How do cross-lingual semantic divergences impact neural machine translation?

Eleftheria Briakou

What is a parallel text?

What is a parallel text?

“a parallel text is a text placed alongside its translation or translations” *

* Wikipedia: https://en.wikipedia.org/wiki/Parallel_text

How to obtain parallel texts?

Human translation
e.g., FLORES

Alignment of translated documents
e.g., ParaCrawl

Mining from monolingual texts
e.g., WikiMatrix

Machine Translation
e.g., XNLI

Parallel texts in the MT pipeline

Human translation
e.g., FLORES

Alignment of translated documents
e.g., ParaCrawl

Mining from monolingual texts
e.g., WikiMatrix

Machine Translation
e.g., XNLI

Parallel texts in the MT pipeline

Human translation
e.g., FLORES

EVALUATION DATA

Alignment of translated documents
e.g., ParaCrawl

Mining from monolingual texts
e.g., WikiMatrix

Machine Translation
e.g., XNLI

Parallel texts in the MT pipeline

Human translation
e.g., FLORES

Alignment of translated documents
e.g., ParaCrawl

Mining from monolingual texts
e.g., WikiMatrix

TRAINING DATA

Machine Translation
e.g., XNLI

Parallel texts in the MT pipeline

Human translation
e.g., FLORES

Alignment of translated documents
e.g., ParaCrawl

Mining from monolingual texts
e.g., WikiMatrix

Machine Translation
e.g., XNLI

PSEUDO TRAINING DATA

Parallel texts in the MT pipeline

Alignment of translated documents

e.g., ParaCrawl

Mining from monolingual texts

e.g., WikiMatrix

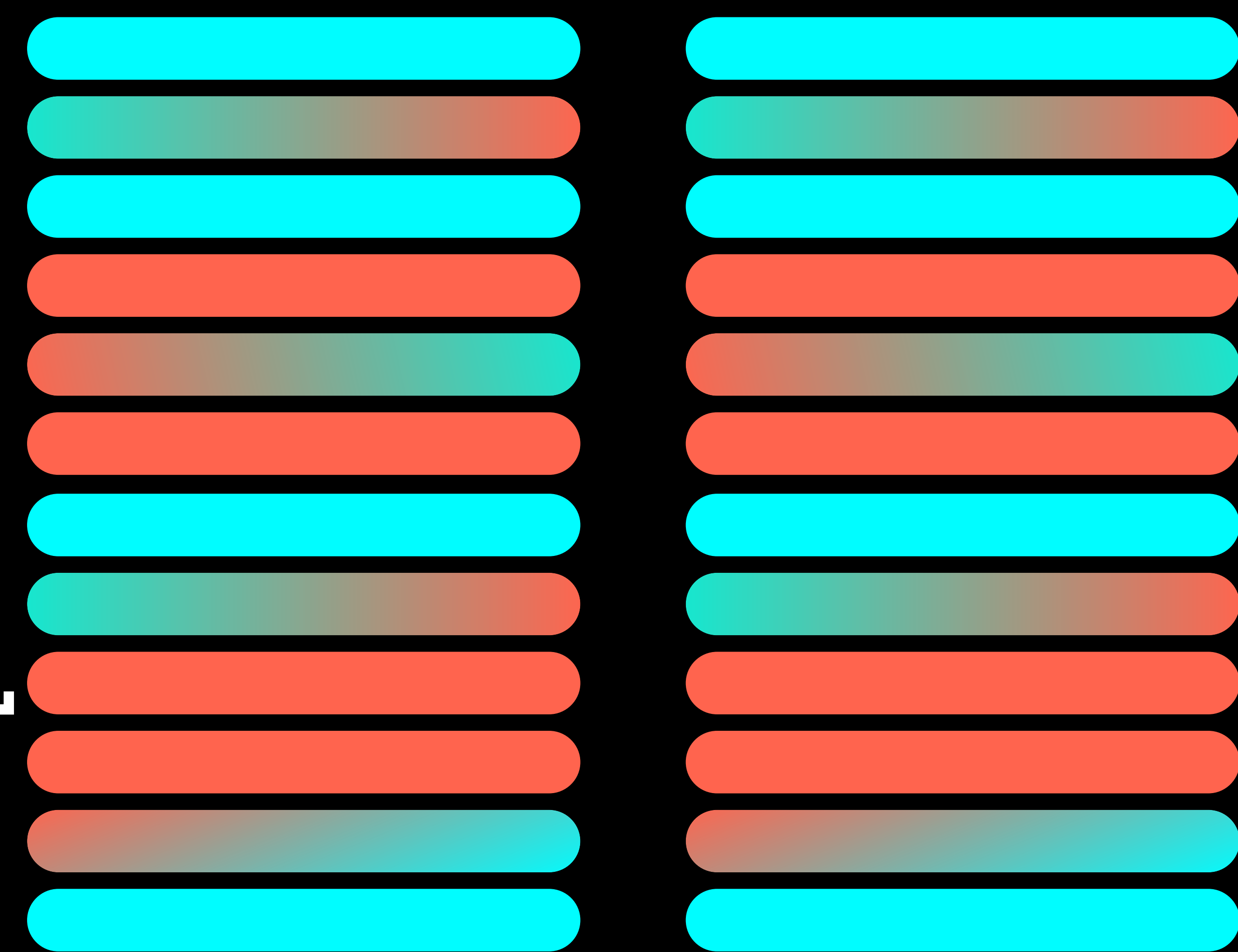
Automatic extraction of parallel texts introduces noise

On the Impact of Various Types of Noise on Neural Machine Translation; Huda Khayrallah and Philipp Koehn; WNLG 2018

Automatic extraction of parallel texts introduces noise

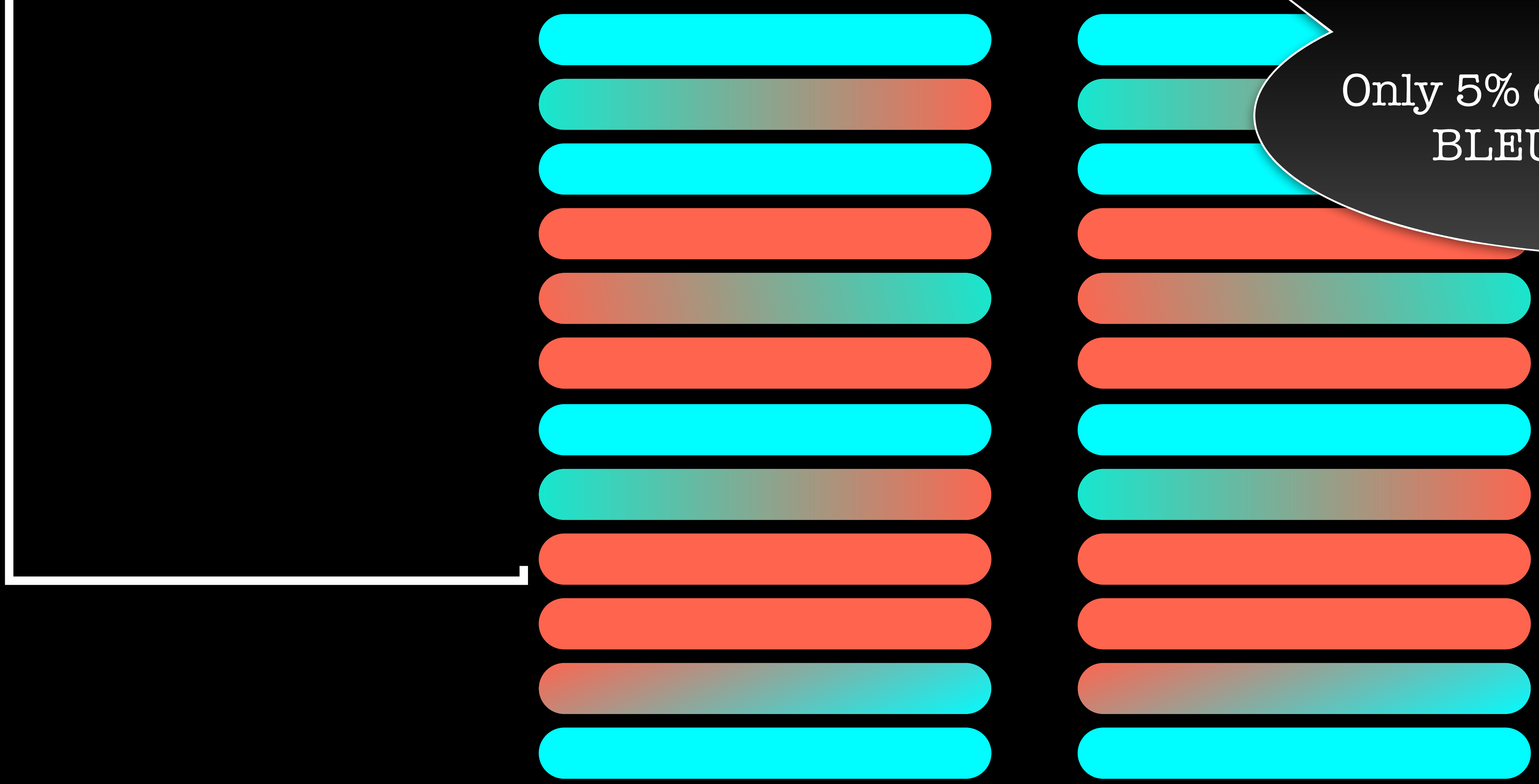
EN All helicopters have adjustments
DE All helicopters have adjustments

Input copy



Automatic extraction of parallel texts introduces noise

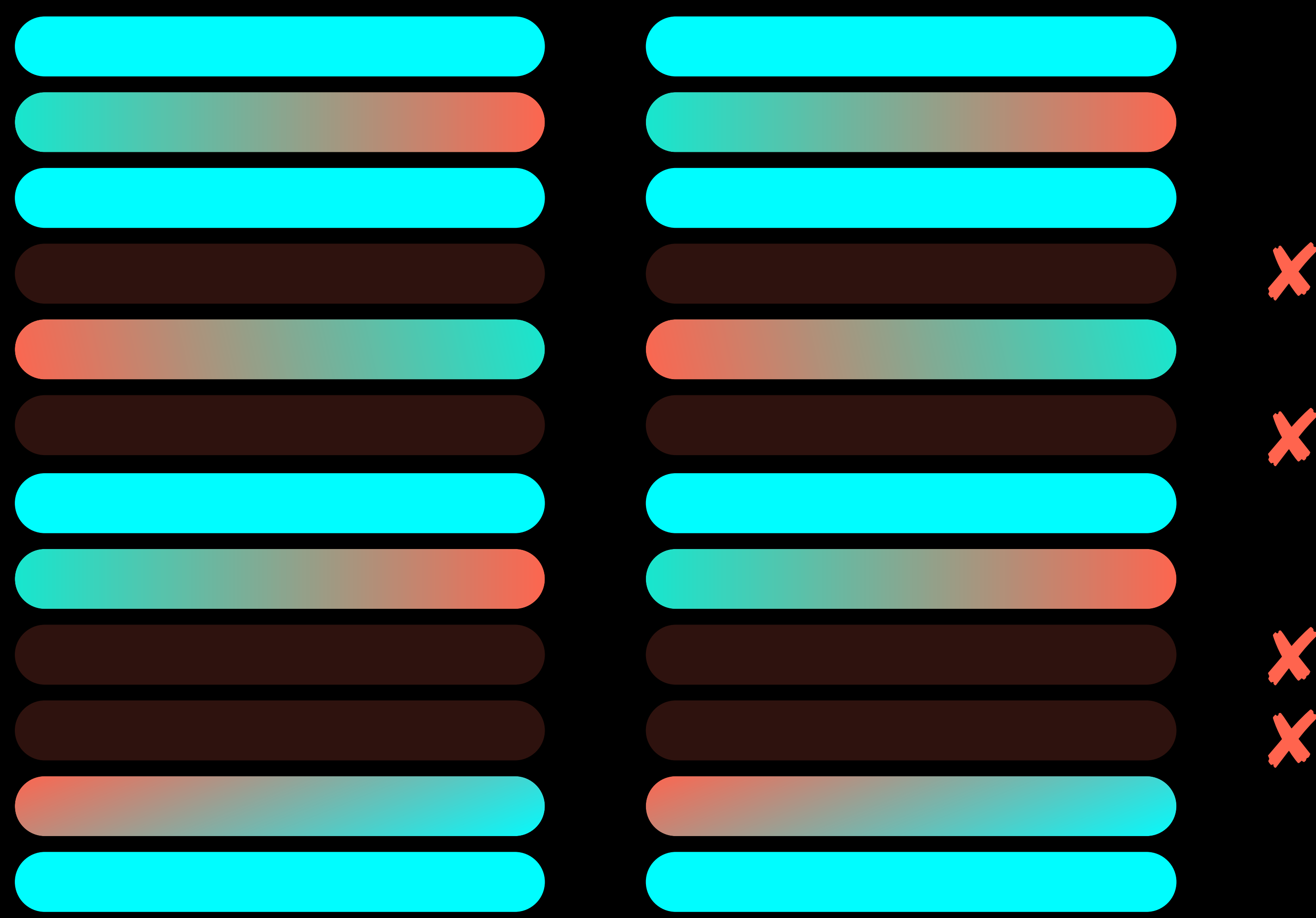
EN All helicopters have adjustments
DE All helicopters have adjustments



