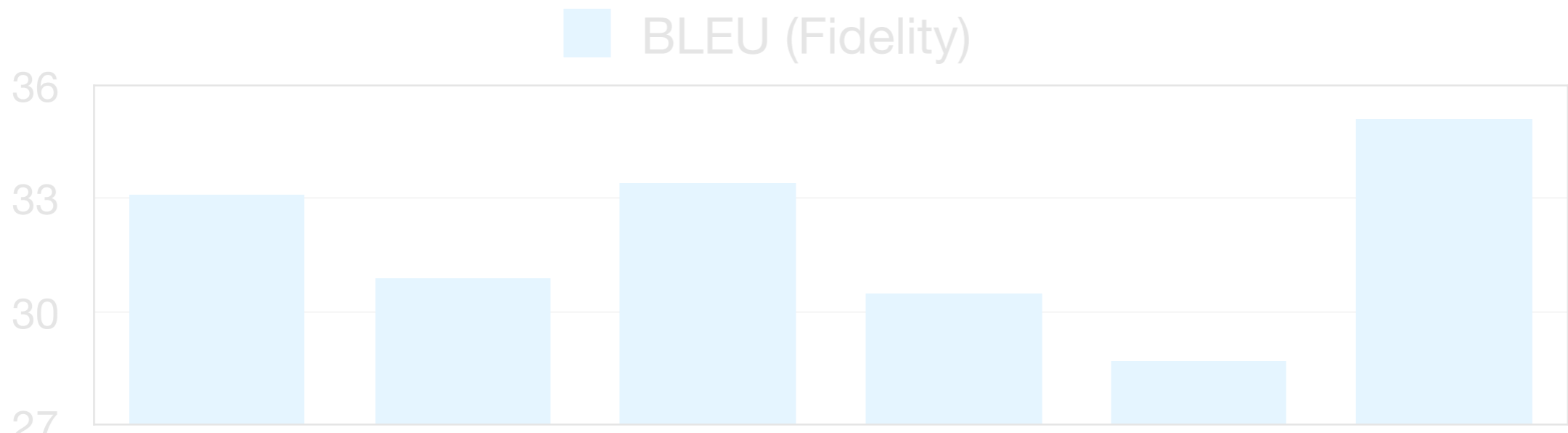
■ 4-Distinct (Diversity)



