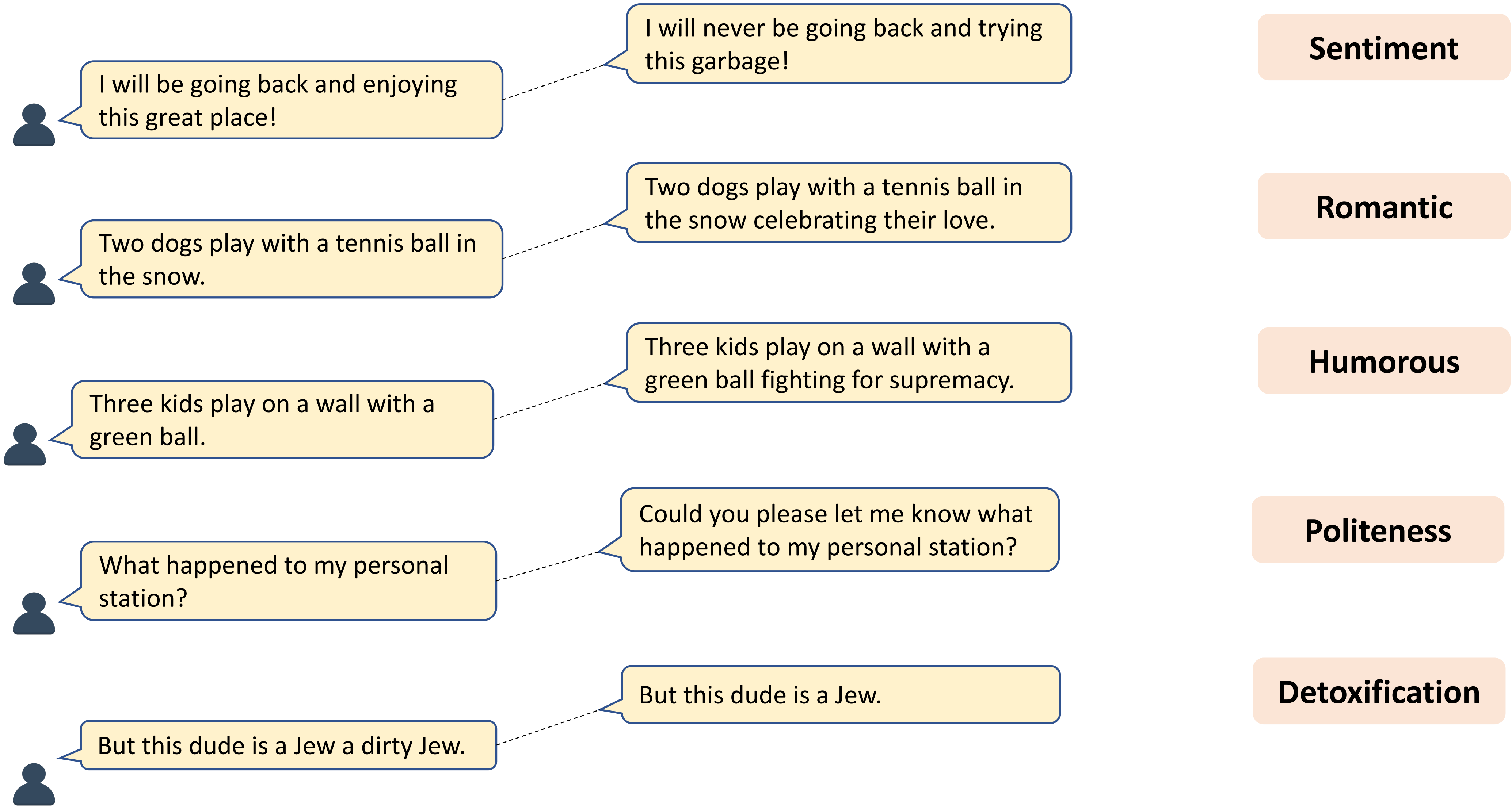
“Slow Service”↛ “Great Food”: Enhancing Content Preservation in Unsupervised Text Style Transfer

Wanzheng Zhu (wz6@illinois.edu), **Suma Bhat**

University of Illinois, at Urbana-Champaign

Text Style Transfer



I will be going back and enjoying this great place!

I will never be going back and trying this garbage!

Sentiment

Two dogs play with a tennis ball in the snow.

Two dogs play with a tennis ball in the snow celebrating their love.

Romantic

Three kids play on a wall with a green ball.

Three kids play on a wall with a green ball fighting for supremacy.

Humorous

What happened to my personal station?

Could you please let me know what happened to my personal station?