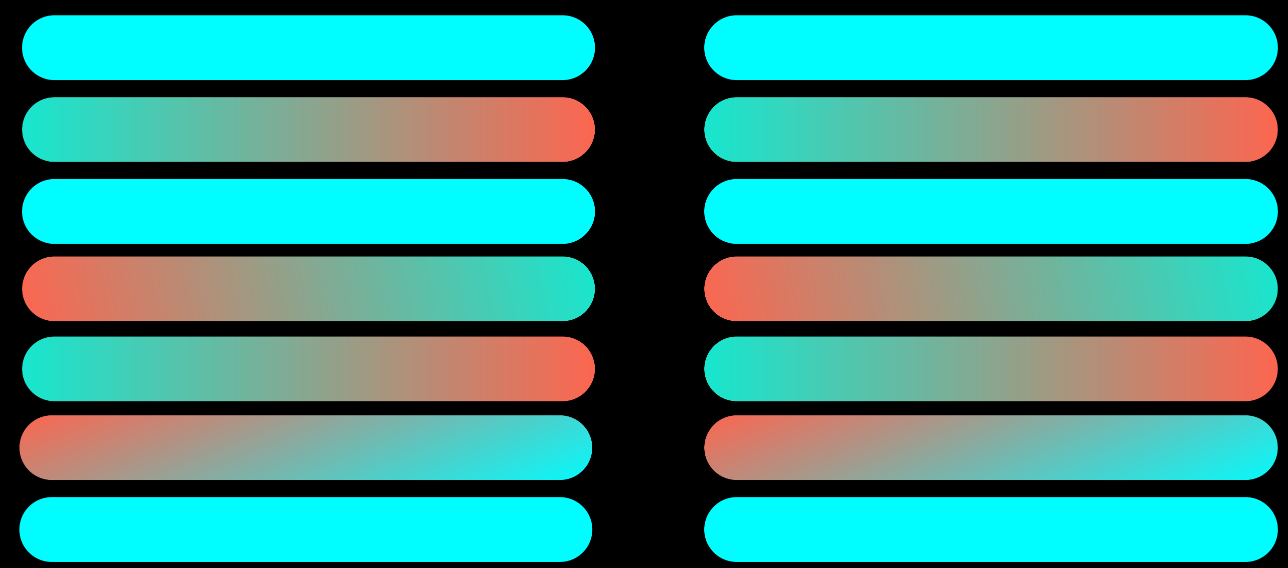
Only 5% causes approx. 10 BLEU degradation

On the Impact of Various Types of Noise on Neural Machine Translation; Huda Khayrallah and Philipp Koehn; WNLG 2018

After noise filtering...



After noise filtering...
How parallel is "parallel" text?



How parallel is "parallel" text?

EN	After Caesar's death, he joined the party of Cassius.
FR	Après la mort du dictateur il est accusé par Cassius de contre Rome.

How parallel is "parallel" text?

EN	After Caesar's death, he joined the party of Cassius.
FR	Après la mort du dictateur il est accusé par Cassius de contre Rome.

After the death of the dictator he is accused by Cassius of conspiring against Rome.

How parallel is "parallel" text?

EN After Caesar's death, he joined the party of Cassius.

FR Après la mort du dictateur il est accusé par Cassius de contre Rome.

After the death of the dictator he is accused by Cassius of conspiring against Rome.

topically related — coarse meaning differences

How parallel is "parallel" text?

EN	"The Maple Leaf Forever" served for many years as an unofficial Canadian national anthem.
FR	"The Maple Leaf Forever" est un chant patriotique pro canadien anglais.

How parallel is "parallel" text?

EN	"The Maple Leaf Forever" served for many years as an unofficial Canadian national anthem.
FR	"The Maple Leaf Forever" est un chant patriotique pro canadien anglais.

The Maple Leaf Forever is an English Canadian patriotic song.

How parallel is "parallel" text?

added content

EN	"The Maple Leaf Forever" served for **many years** as an **unofficial** Canadian national anthem.
FR	"The Maple Leaf Forever" est un chant patriotique pro canadien **anglais**.

The Maple Leaf Forever is an English Canadian patriotic song.

How parallel is "parallel" text?

mistranslated content

EN	"The Maple Leaf Forever" served for many years as an unofficial Canadian national anthem.
FR	"The Maple Leaf Forever" est un chant patriotique pro canadien anglais.

The Maple Leaf Forever is an English Canadian patriotic song.

How parallel is “parallel” text?

EN “The Maple Leaf Forever" served for **many years** as an **unofficial** Canadian **national anthem.**

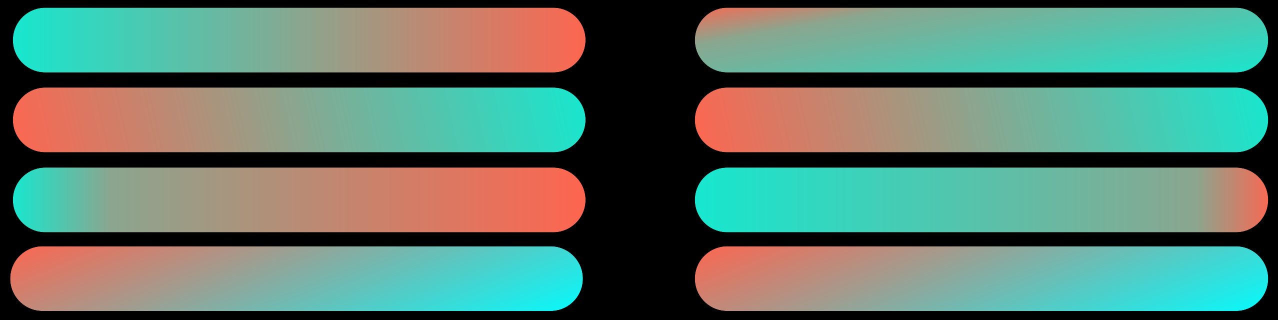
FR “The Maple Leaf Forever” est un **chant patriotique** pro canadien **anglais**.

The Maple Leaf Forever is an English Canadian patriotic song.

shared content — fine-grained meaning differences

Cross-lingual Semantic Divergences

Parallel sentences where source and target do not convey the same meaning



Cross-lingual Semantic Divergences

OUTLINE

- CHAPTER A: How **frequent** are they?
- CHAPTER B: How can we **detect** them?
- CHAPTER C: How do they **impact NMT**?

Cross-lingual Semantic Divergences

OUTLINE

CHAPTER A: How **frequent** are they?

CHAPTER B: How can we **detect** them?

CHAPTER C: How do they **impact NMT**?

Annotating
Cross-lingual Semantic Divergences

CHALLENGES

- annotators without expert knowledge
- divergences vary in their granularity
- annotator agreement

Annotating
Cross-lingual Semantic Divergences

Annotation Protocol

Goal: encourage annotator's sensitivity to subtle meaning differences

Annotating Cross-lingual Semantic Divergences

Annotation Protocol

Goal: encourage annotator's sensitivity to subtle meaning differences

Rationalized
English
FREnch
Semantic
Divergences

REFRESD: Annotation Protocol

Given an English-French WikiMatrix sentence-pair

She made a courtesy call to the Hawaiian Islands.

Il fait une escale aux îles Hawaï.

REFRESD: Annotation Protocol

Given an English-French WikiMatrix sentence-pair

She made a courtesy call to the Hawaiian Islands.

Il fait une escale aux îles Hawaï.

rationales

A. highlight spans that differ in meaning

REFRESD: Annotation Protocol

Given an English-French WikiMatrix sentence-pair

She made a courtesy call to the Hawaiian Islands.

Il fait une escale aux îles Hawaï.

rationales

A. highlight spans that differ in meaning

B. make sentence-level judgment

NO MEANING DIFFERENCE
SOME MEANING DIFFERENCE
UNRELATED

distinct classes

REFRESD: Annotation findings

- Rationales improve annotator agreement

  Krippendorf's α: 0.60

- Semantic divergences are frequent in REFRESD

  64% semantic divergences

CHAPTER A REVISITED:

How frequent are semantic divergences?

24% Unrelated

40% Some meaning difference

36% No meaning difference

Cross-lingual Semantic Divergences

OUTLINE

~~CHAPTER A: How frequent are they?~~

CHAPTER B: How can we detect them?

CHAPTER C: How do they impact NMT?

Detecting Semantic Divergences: Problem definition

INPUT

She made a courtesy call to the Hawaiian Islands.
Il fait une escale aux îles Hawaï.

OUTPUT

EQUIVALENCE VS. DIVERGENCE

Detecting Semantic Divergences: Problem definition

INPUT

She made a courtesy call to the Hawaiian Islands.
Il fait une escale aux îles Hawaï.

OUTPUT

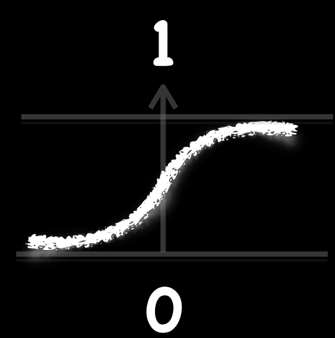
EQUIVALENCE VS. DIVERGENCE

✓ no human-annotated training data
✓ divergences can be fine-grained

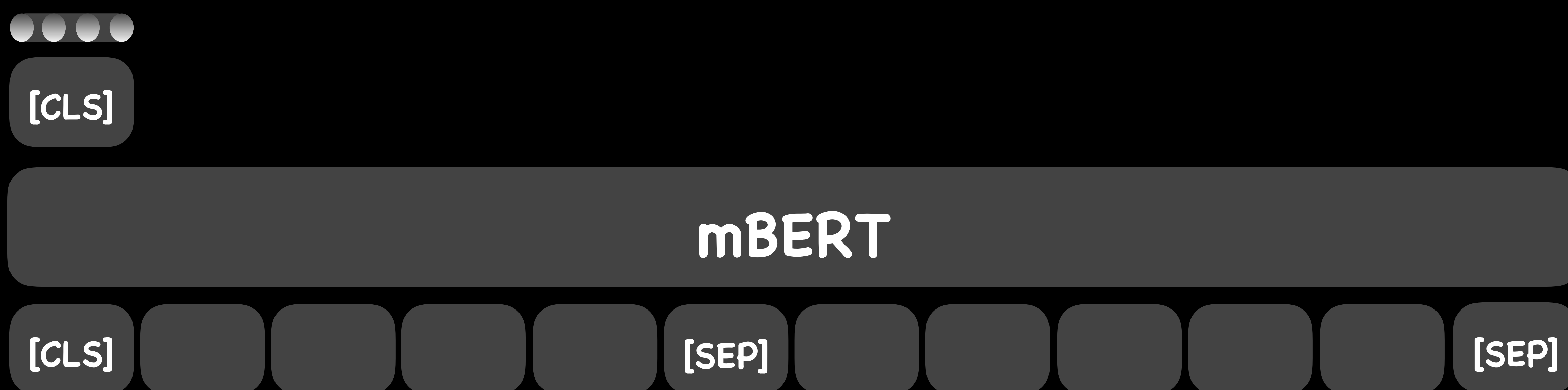
Divergent mBERT

parallel

Divergent mBERT



F(**parallel**) ⟶ probability of being equivalent



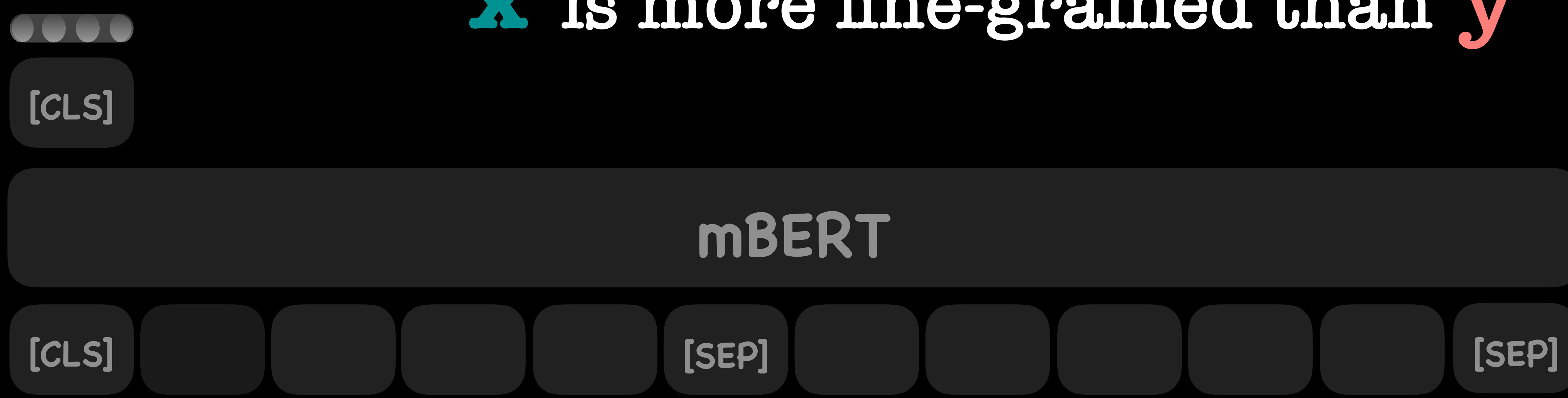
parallel

Divergent mBERT

$$D = \{(\mathbf{x}, \mathbf{y})\}$$

contrastive pair

$\mathbf{x}$ is more fine-grained than $\mathbf{y}$

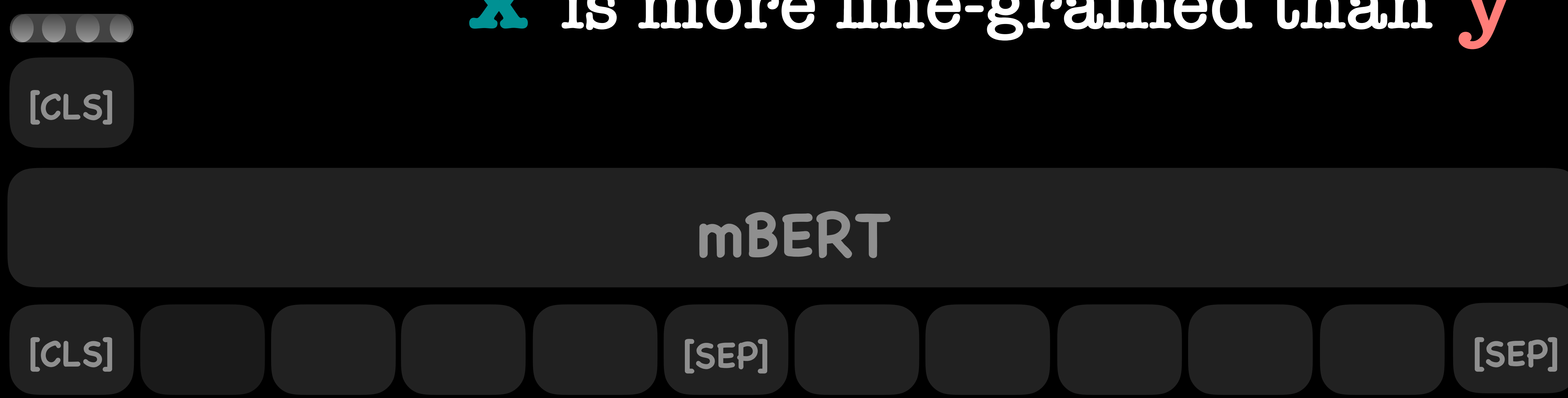


Divergent mBERT

Learning to rank contrastive pairs

$$\max\{0, \xi - \mathrm{F}(\mathbf{x}) - \mathrm{F}(\mathbf{y})\}$$

x is more fine-grained than **y**



Predicting token divergences:
Problem definition

INPUT

She made a courtesy call to the Hawaiian Islands.
Il fait une escale aux îles Hawaï.