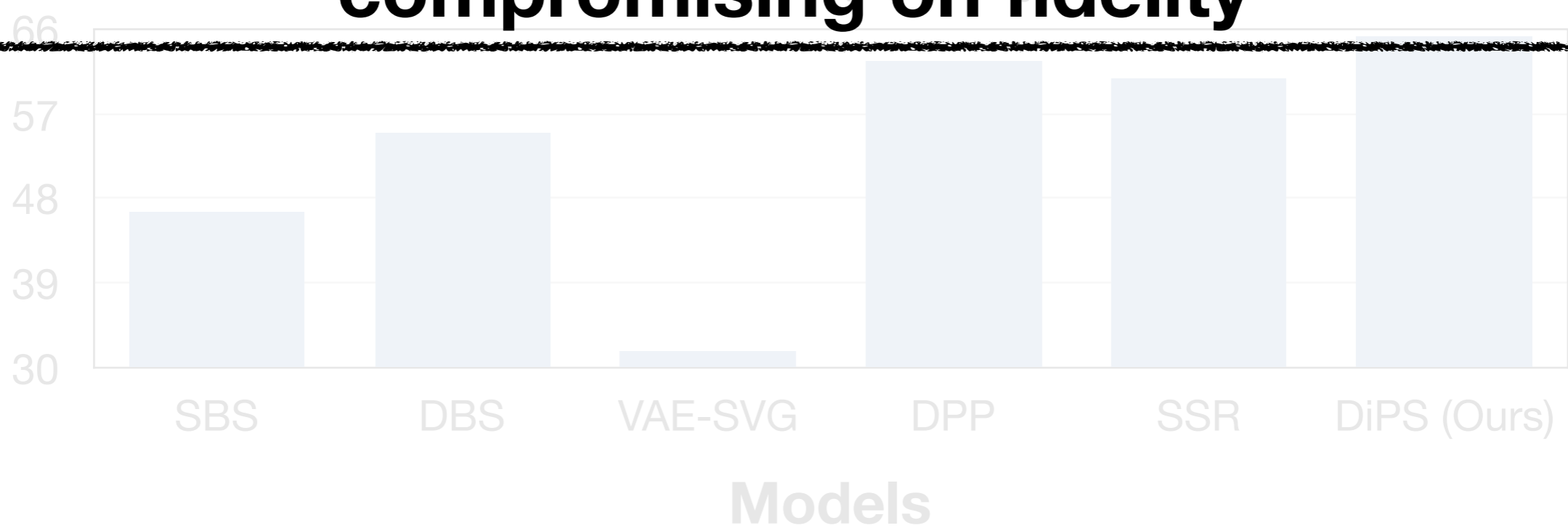
Models

Fidelity & Diversity

(Quora Dataset)