Politeness

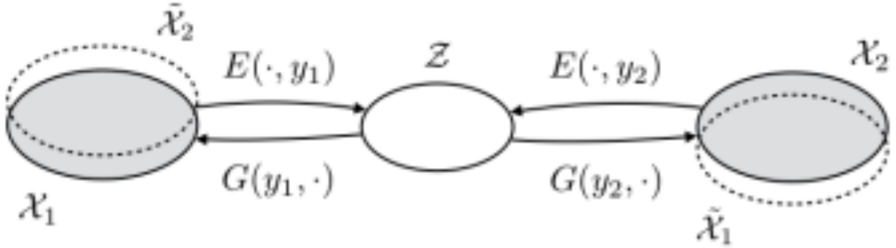
But this dude is a Jew a dirty Jew.

But this dude is a Jew.

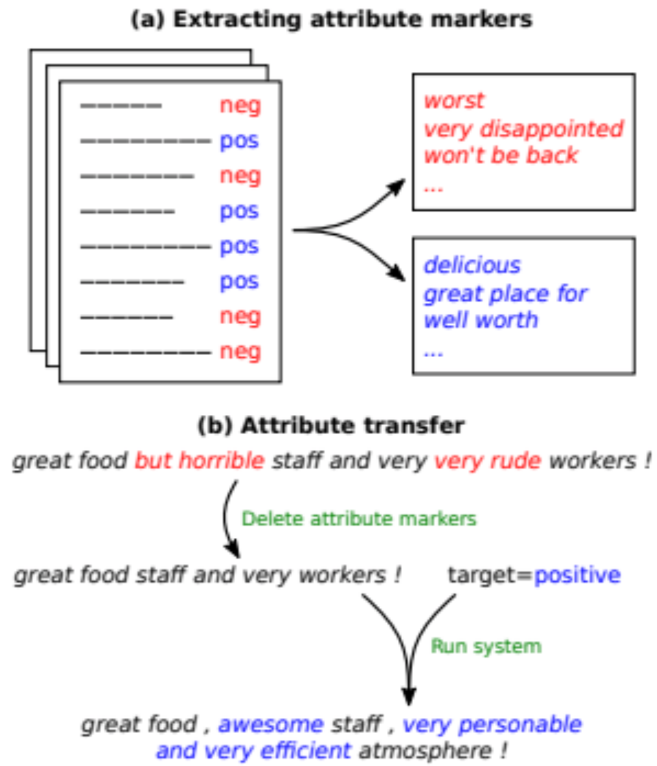
Detoxification

Existing Work

- Latent Representation
- Prototype-editing



[1] Shen et al. 2017. Style Transfer from Non-Parallel Text by Cross-Alignment



(a) Extracting attribute markers
neg
pos
neg
pos
pos
pos
neg
neg
worst
very disappointed
won't be back
...
delicious
great place for
well worth
...
(b) Attribute transfer
great food but horrible staff and very very rude workers !
Delete attribute markers
great food staff and very workers !
target=positive
Run system
great food , awesome staff , very personable and very efficient atmosphere !

[2] Li et al. 2018. Delete, Retrieve, Generate: A Simple Approach to Sentiment and Style Transfer