OUTPUT

EQ EQ EQ DIV DIV EQ EQ EQ DIV

EQ EQ EQ DIV DIV EQ EQ

Divergent mBERT
Token-level prediction

Η συνθήκη είναι φτωχή The economic situation was poor

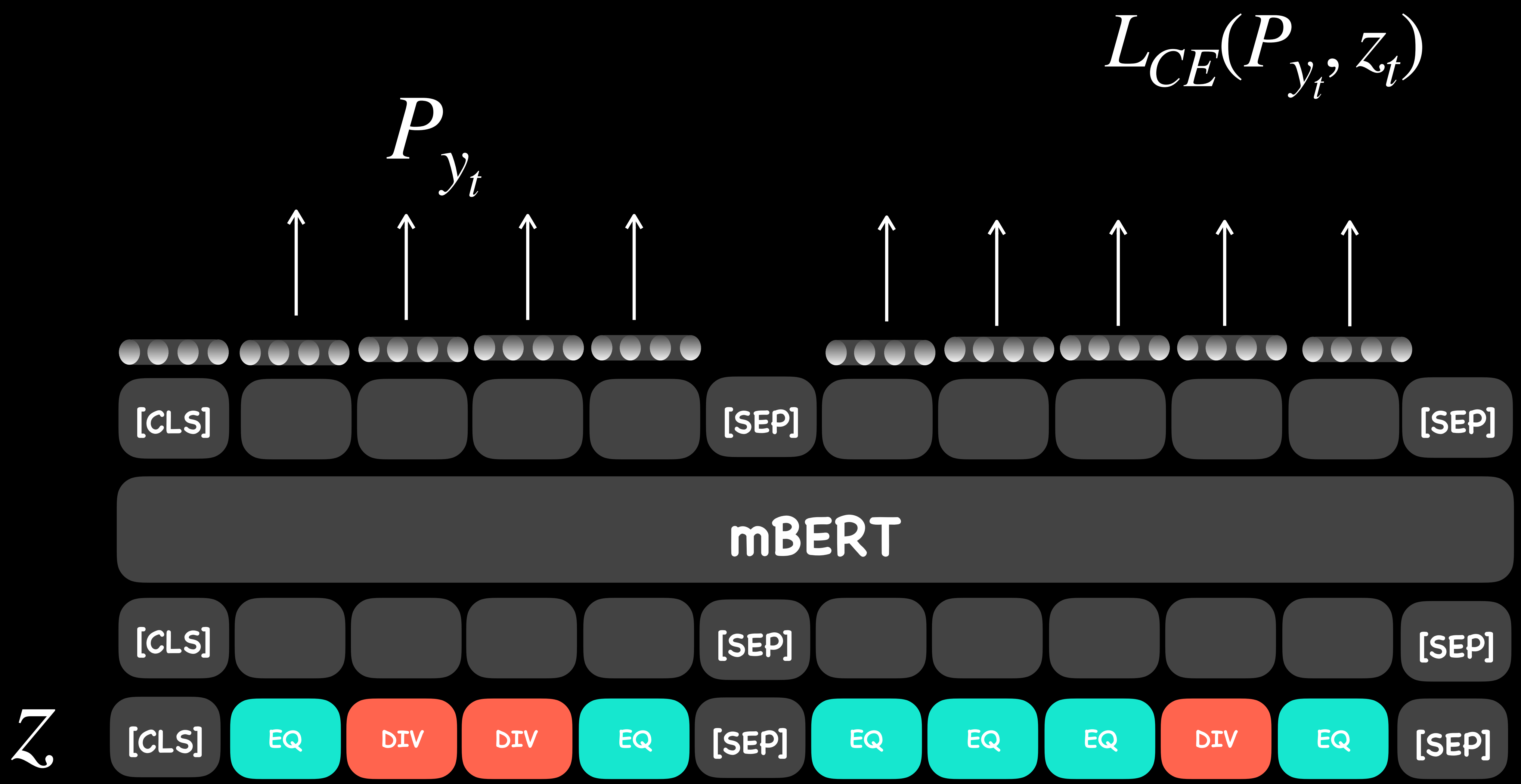
Z [CLS] EQ DIV DIV EQ [SEP] EQ EQ EQ DIV EQ [SEP]

Divergent mBERT

Token-level prediction

Divergent mBERT

Token-level prediction



Divergent mBERT

Multi-task variant

Sentence prediction

Token prediction

$$\max\{0, \xi, -F(x) + F(y)\} + \frac{1}{|y|} \sum_{t=1}^{|y|} L_{CE}(P_{y_t}, z_t)$$

Learn to rank contrastive pairs & predict divergent tokens

Synthetic training data

Synthetic training data

Seed equivalent

Now however one of them is suddenly asking your help and you can see from this how weak they are.

Maintenant cependant l'un d'eux vient soudainement demander votre aide et vous pouvez voir à quel point ils sont faibles

Synthetic training data

Subtree Deletion

Now however one of them is suddenly asking your help and you can see from this ~~how weak they are.~~

Maintenant cependant l'un d'eux vient soudainement demander votre aide et vous pouvez voir à quel point ils sont faibles

Synthetic training data

Phrase Replacement

Now however one of them is suddenly asking your help and you can see from this how weak they are.

Maintenant cependant l'un d'eux vient soudainement demander votre aide et vous pouvez voir à quel point ils sont faibles

Synthetic training data
Phrase Replacement

Now however one of them is absolutely fighting his policy and you can see from this how weak they are.

Maintenant cependant l'un d'eux vient soudainement demander votre aide et vous pouvez voir à quel point ils sont faibles

Synthetic training data

Lexical Substitution

Now however one of them is suddenly asking your **help** and you can see from this how weak they are.

Maintenant cependant l'un d'eux vient soudainement demander votre **aide** et vous pouvez voir à quel point ils sont faibles

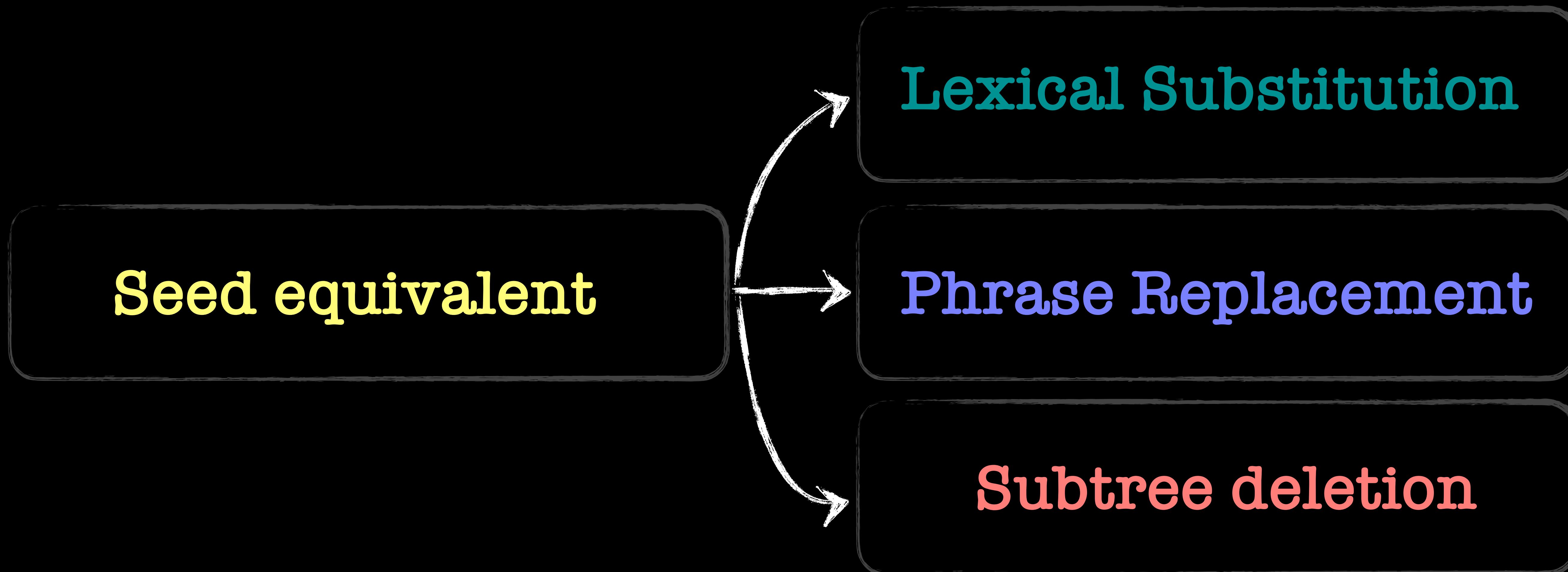
Synthetic training data

Lexical Substitution

Now however one of them is suddenly asking your **mercy** and you can see from this how weak they are.

Maintenant cependant l'un d'eux vient soudainement demander votre **aide** et vous pouvez voir à quel point ils sont faibles

Contrastive pairs: Divergences contrast with specific seed



Divergence ranking:
Learning to rank contrastive divergences

Rank contrastive divergences of increasing granularity

Seed equivalent	>	Lexical Substitution
Lexical Substitution	>	Phrase Replacement
Lexical Substitution	>	Subtree deletion