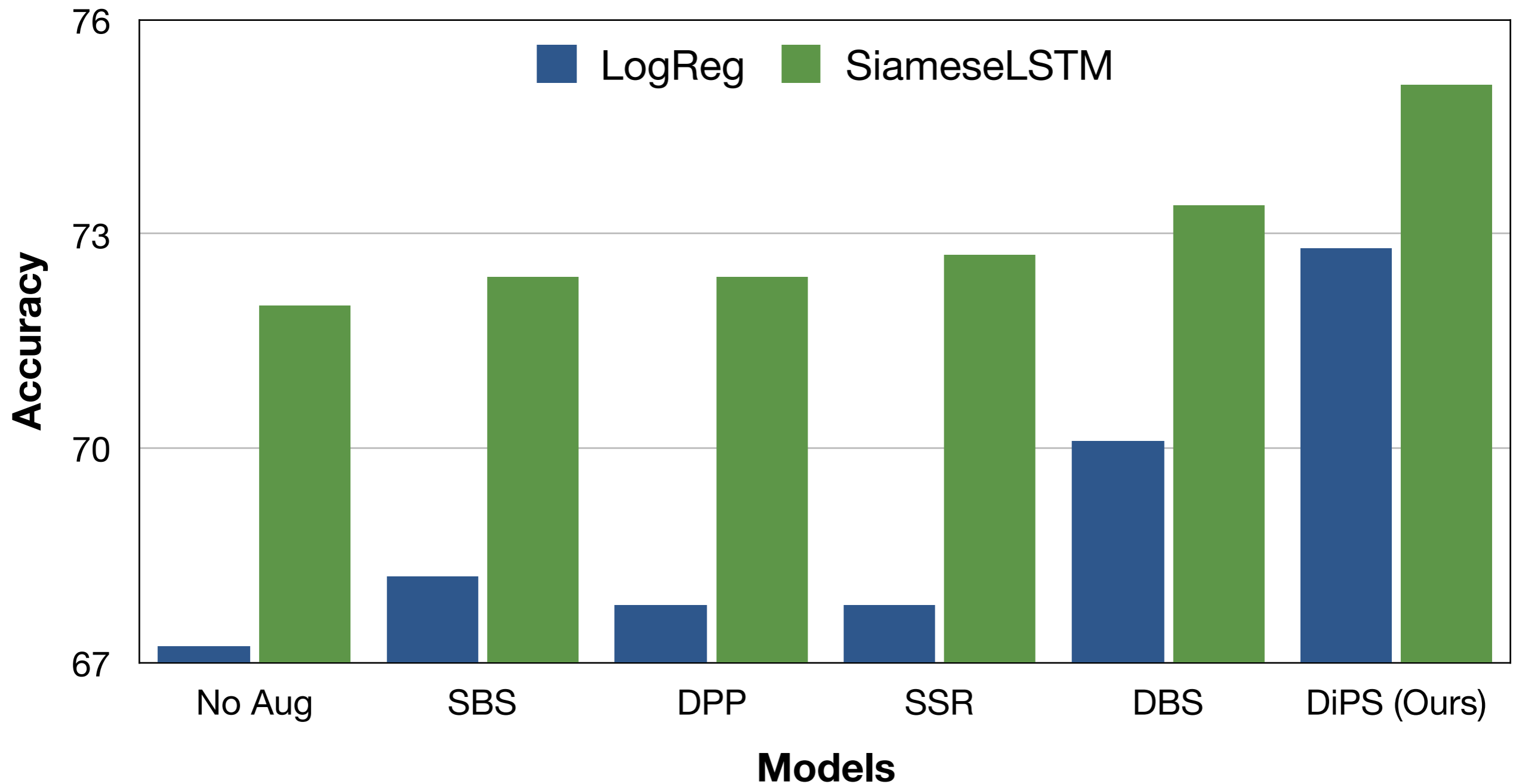


DiPS induces diversity without compromising on fidelity



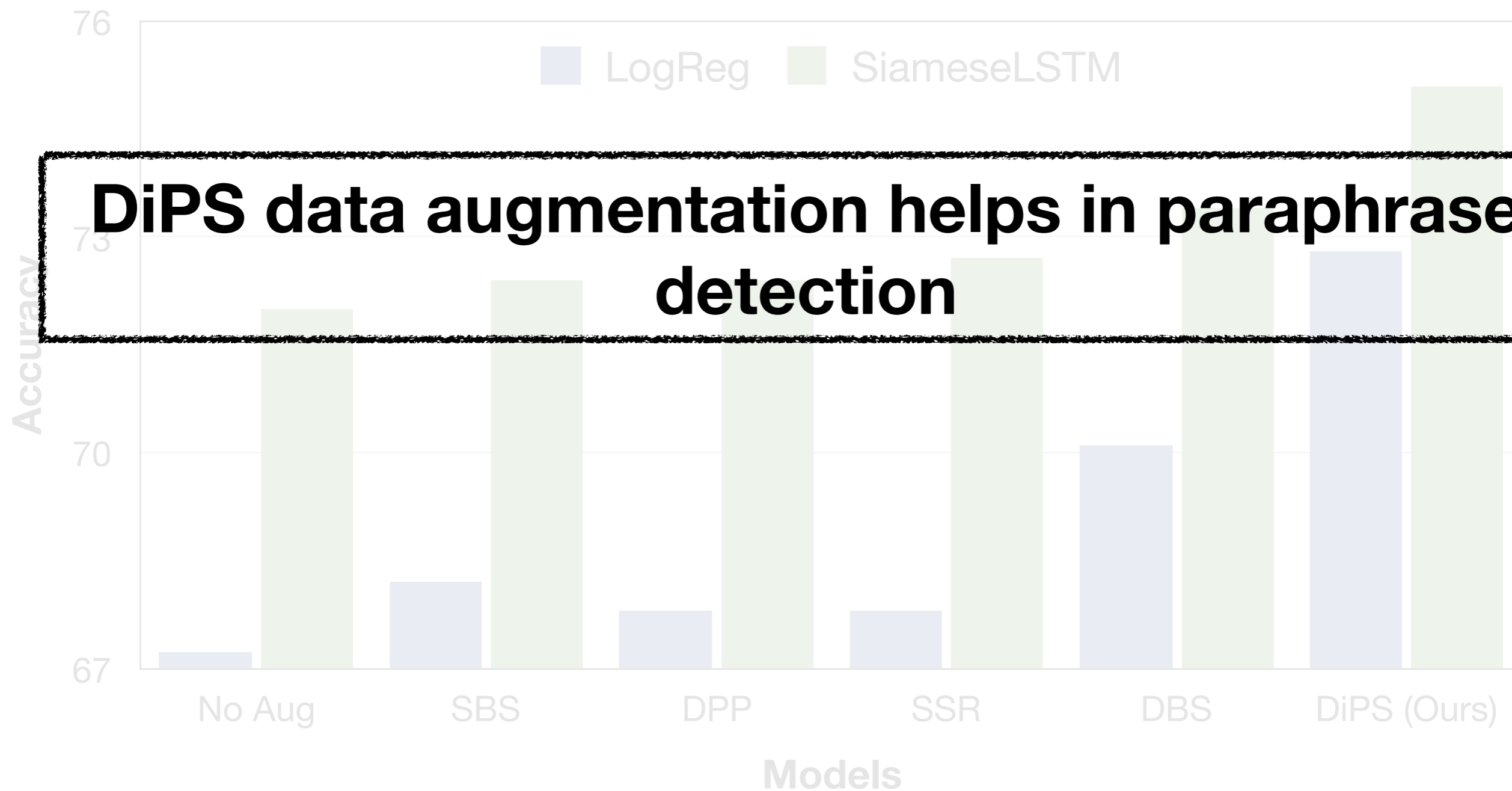