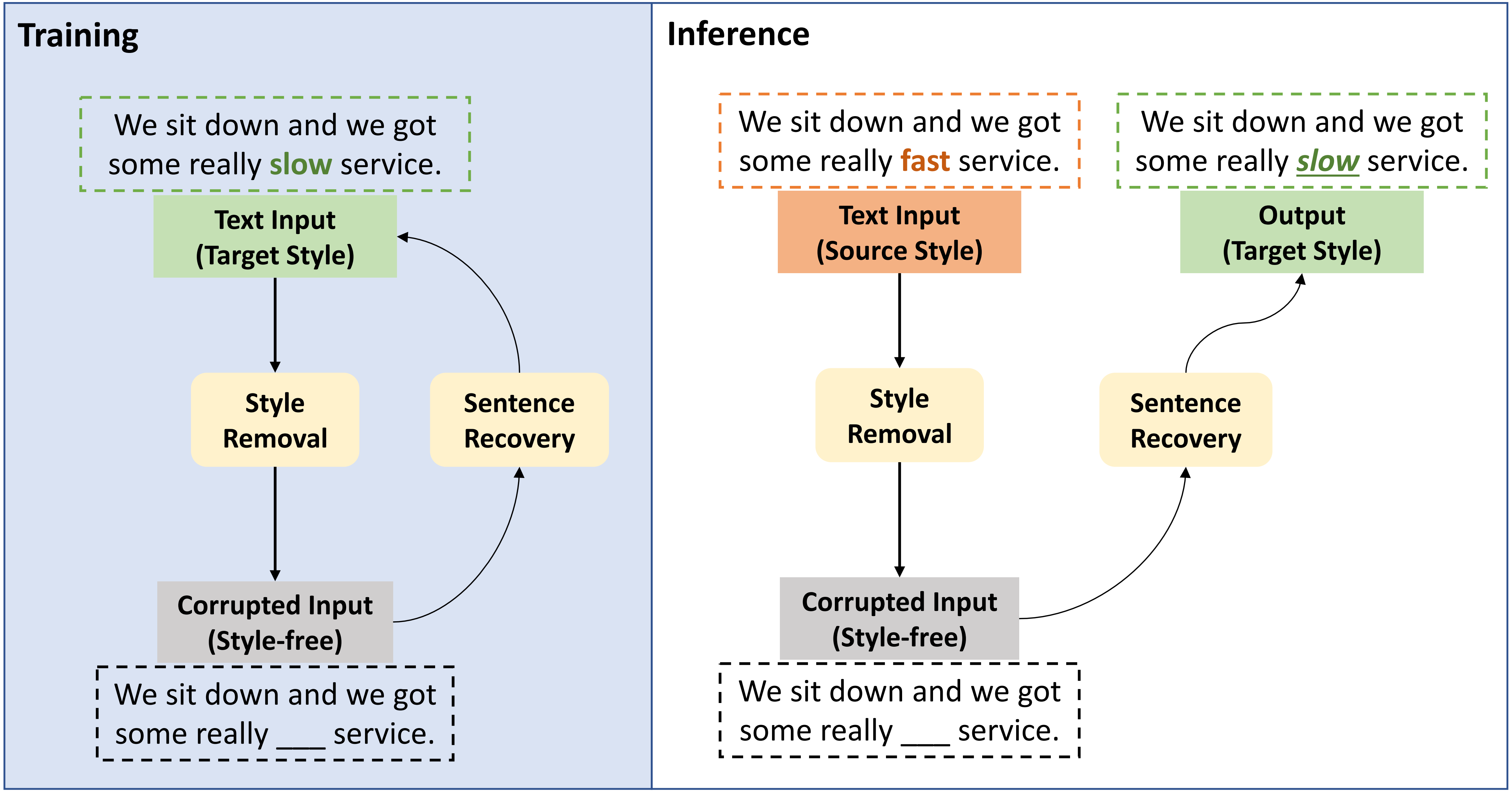
Challenge: Content Preservation

- Many content-related tokens are masked.
  - ✓ BERT-based keyword extraction model with syntactic information.
- Irrelevant words associated with the target style are infilled.
  - ✓ Training a T5 model on a pseudo-parallel dataset.

we sit down and we got some really slow and lazy **service**.

we sit down and we got some really good **food** and loved it.

Model - Overview



Training

We sit down and we got some really **slow** service.

Text Input (Target Style)

Style Removal

Corrupted Input (Style-free)

We sit down and we got some really ___ service.

Sentence Recovery

Inference

We sit down and we got some really **fast** service.

Text Input (Source Style)

Style Removal

Corrupted Input (Style-free)

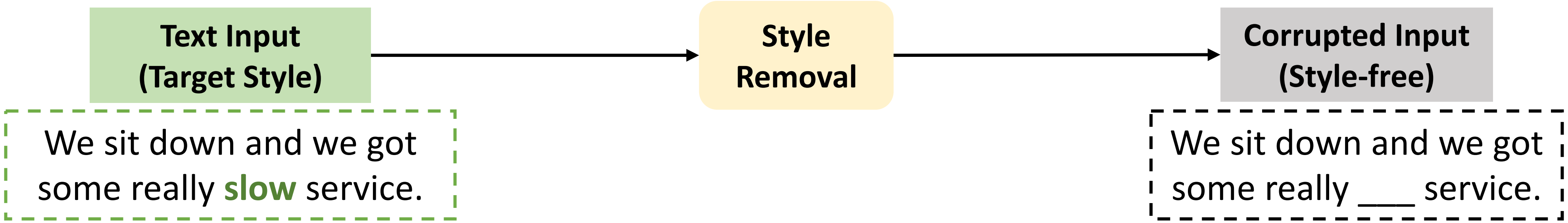
We sit down and we got some really ___ service.

Sentence Recovery

Output (Target Style)

We sit down and we got some really ***slow*** service.