Divergence detection:
Evaluation on REFRESD

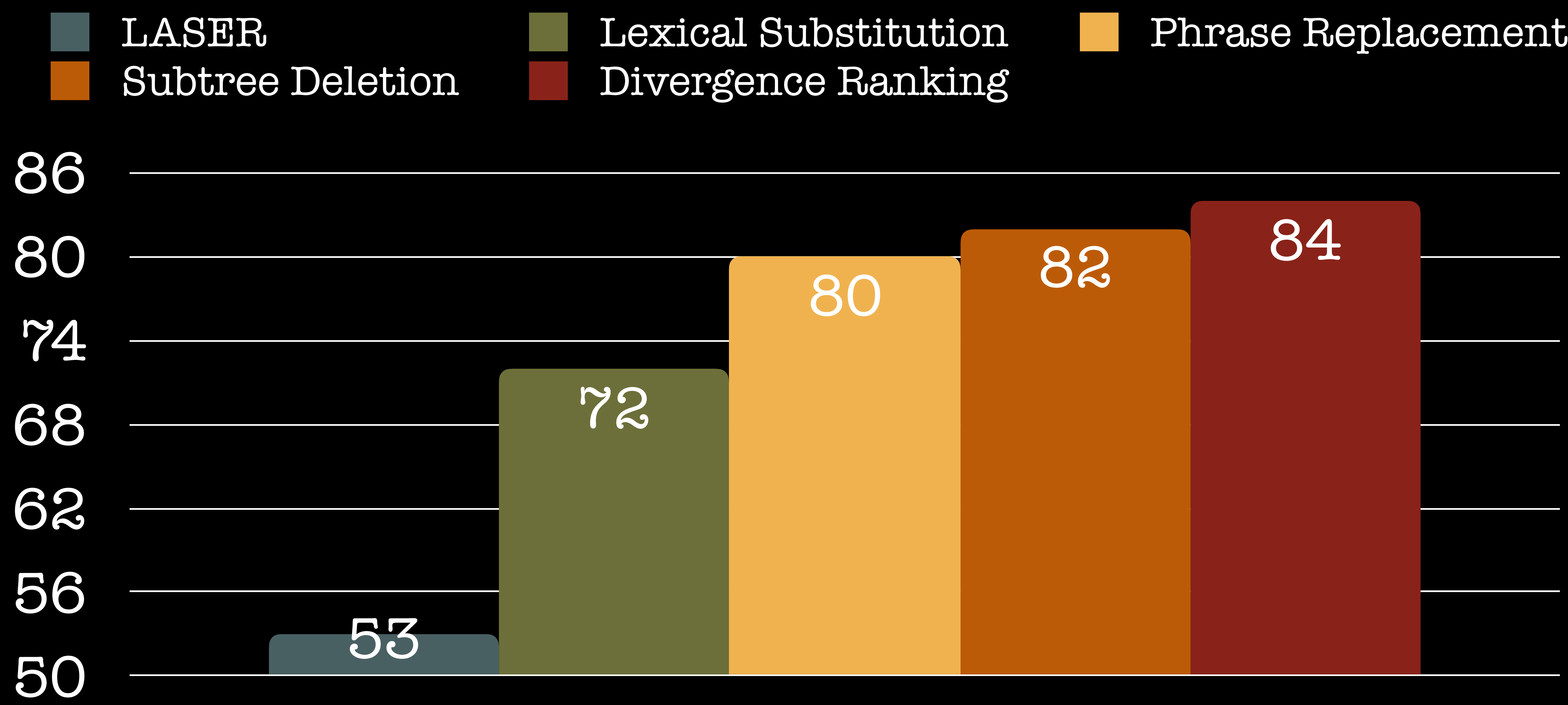
Sentence prediction

NO MEANING DIFFERENCE	Equivalent
SOME MEANING DIFFERENCE UNRELATED	} Divergent

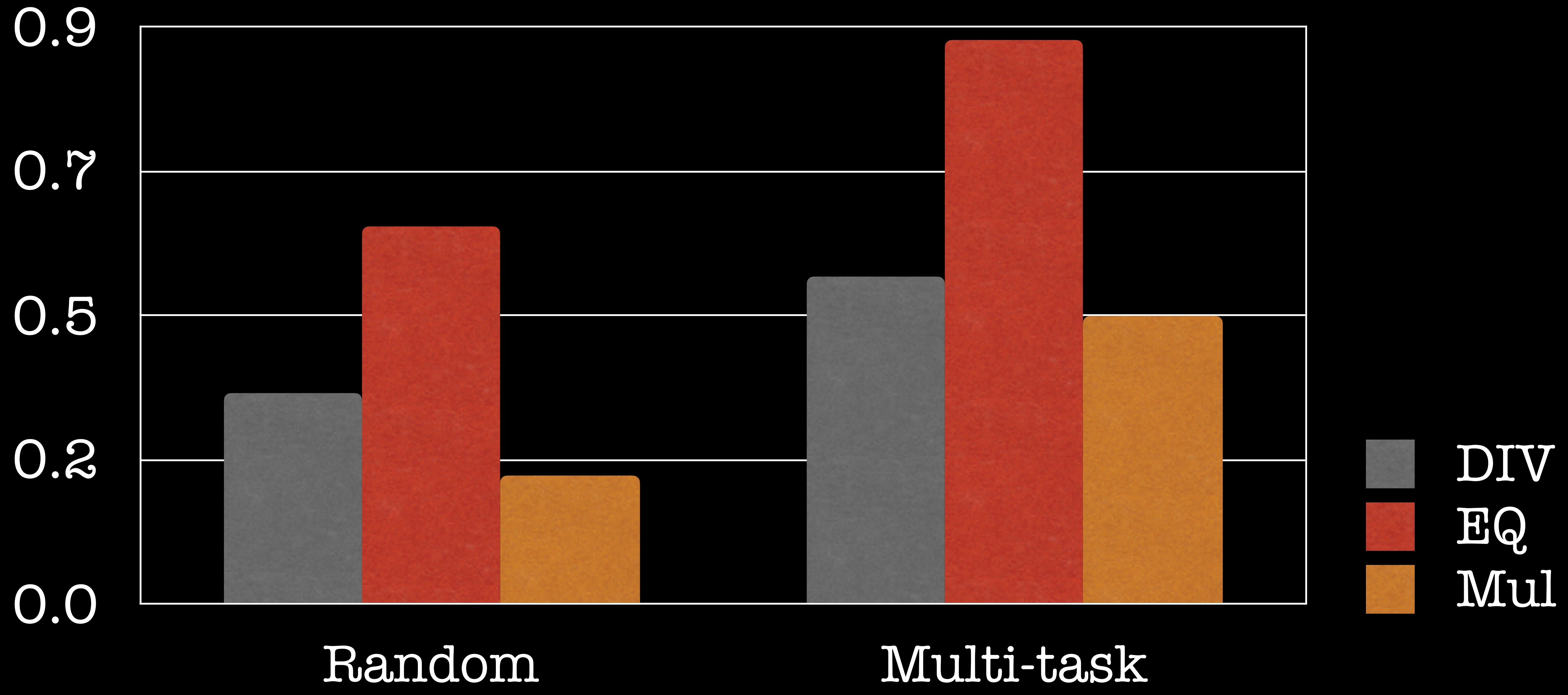
Token prediction

SOME MEANING DIFFERENCE	} Rationales

Divergence Ranking Exploits Diverse Synthetic Samples Better



Divergences Ranking yields moderate results on token prediction



Detecting Fine-Grained Cross-Lingual Semantic Divergences without Supervision by Learning to Rank; Eleftheria Briakou and Marine Carpuat; EMNLP 2020

CHAPTER B REVISITED:

Can we automatically detect divergences?

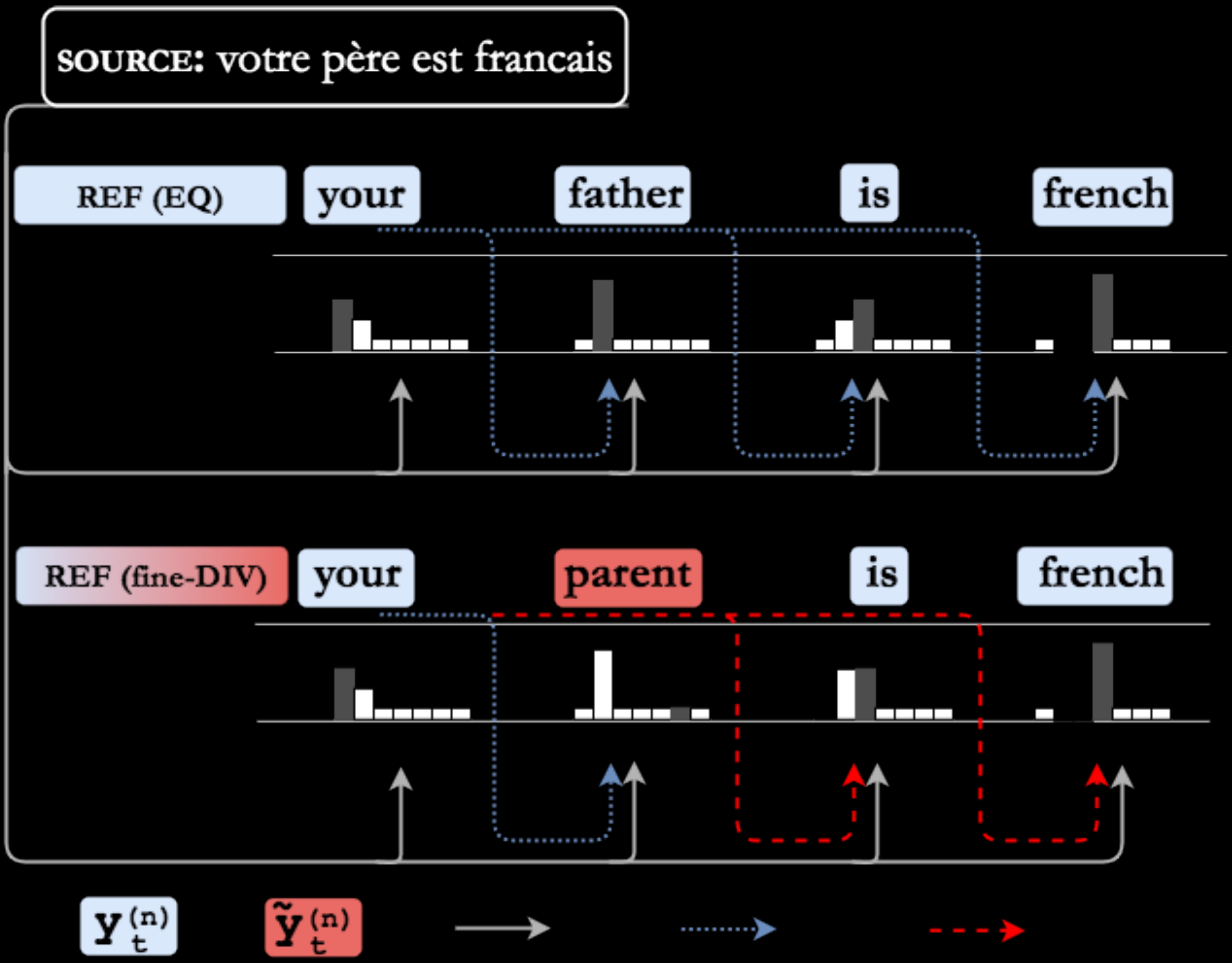
- without supervision
- by learning to rank divergences
- at sentence & token level

Cross-lingual Semantic Divergences

OUTLINE

- ~~CHAPTER A: How frequent are they?~~
- ~~CHAPTER B: How can we detect them?~~
- **CHAPTER C: How do they impact NMT?**

Assumptions of semantic equivalence in Neural Machine Translation



Beyond noise: Mitigating the impact of Fine-grained Divergences on Neural Machine Translation; Eleftheria Briakou and Marine Carpuat; ACL 2021

Divergences matter for NMT because they yield unreliable training signals

$$J(\theta) = \sum_{n=1}^{N} \sum_{t=1}^{T} \log p(y_t^{(n)} \mid y_{<t}^{(n)}, x^{(n)}; \theta)$$

votre père est français

your parent is french

Divergences matter for NMT because they yield unreliable training signals

$$J(\theta) = \sum_{n=1}^{N} \sum_{t=1}^{T} \log p(y_t^{(n)} \mid y_{<t}^{(n)}, \boxed{x^{(n)}}; \theta)$$

votre père est français

your parent is french

Divergences matter for NMT because they yield unreliable training signals

$$J(\theta) = \sum_{n=1}^{N} \sum_{t=1}^{T} \log p(y_t^{(n)} \mid y_{<t}^{(n)}, x^{(n)}; \theta)$$

$$t = 1$$

votre père est français

your parent is french

Divergences matter for NMT because they yield unreliable training signals

$$J(\theta) = \sum_{n=1}^{N} \sum_{t=1}^{T} \log p(y_t^{(n)} \mid y_{<t}^{(n)}, x^{(n)}; \theta)$$

$t = 2$

votre père est français

your parent is french

Divergences matter for NMT because they yield unreliable training signals

$$J(\theta) = \sum_{n=1}^{N} \sum_{t=1}^{T} \log p(\boxed{y_t^{(n)}} \mid \boxed{y_{<t}^{(n)}}, \boxed{x^{(n)}}; \theta)$$

$t = 3$

votre père est français	your parent is french

Divergences matter for NMT because they yield unreliable training signals

$$J(\theta) = \sum_{n=1}^{N} \sum_{t=1}^{T} \log p(y_t^{(n)} \mid y_{<t}^{(n)}, x^{(n)}; \theta)$$

$t = 4$

votre père est français | your parent is french

How do fine-grained divergences impact NMT?

Controlled analysis on artificial divergences

Experimental Setting

Training bitext	WikiMatrix (mined)
Test set	TED
Language-pair	French to English
NMT architecture	Transformer