Data Augmentation Paraphrase Detection

Quora Dataset



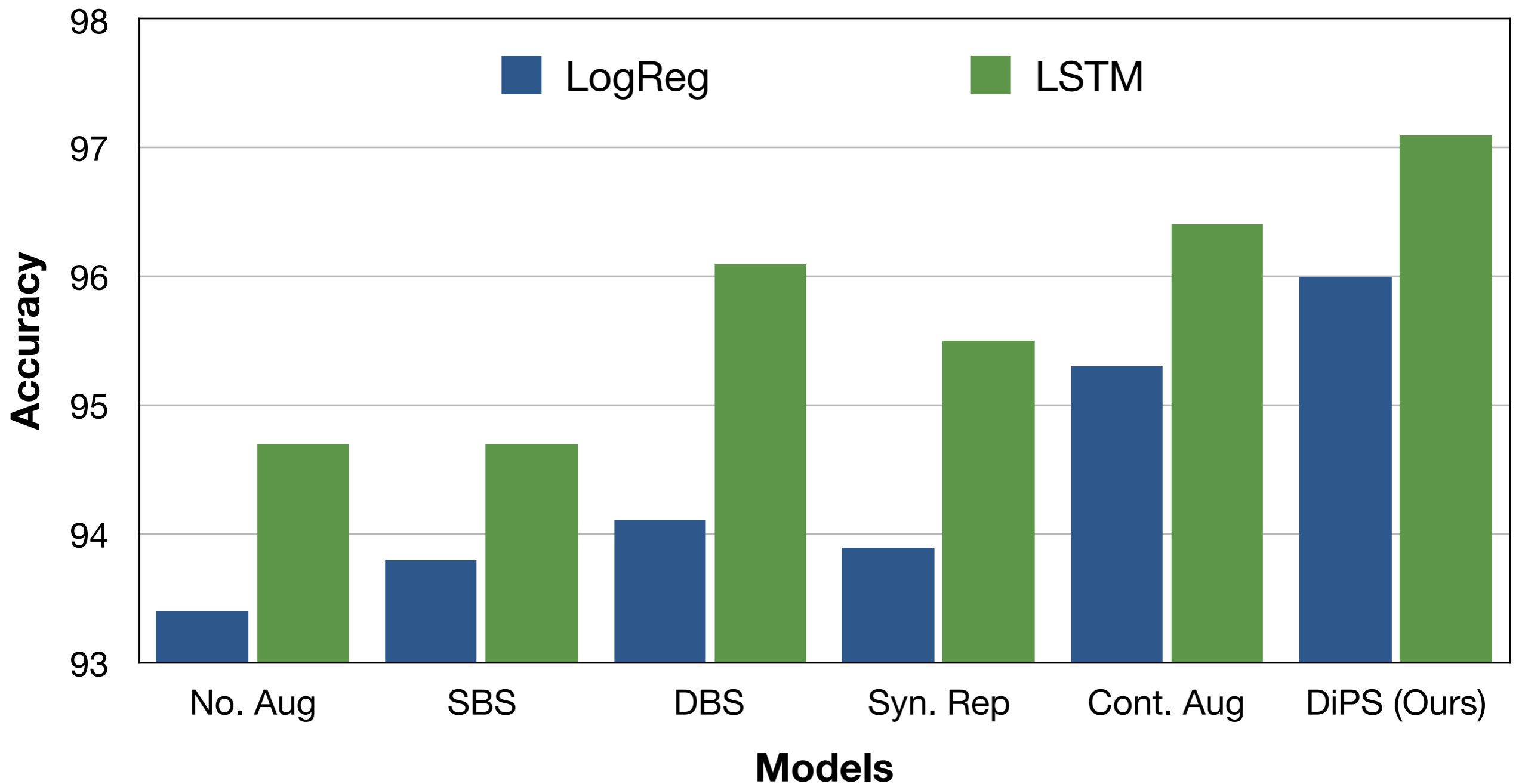