Model – Style Removal



Text Input (Target Style)

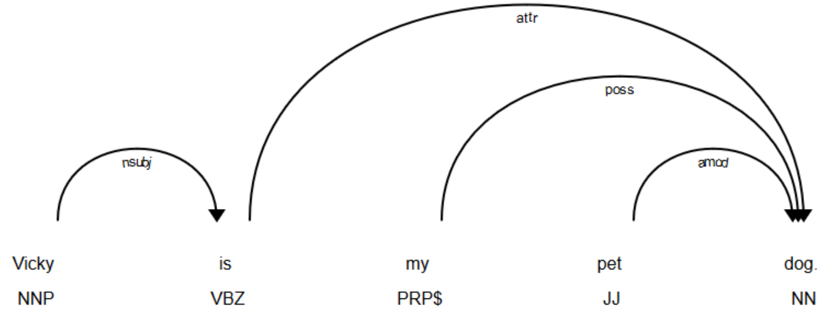
We sit down and we got some really **slow** service.

Style Removal

Corrupted Input (Style-free)

We sit down and we got some really ___ service.

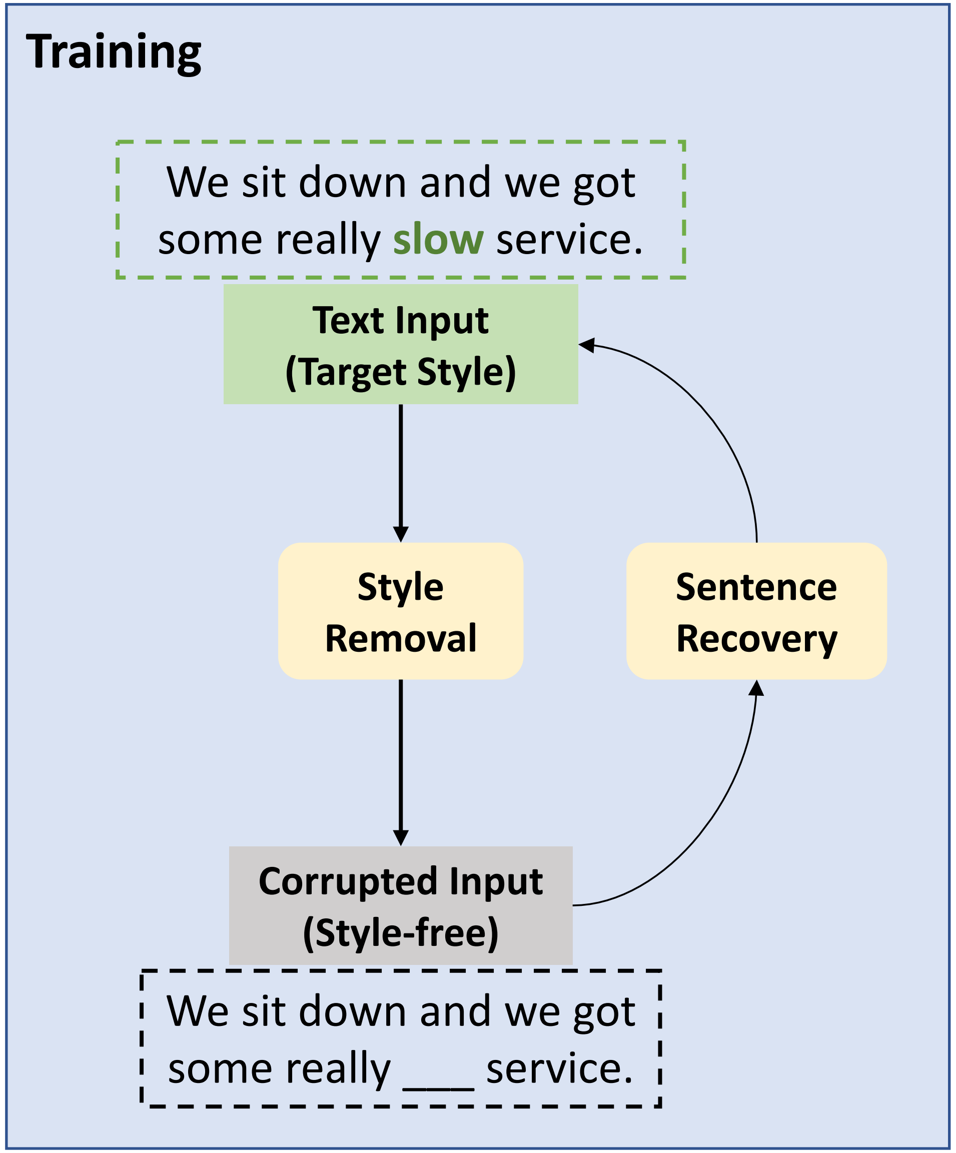
- **Keyword Extraction**
  - BERT Embedding $(e_{t1}, e_{t2}, e_{t3}, \dots, e_{tn}, e_s)$
    - Ranking: $\mathrm{r_{ti}} = \alpha \cdot \cos(e_{ti}, e_s)$
  - Dependency Parsing
  - Ranking: $\mathrm{r_{ti}} = \alpha \cdot \cos(e_{ti}, e_s) + \beta \cdot d_i + \gamma \cdot o_i$
- **Attention**



In dependency parsing, the head word of a constituent was the central organizing word of a larger constituent [1].