Measuring the impact of synthetic divergences on NMT

EQUIVALENT

PHRASE DELETION

LEXICAL SUBSTITUTION

PHRASE REPLACEMENT

Measuring the impact of synthetic divergences on NMT

EQUIVALENT

PHRASE DELETION

LEXICAL SUBSTITUTION

PHRASE REPLACEMENT

Gradually corrupt equivalents

20%

20%

20%

Measuring the impact of synthetic divergences on NMT

EQUIVALENT

PHRASE DELETION

LEXICAL SUBSTITUTION

PHRASE REPLACEMENT

Gradually corrupt equivalents

50%

50%

50%

Measuring the impact of synthetic divergences on NMT

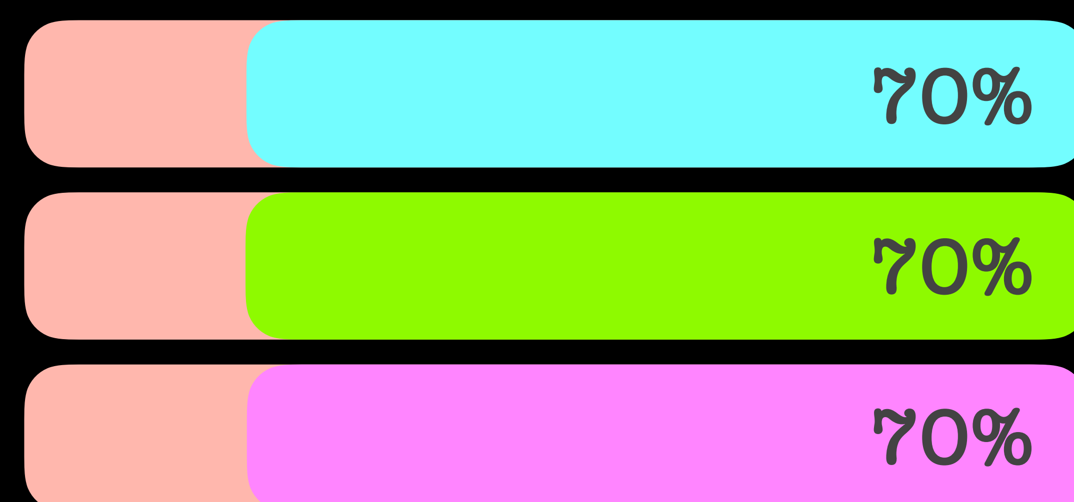
EQUIVALENT

PHRASE DELETION

LEXICAL SUBSTITUTION

PHRASE REPLACEMENT

Gradually corrupt equivalents



Measuring the impact of synthetic divergences on NMT

EQUIVALENT

PHRASE DELETION

LEXICAL SUBSTITUTION

PHRASE REPLACEMENT

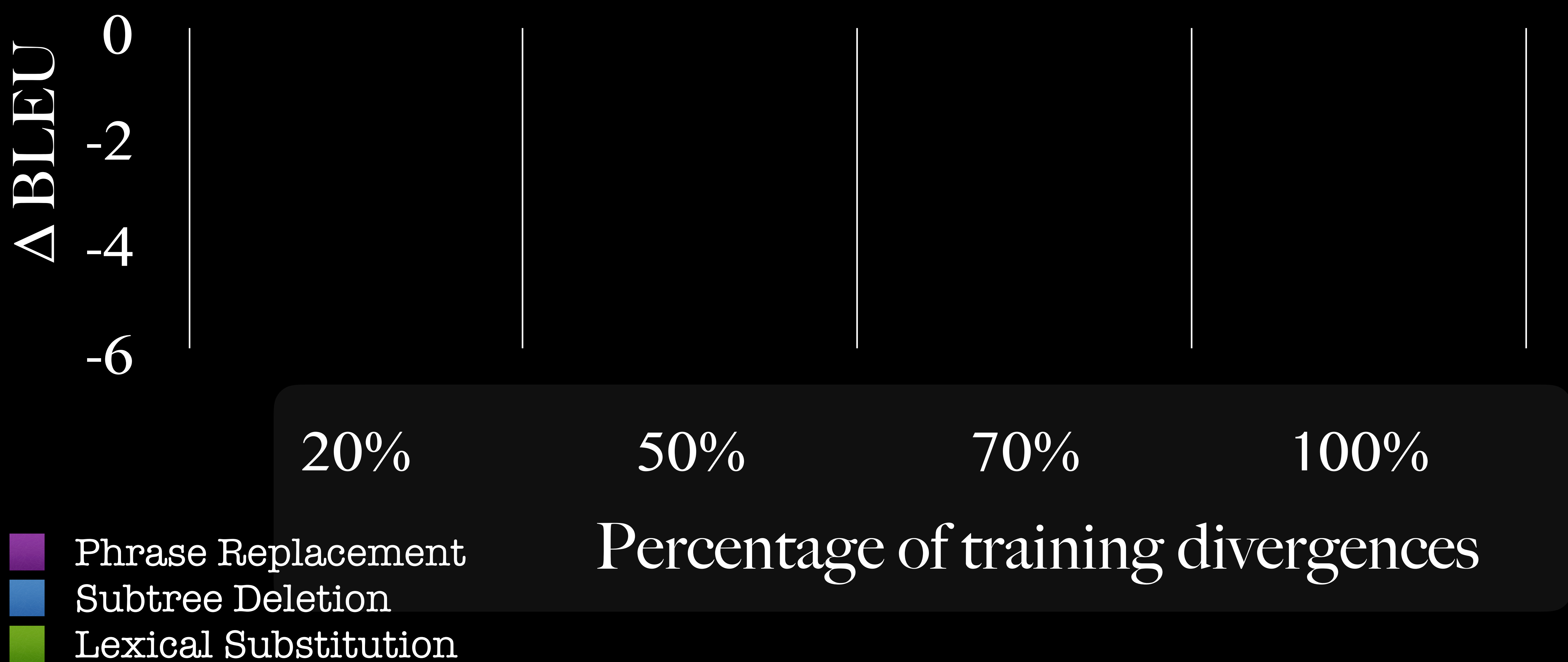
Gradually corrupt equivalents

100%

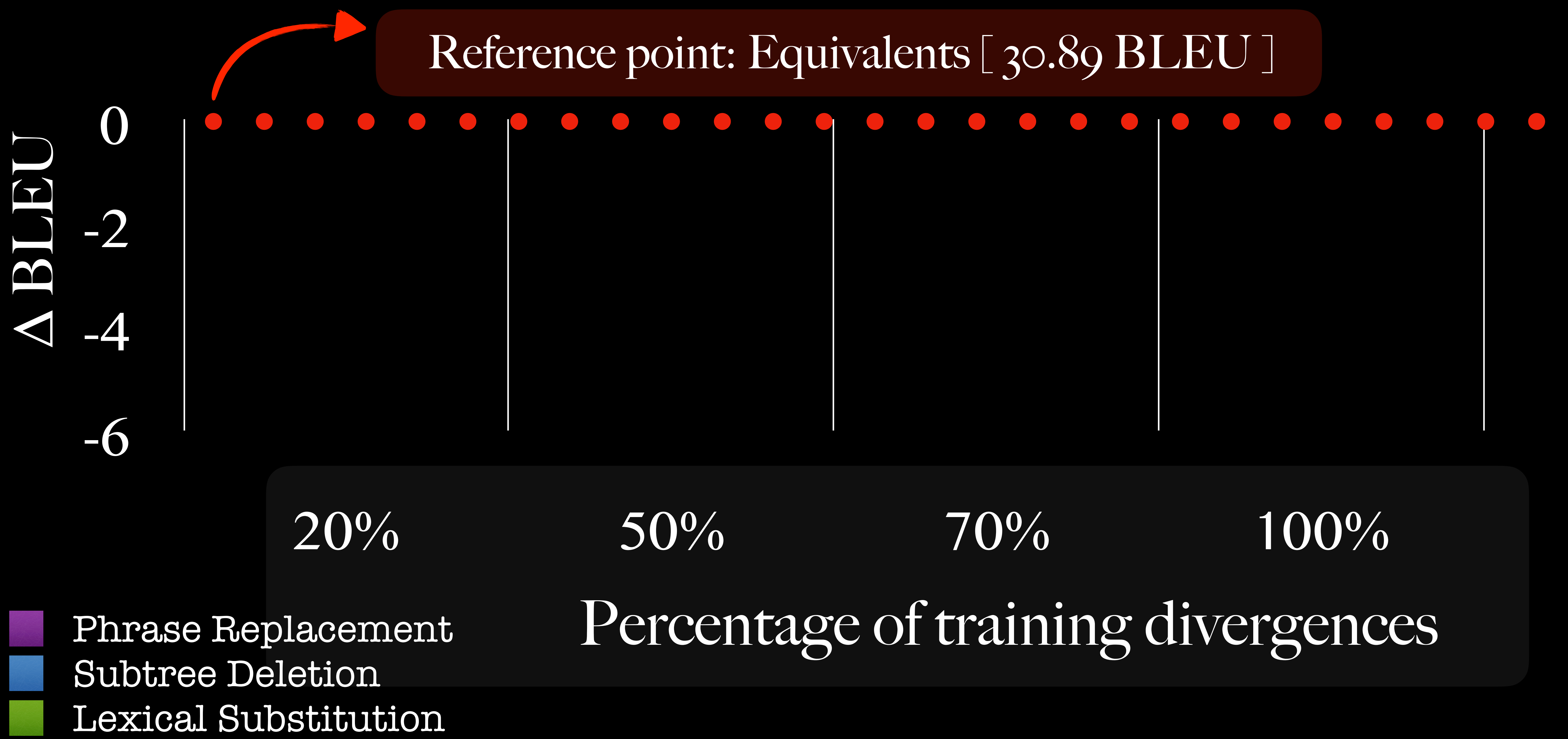
100%

100%

Fine-grained Divergences degrade BLEU when they overwhelm the training data

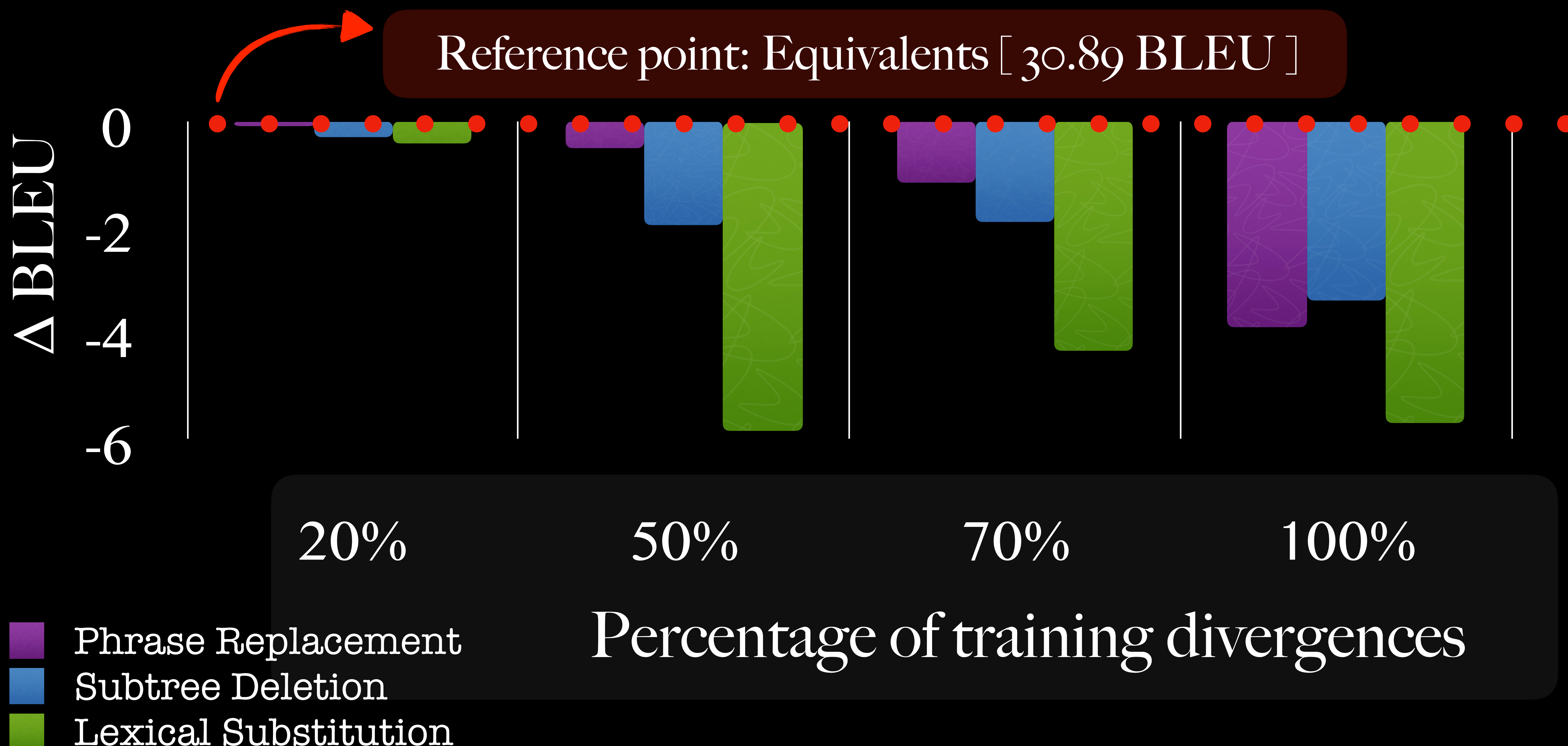


Fine-grained Divergences degrade BLEU when they overwhelm the training data



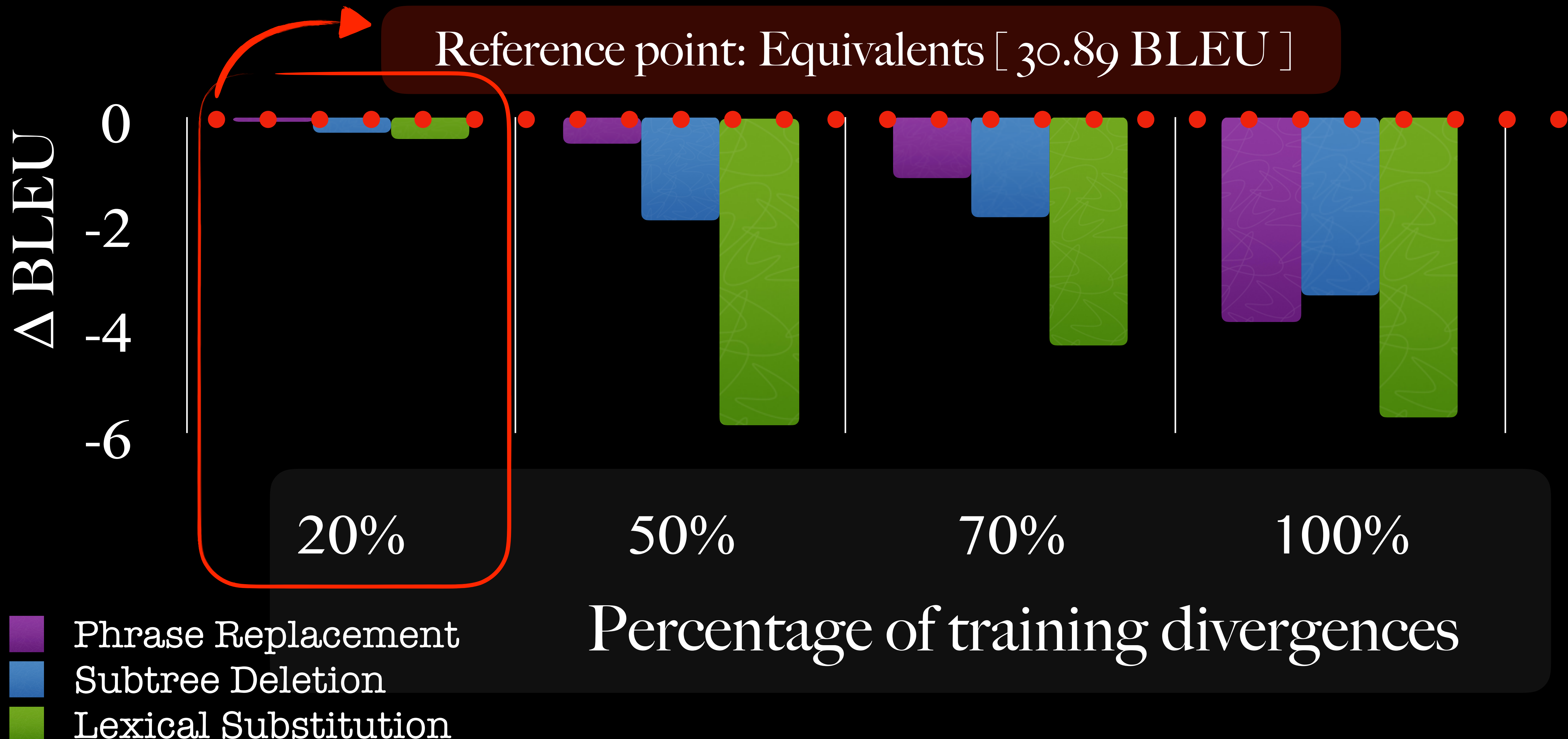
Beyond noise: Mitigating the impact of Fine-grained Divergences on Neural Machine Translation; Eleftheria Briakou and Marine Carpuat; ACL 2021