Data Augmentation Paraphrase Detection

Quora Dataset



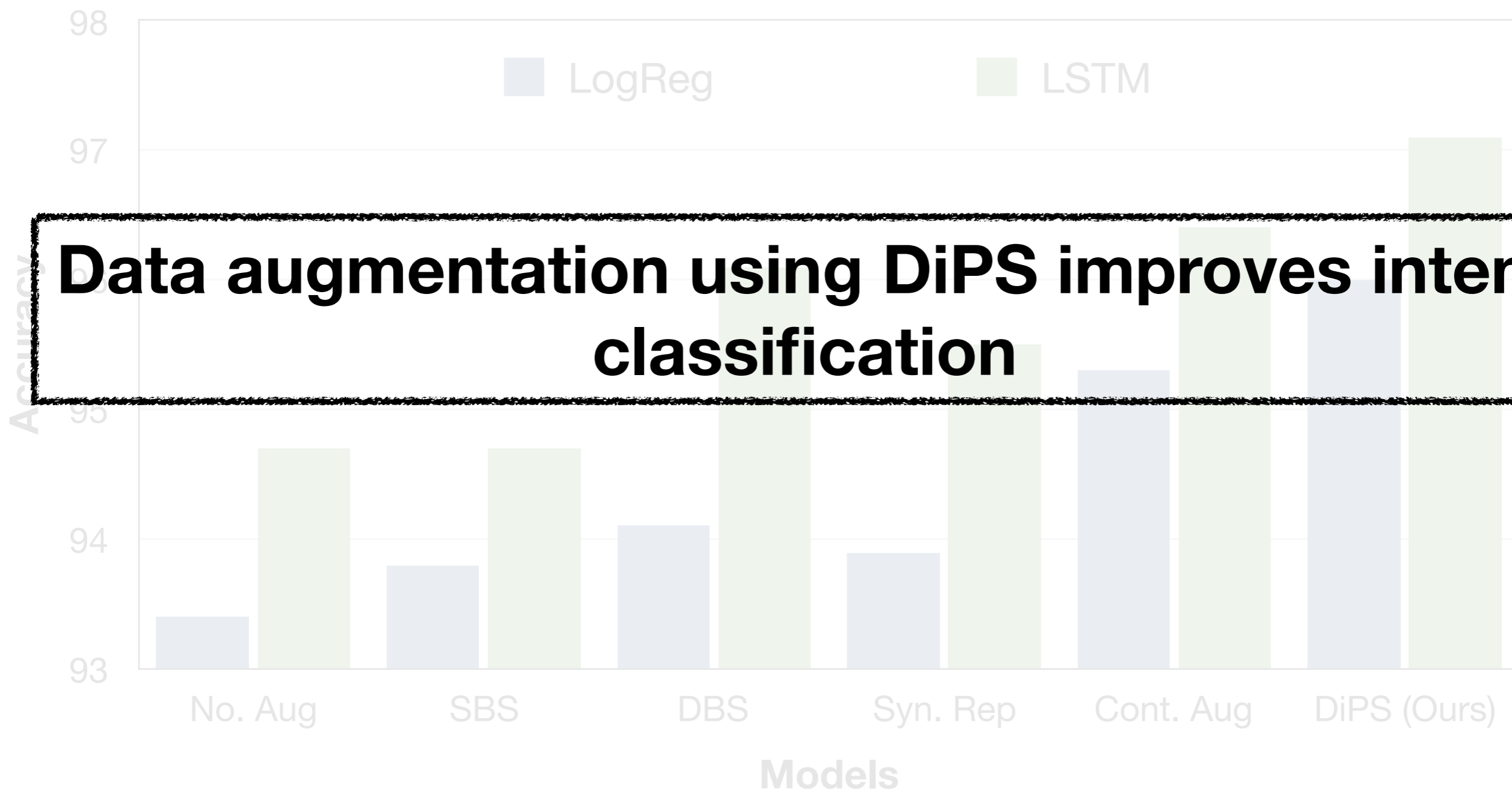
Data Augmentation for Intent Classification