[1] Speech & Language Processing. Jurafsky, 2000

Model – Sentence Recovery



Training

We sit down and we got some really **slow** service.

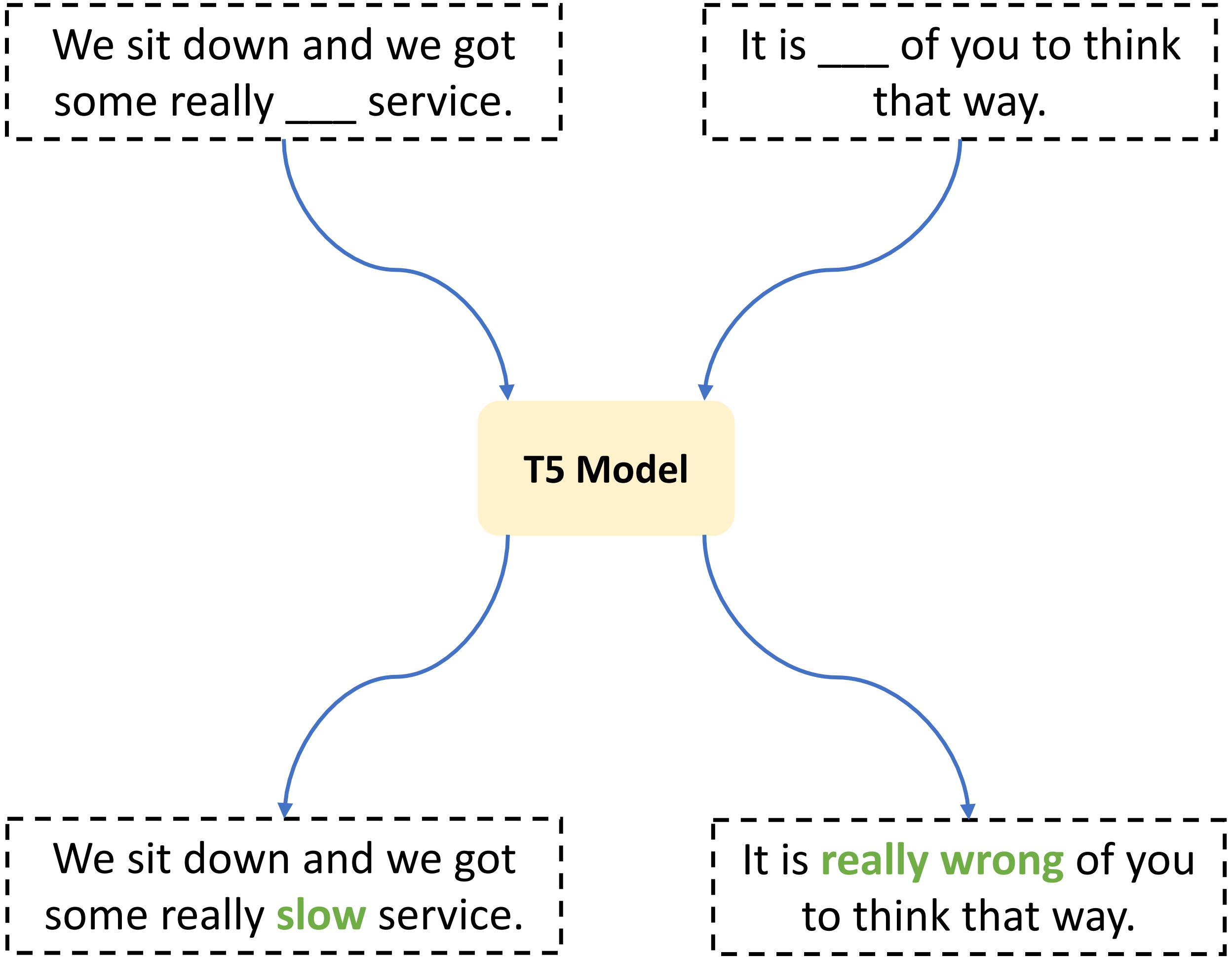
Text Input (Target Style)

Style Removal

Sentence Recovery

Corrupted Input (Style-free)

We sit down and we got some really ___ service.



We sit down and we got some really ___ service.