Fine-grained Divergences degrade BLEU when they overwhelm the training data



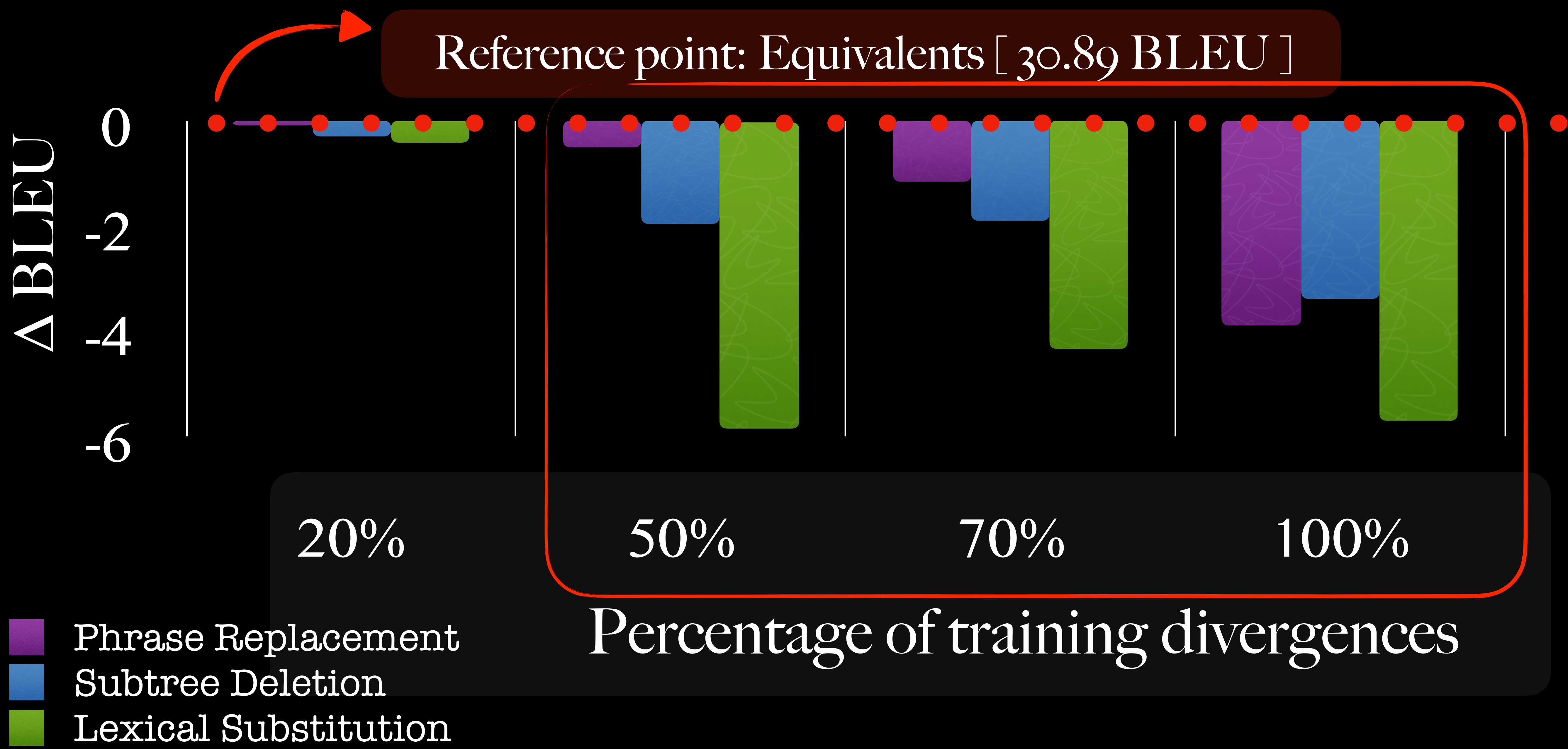
Beyond noise: Mitigating the impact of Fine-grained Divergences on Neural Machine Translation; Eleftheria Briakou and Marine Carpuat; ACL 2021

Fine-grained Divergences degrade BLEU when they overwhelm the training data



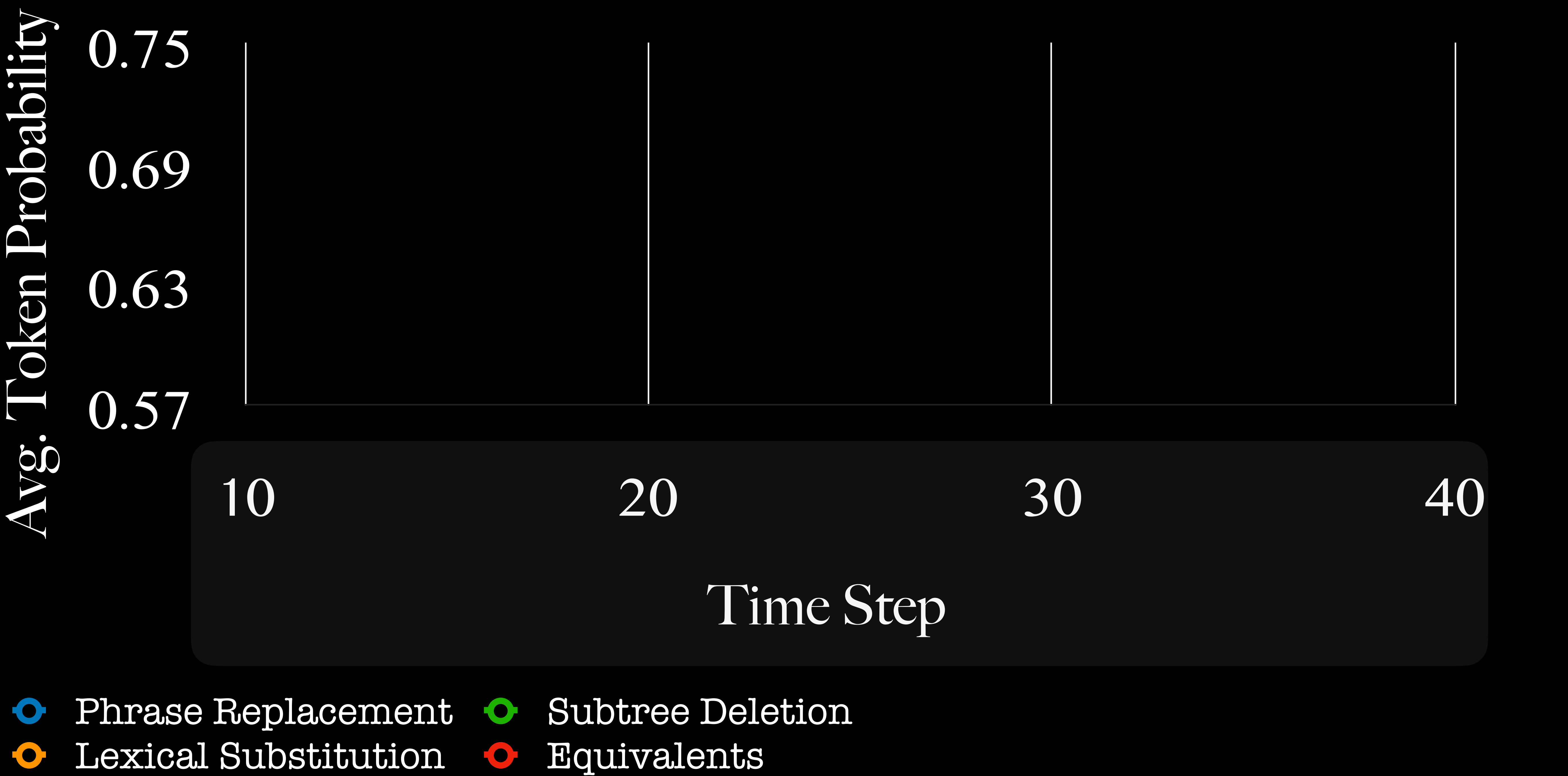
Beyond noise: Mitigating the impact of Fine-grained Divergences on Neural Machine Translation; Eleftheria Briakou and Marine Carpuat; ACL 2021

Fine-grained Divergences degrade BLEU when they overwhelm the training data



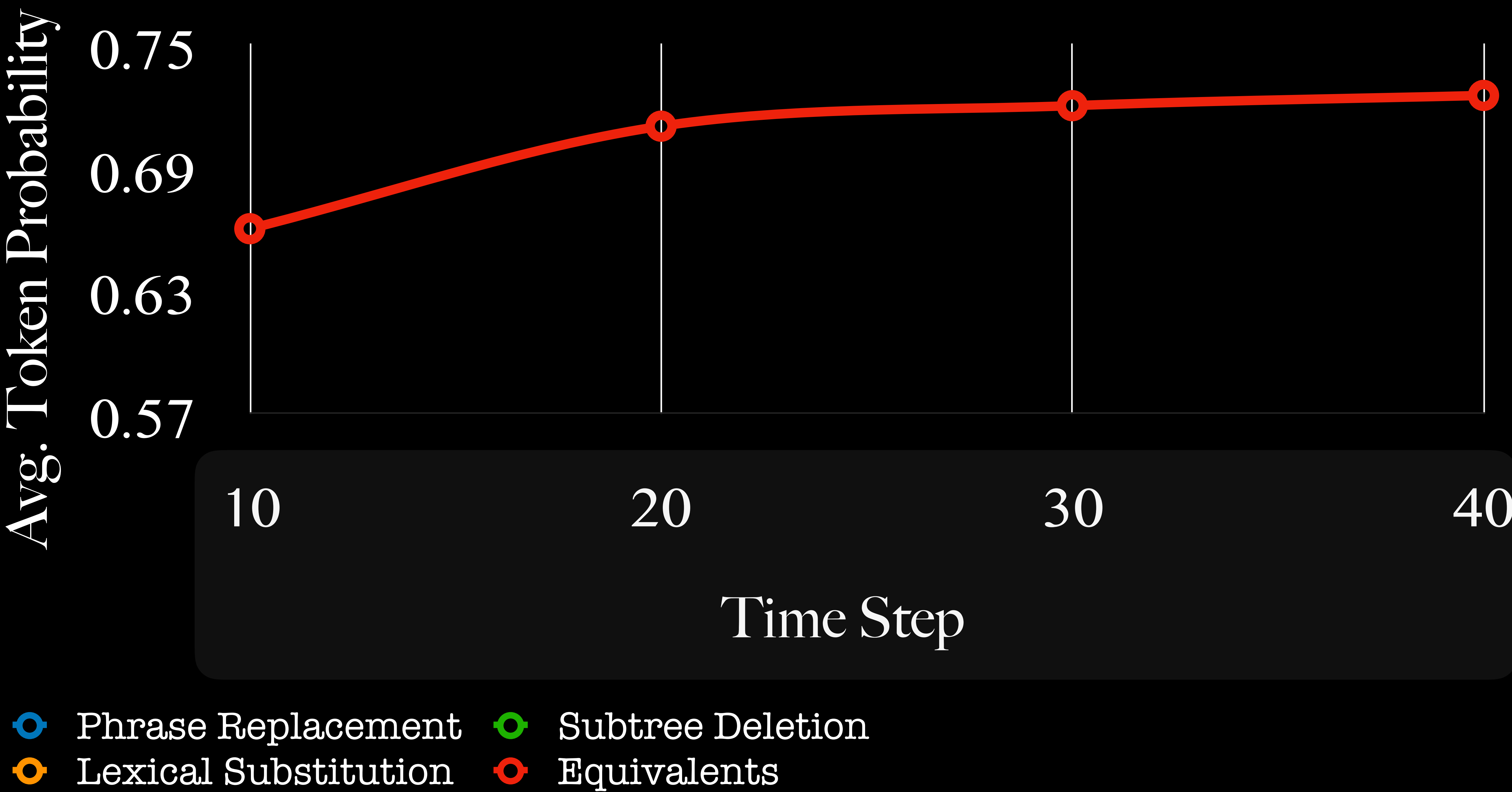
Beyond noise: Mitigating the impact of Fine-grained Divergences on Neural Machine Translation; Eleftheria Briakou and Marine Carpuat; ACL 2021