Dataset : SNIPS



Data Augmentation for Intent Classification

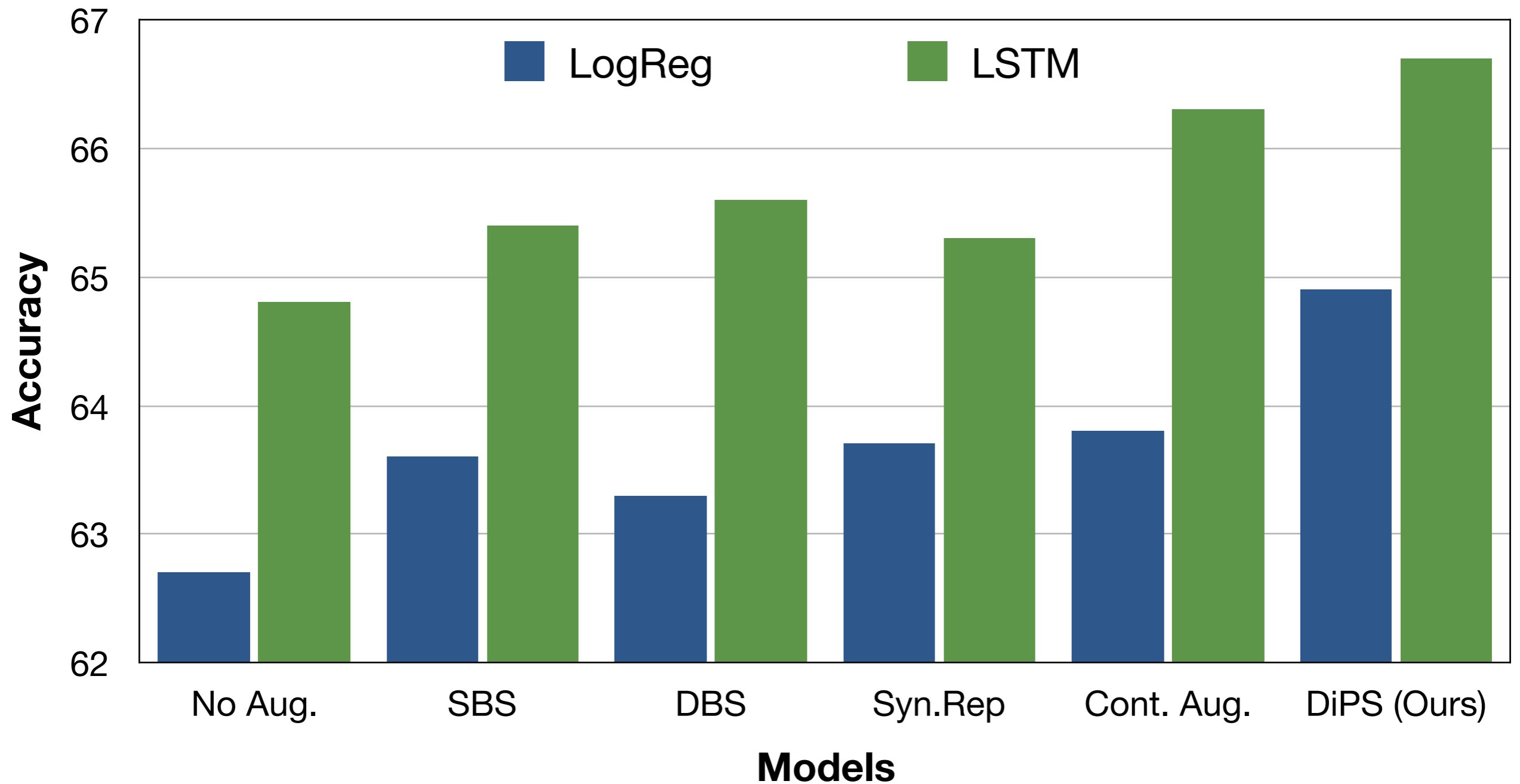
Dataset : SNIPS



Data augmentation using DiPS improves intent classification

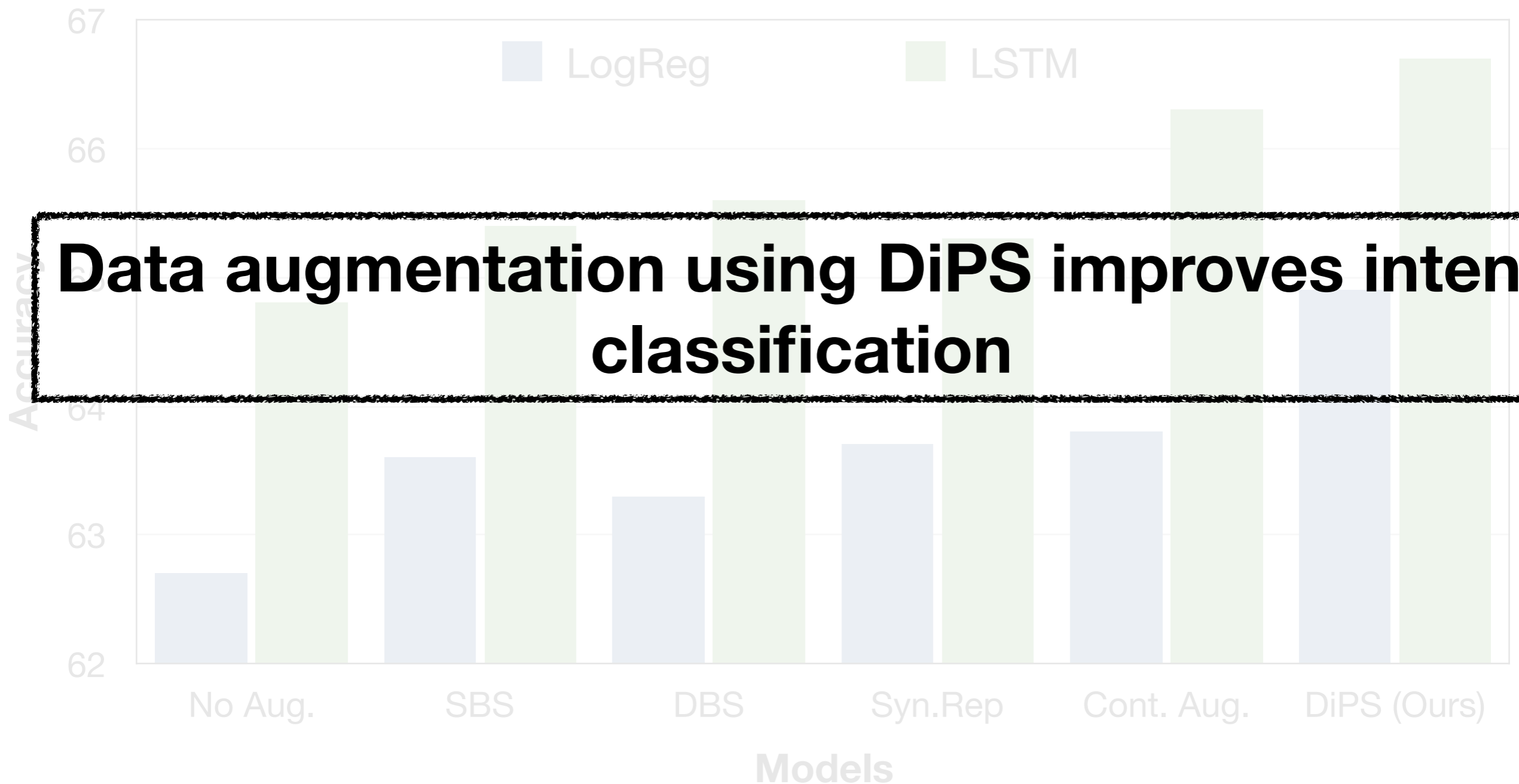
Data Augmentation for Intent Classification

Dataset : Yahoo-L31



Data Augmentation for Intent Classification

Dataset : Yahoo-L31



Data augmentation using DiPS improves intent classification

Conclusion

Conclusion

Problem

Diversity in
Paraphrases

Without
compromising
on fidelity

Conclusion

Problem

Diversity in
Paraphrases

Without
compromising
on fidelity

Method

DiPS

Sub-modular
optimisation

Conclusion

Problem

Diversity in
Paraphrases

Without
compromising
on fidelity

Method

DiPS

Sub-modular
optimisation

Take-Aways

Seq2Seq + Diversity

Data Augmentation
Using Paraphrasing

Code

<https://github.com/mallabiisc/DiPS>



Code

<https://github.com/mallabiisc/DiPS>



Acknowledgement

Code

<https://github.com/mallabiisc/DiPS>



Acknowledgement



सत्यमेव जयते
Government of India
Ministry of Human Resource
Development



Microsoft®
Research

amazon

Code

<https://github.com/mallabiisc/DiPS>



Acknowledgement



Microsoft®
Research

amazon

Find us on 

ashutosh@iisc.ac.in, satwik55@gmail.com

Code

<https://github.com/mallabiisc/DiPS>



Acknowledgement



Microsoft®
Research

amazon

Find us on 

ashutosh@iisc.ac.in, satwik55@gmail.com

Thank you