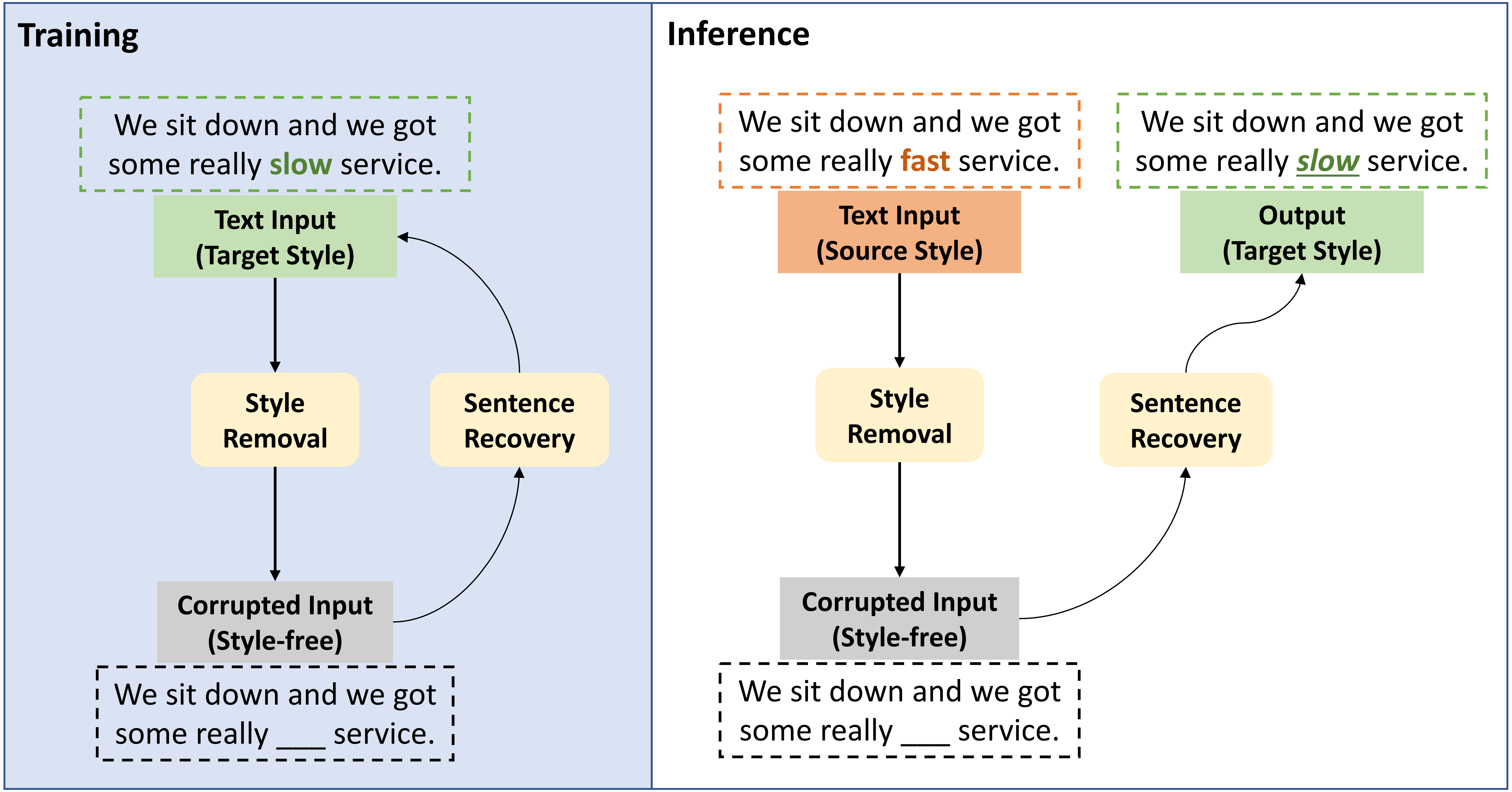
It is ___ of you to think that way.

T5 Model

We sit down and we got some really **slow** service.

It is **really wrong** of you to think that way.

Model - Overview



Training

We sit down and we got some really **slow** service.

Text Input (Target Style)

Style Removal

Sentence Recovery

Corrupted Input (Style-free)

We sit down and we got some really ___ service.

Inference

We sit down and we got some really **fast** service.

Text Input (Source Style)

Style Removal

Corrupted Input (Style-free)

We sit down and we got some really ___ service.

Sentence Recovery

Output (Target Style)

We sit down and we got some really ***slow*** service.