Fine-grained Divergences increase the uncertainty of token predictions



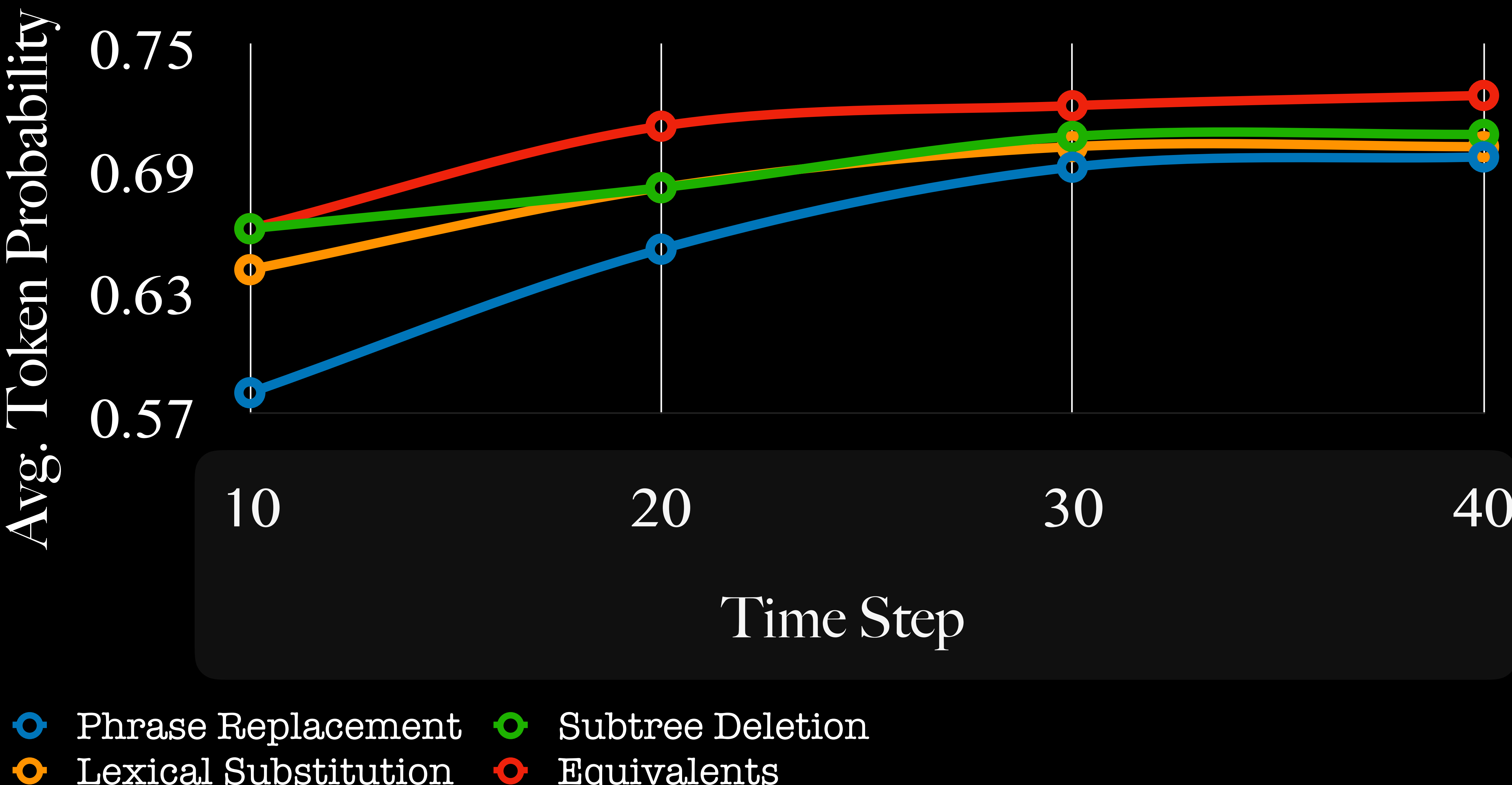
Beyond noise: Mitigating the impact of Fine-grained Divergences on Neural Machine Translation; Eleftheria Briakou and Marine Carpuat; ACL 2021

Fine-grained Divergences increase the uncertainty of token predictions



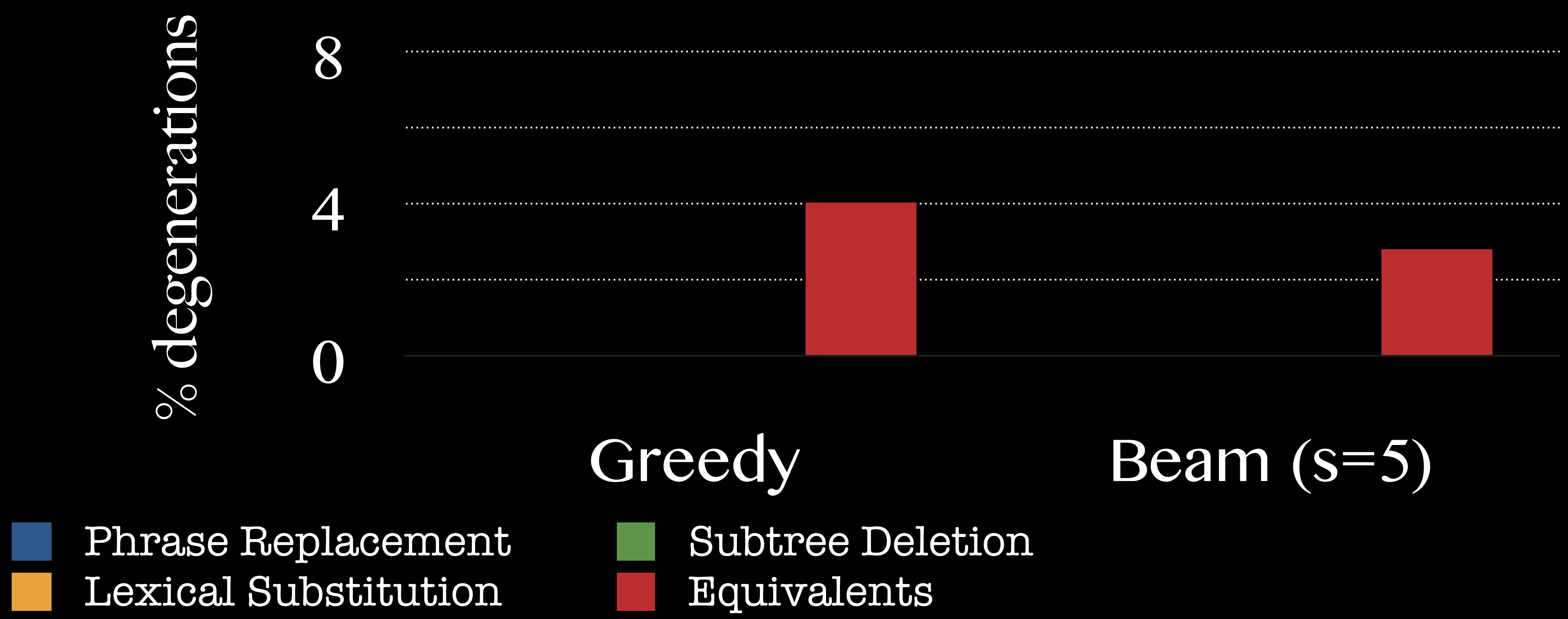
Beyond noise: Mitigating the impact of Fine-grained Divergences on Neural Machine Translation; Eleftheria Briakou and Marine Carpuat; ACL 2021

Fine-grained Divergences increase the uncertainty of token predictions



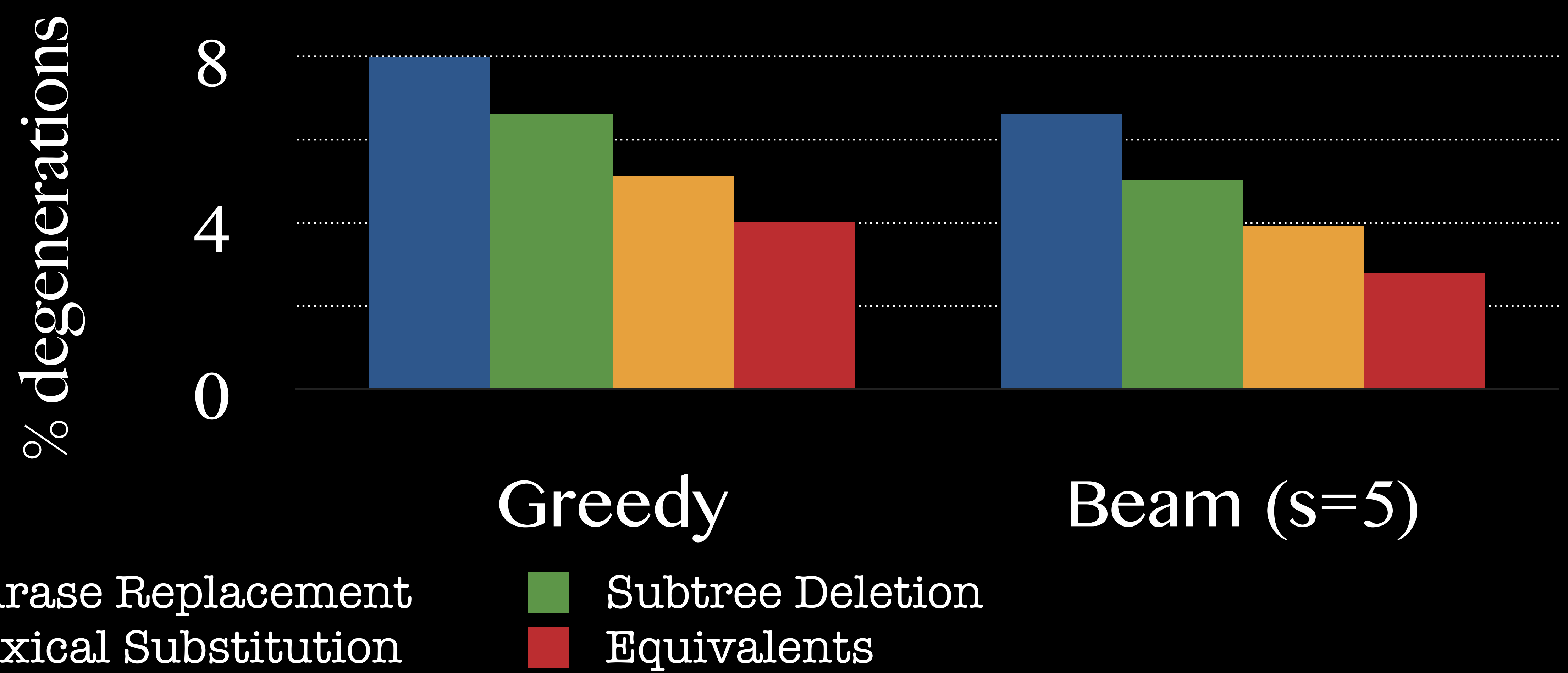
Beyond noise: Mitigating the impact of Fine-grained Divergences on Neural Machine Translation; Eleftheria Briakou and Marine Carpuat; ACL 2021

Fine-grained Divergences increase the frequency of degenerated hypotheses



Beyond noise: Mitigating the impact of Fine-grained Divergences on Neural Machine Translation; Eleftheria Briakou and Marine Carpuat; ACL 2021

Fine-grained Divergences increase the frequency of degenerated hypotheses



Beyond noise: Mitigating the impact of Fine-grained Divergences on Neural Machine Translation; Eleftheria Briakou and Marine Carpuat; ACL 2021

CHAPTER C REVISITED:
How do semantic divergences impact NMT?

- hurt translation quality
- more repetitive loops
- increase prediction uncertainty

CHAPTER D **BONUS**:

How can we mitigate the negative **impact Of semantic divergences NMT?**

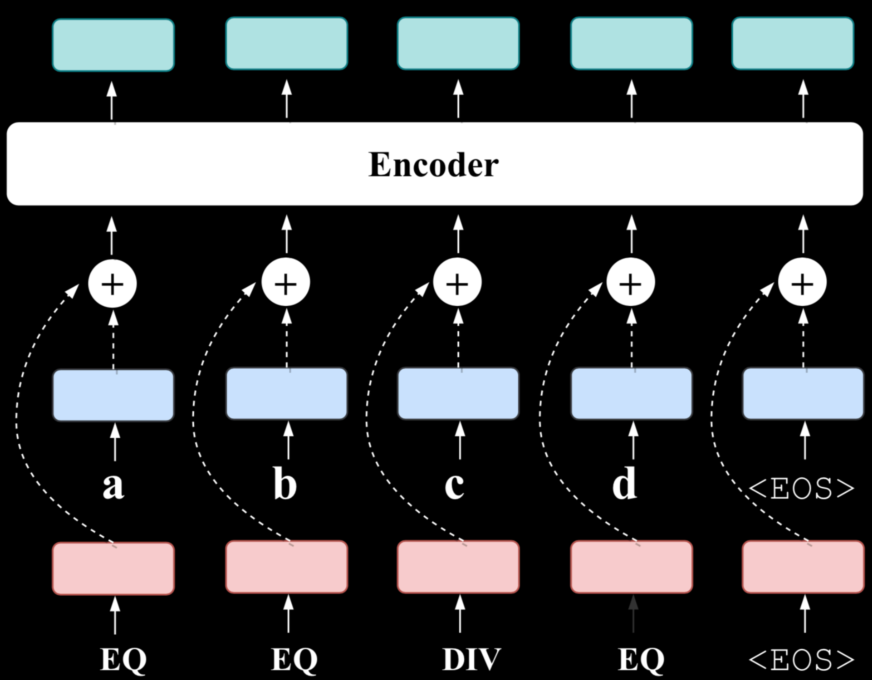
by encoding divergences as token factors