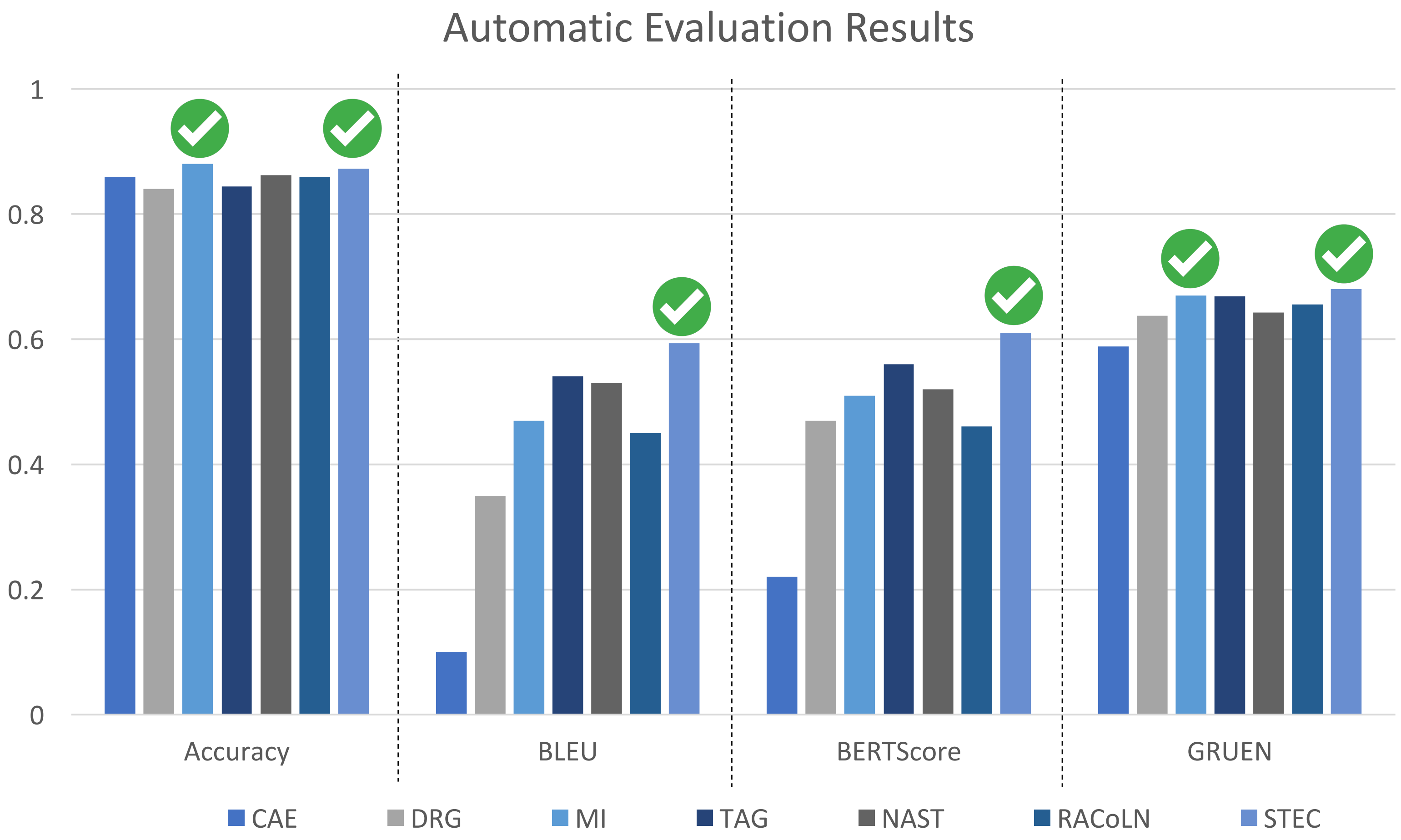
Results - Automatic

Dataset

- Yelp (Li et al. 2018)
- Amazon (Li et al. 2018)
- Captions (Gan et al. 2017)
- Politeness (Madaan et al. 2020)
- Detoxification (Dale et al. 2021)

Evaluation Metric

- Transfer Effectiveness
  - Accuracy
- Content Preservation
  - BLEU
  - BERTScore
- Language Quality
  - GRUEN



Results are averaged across five datasets and are scaled for better presentation.