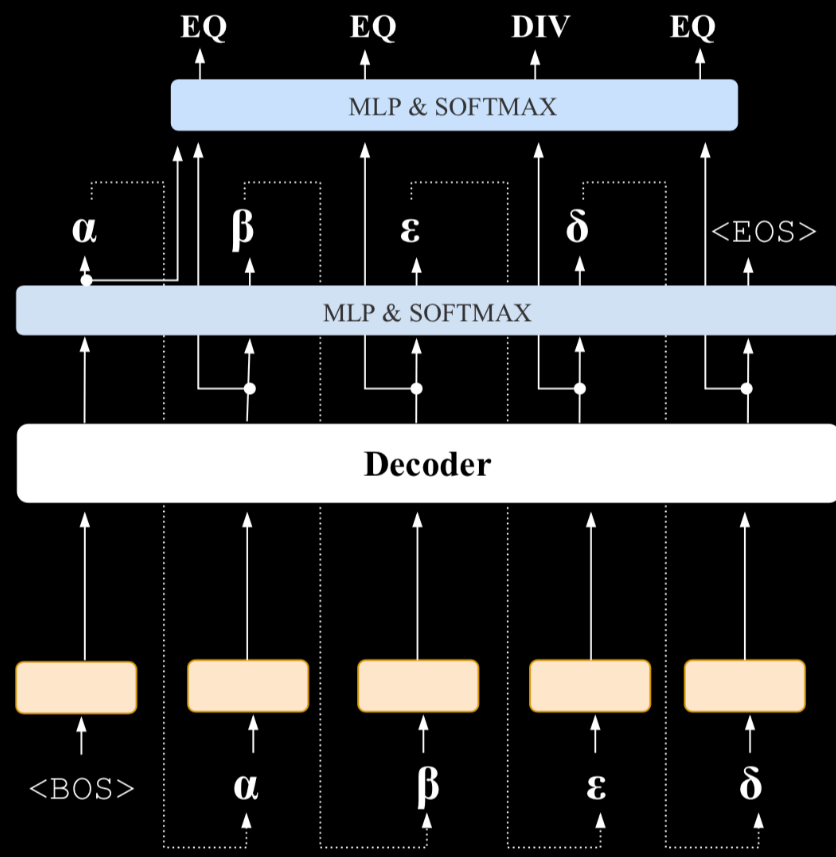
Beyond noise: Mitigating the impact of Fine-grained Divergences on Neural Machine Translation; Eleftheria Briakou and Marine Carpuat; ACL 2021

CHAPTER D **BONUS**:

How can we mitigate the negative **impact** **Of semantic divergences NMT?**

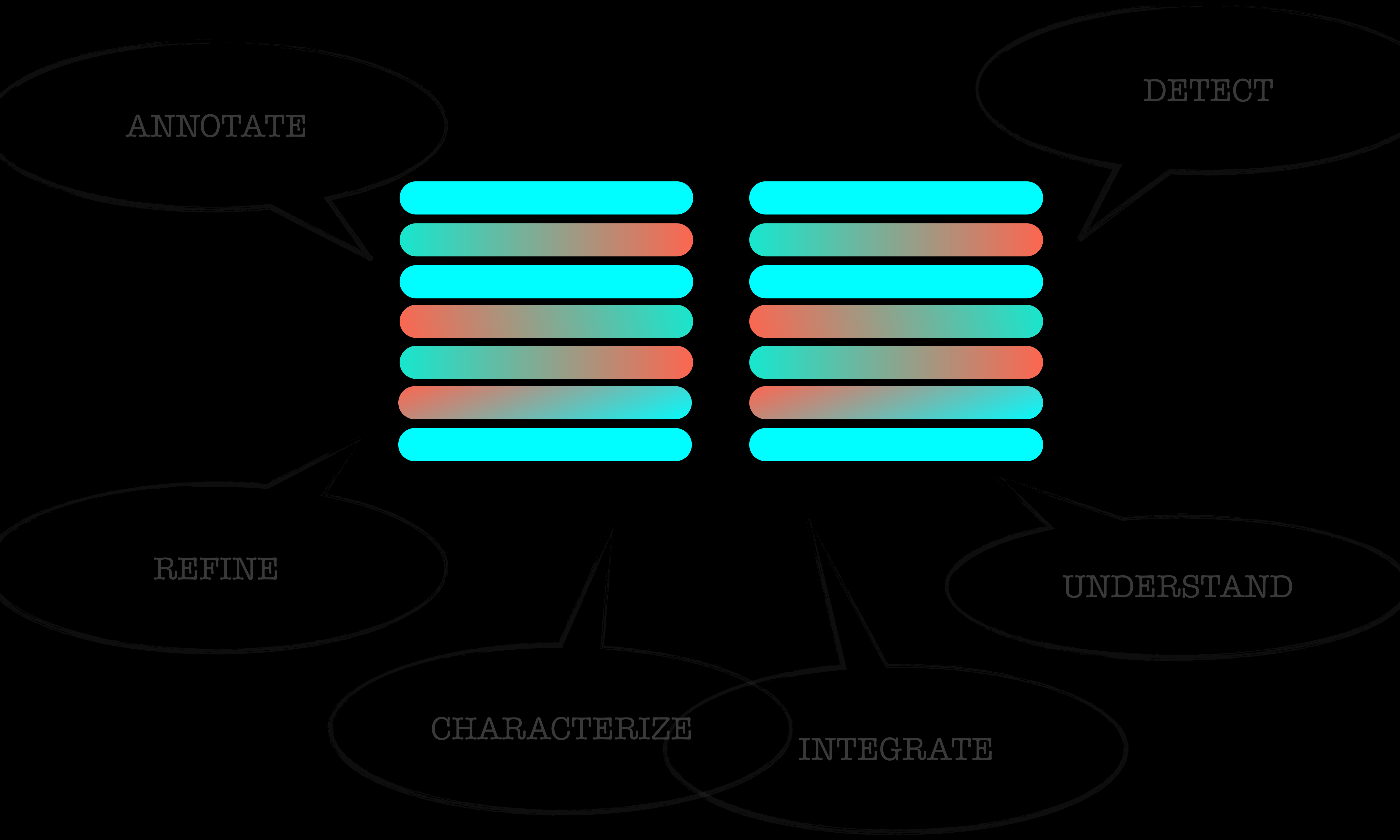
by encoding divergences as token factors



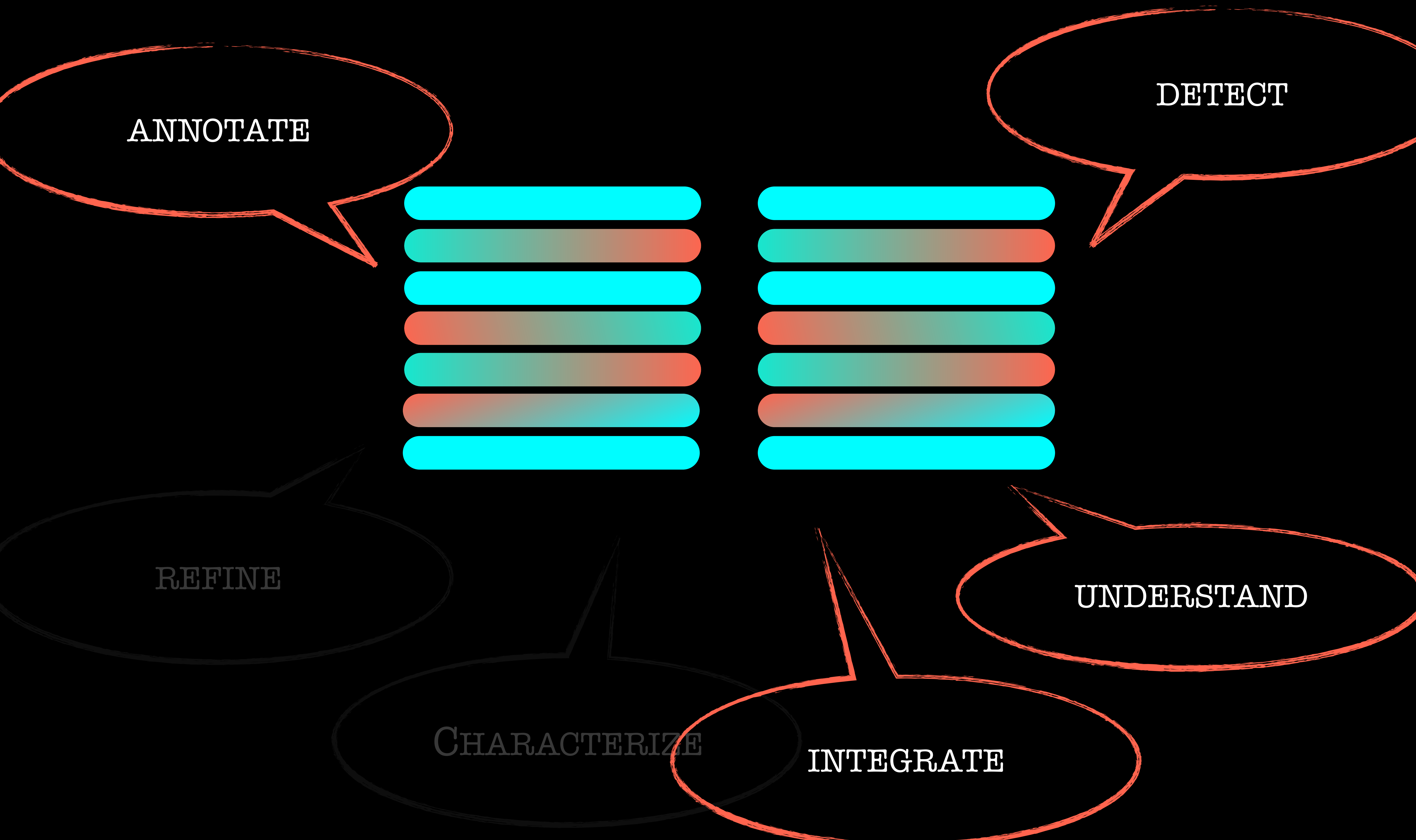


Beyond noise: Mitigating the impact of Fine-grained Divergences on Neural Machine Translation; Eleftheria Briakou and Marine Carpuat; ACL 2021

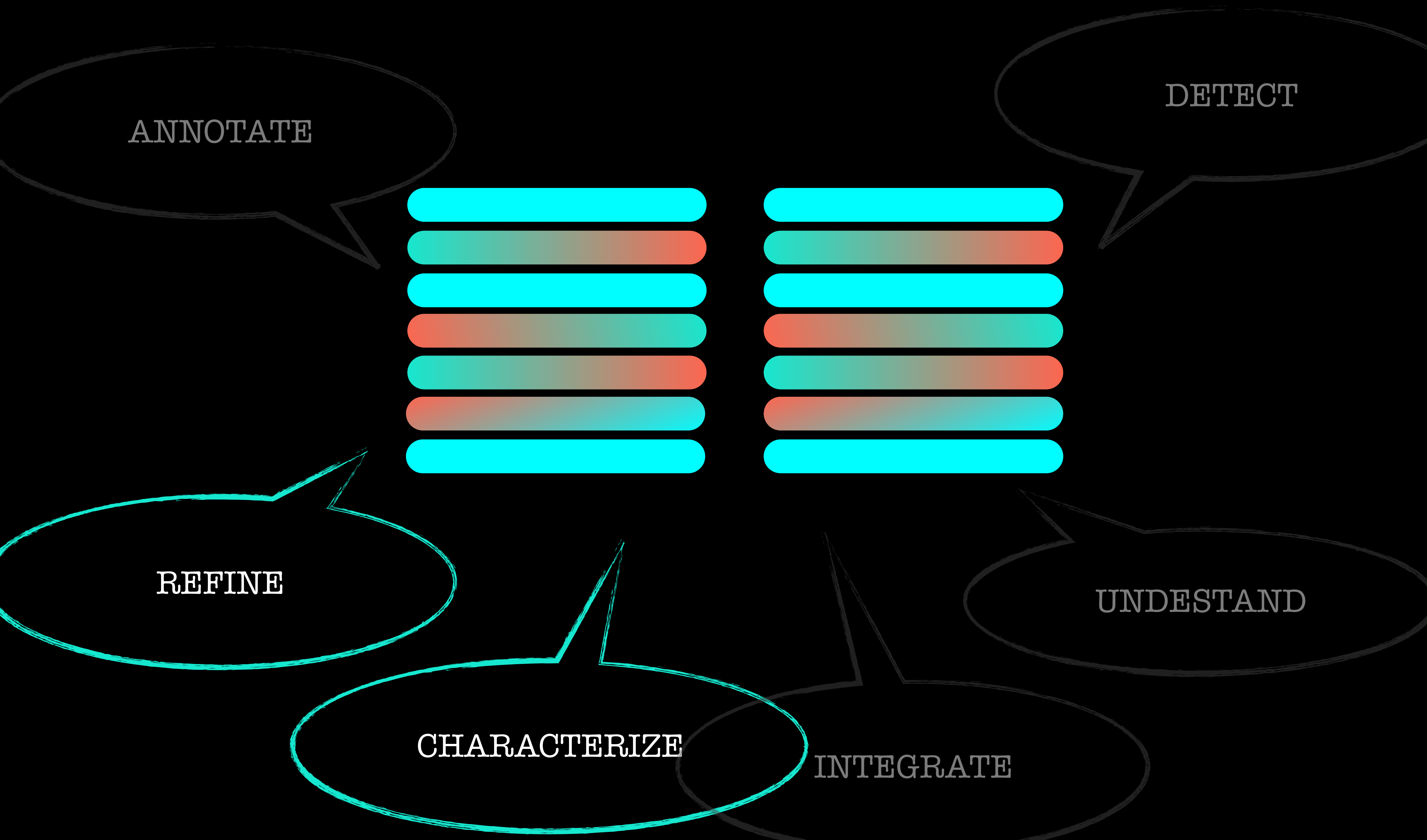
Big Picture Revisited



Big Picture Revisited



Big Picture Revisited



Questions?

ebriakou@cs.umd.edu