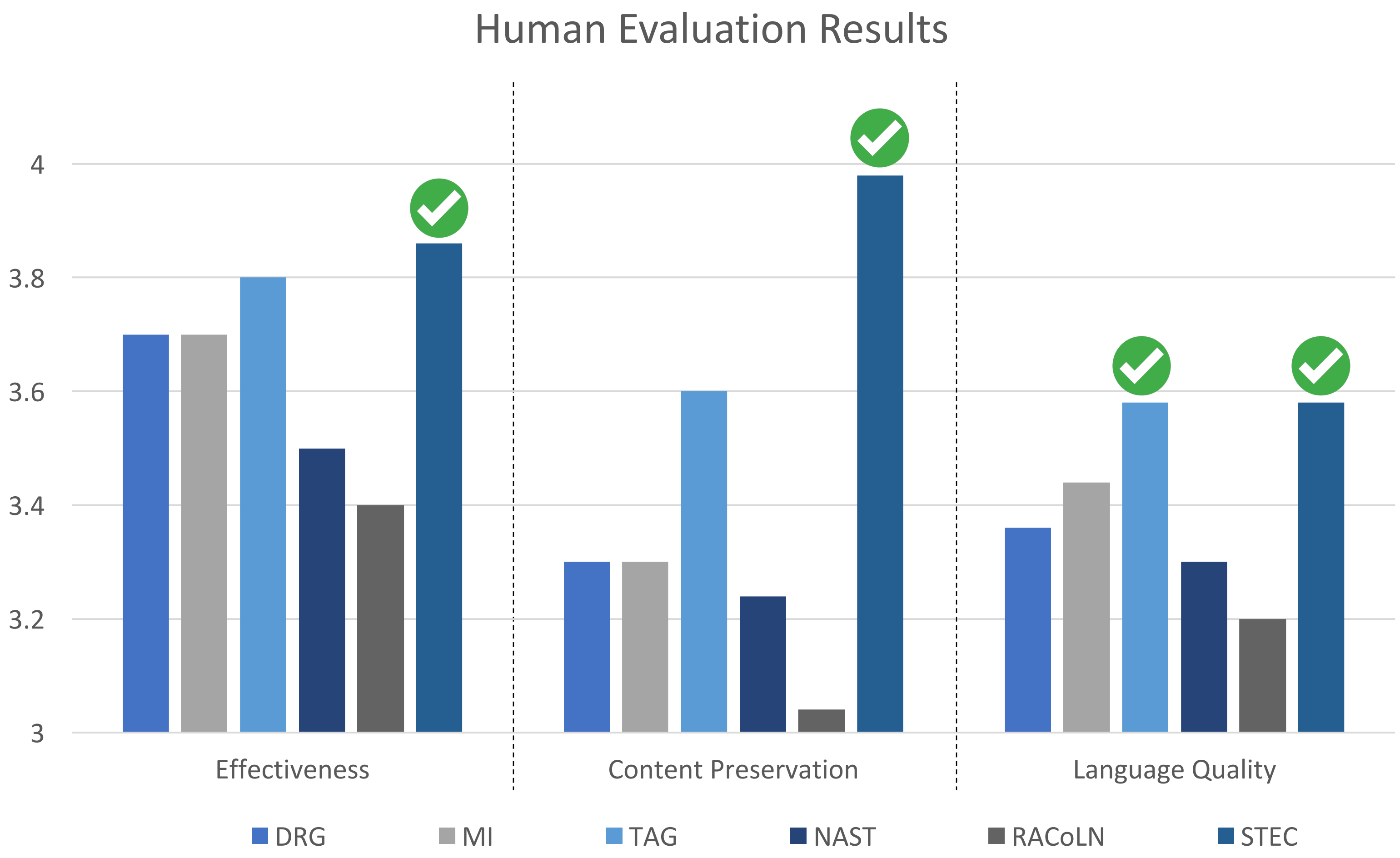
Results - Human

Dataset

- Detoxification (Dale et al. 2021)
- Yelp (Li et al. 2018)
- Amazon (Li et al. 2018)
- Captions (Gan et al. 2017)
- Politeness (Madaan et al. 2020)

Evaluation Metric

- Transfer Effectiveness
- Content Preservation
- Language Quality



Results are averaged across five datasets and are scaled for better presentation.

Case Study

Negative -> Positive	Input	We sit down and we got some really slow and lazy **service**.
	TAG	We sit down and we got some really good **food** and loved it.
	Our model	We sit down and we got some really great service.

Positive -> Negative	Input	The taste is awesome.
	TAG	The taste is not good and **the service is slow**.
	Our model	The taste is really bad.

Factual -> Humorous	Input	The group of **hikers** is resting in front of a **mountain**.
	TAG	The group of **people** is resting in front of a **cliff**.
	Our model	The group of hikers is being pulled in front of a mountain.

Toxic -> Civil	Input	Suggesting that people change their **commute times** is f*cking stupid.
	TAG	Suggesting that people change their **schedules** are not desired.
	Our model	Suggesting that people change their commute times is useless.

"Slow Service”↛ "Great Food": Enhancing Content Preservation in Unsupervised Text Style Transfer

Wanzheng Zhu (wz6@illinois.edu), **Suma Bhat**

University of Illinois, at Urbana-Champaign