Delete, Retrieve, Generate: A Simple Approach to Sentiment and Style Transfer

(Me)

Juncen Li[1], Robin Jia[2], He He[2], and Percy Liang[2]
[1]Tencent [2]Stanford University

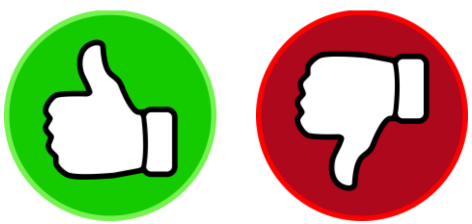
Text Attribute Transfer

Original Sentence: *"The gumbo was bland."*

Original Attribute: *negative* sentiment

Target Attribute: *positive* sentiment

New Sentence: *"The gumbo was tasty."*



Attribute Transfer

Content Preservation

Grammaticality

No parallel data

English		French
The blue house is old.	→	La maison bleue est vieille.
The music was loud.	→	La musique était forte
The boat left.	→	Le bateau est parti
…		…

Negative

The gumbo was bad

Very rude staff

Poorly lit

...

Positive

The beignets were tasty

I like their jambalaya

Very affordable

...

Delete, Retrieve, Generate

I ~~hated~~ the gumbo

Delete

love it

Retrieve

I love the gumbo

Generate

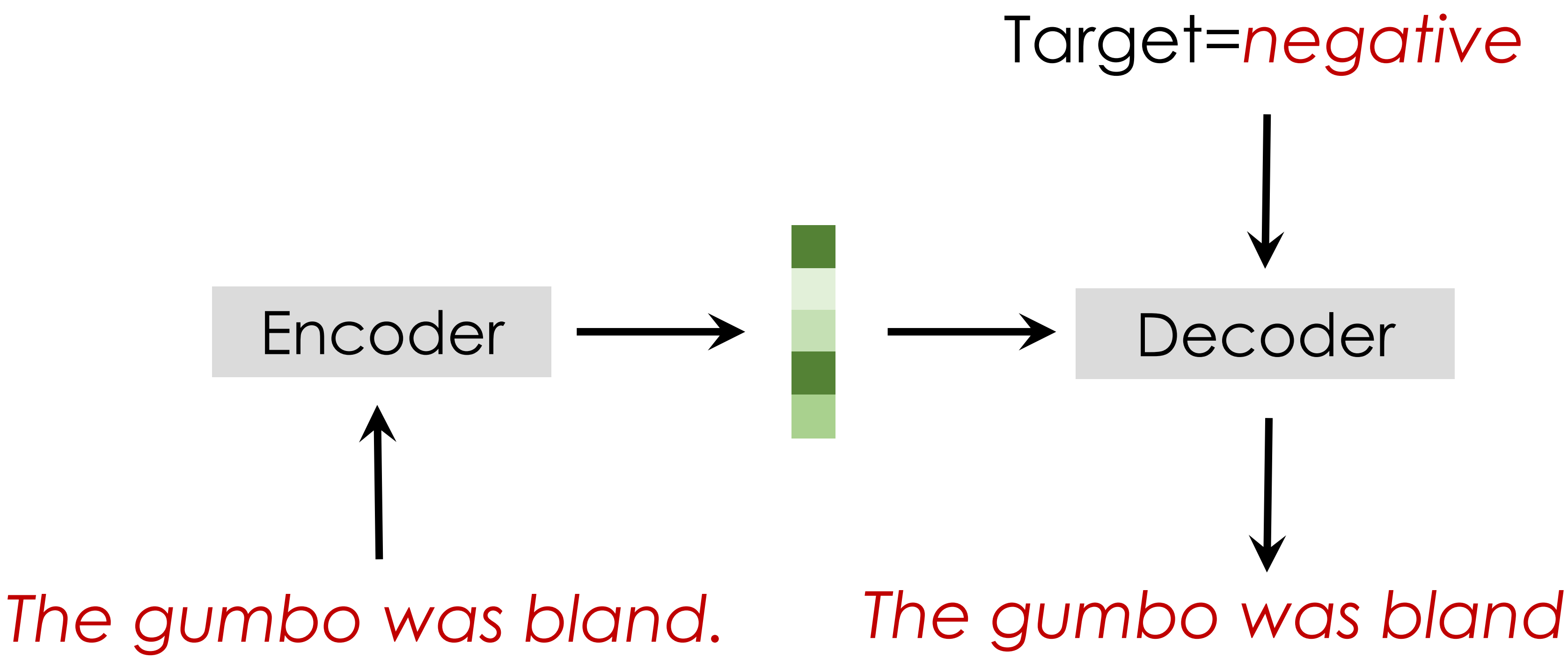
Outline

- Prior work with adversarial methods
- Simple baselines
- Simple neural methods

Outline

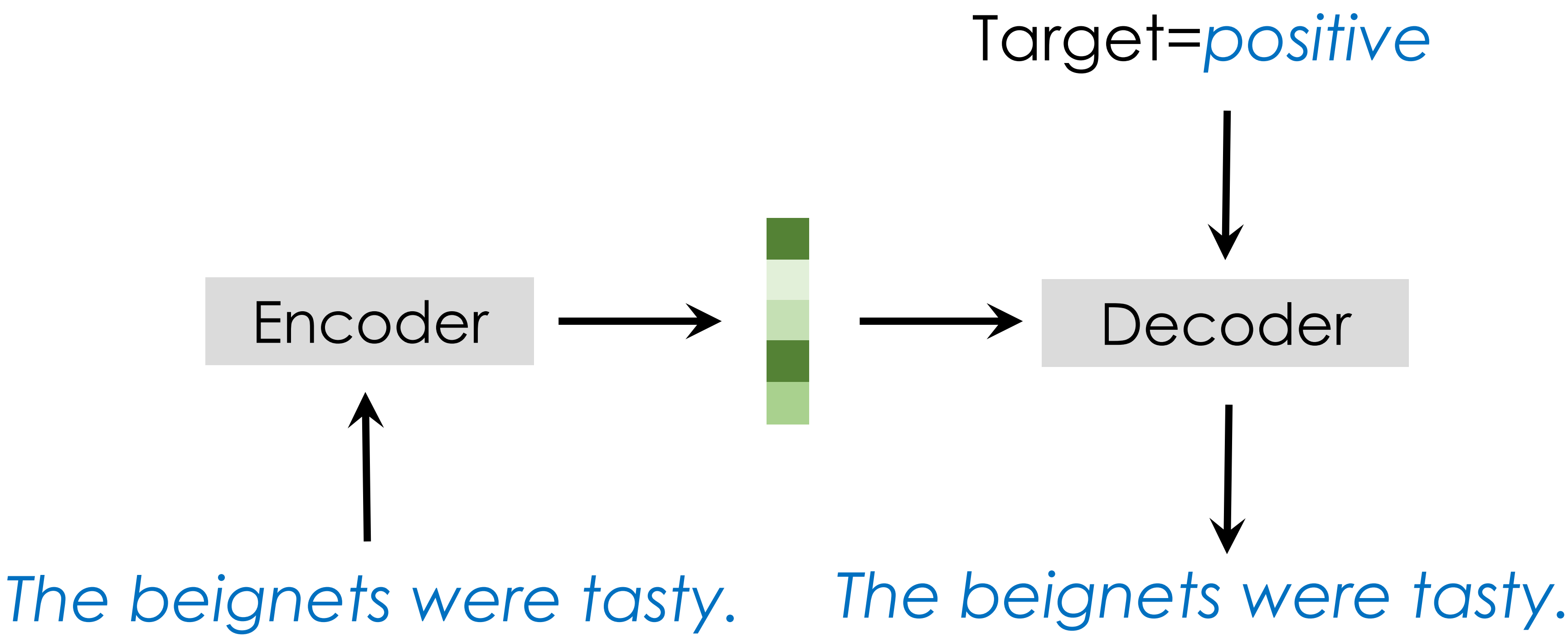
- Prior work with adversarial methods
- Simple baselines
- Simple neural methods

Basic auto-encoder



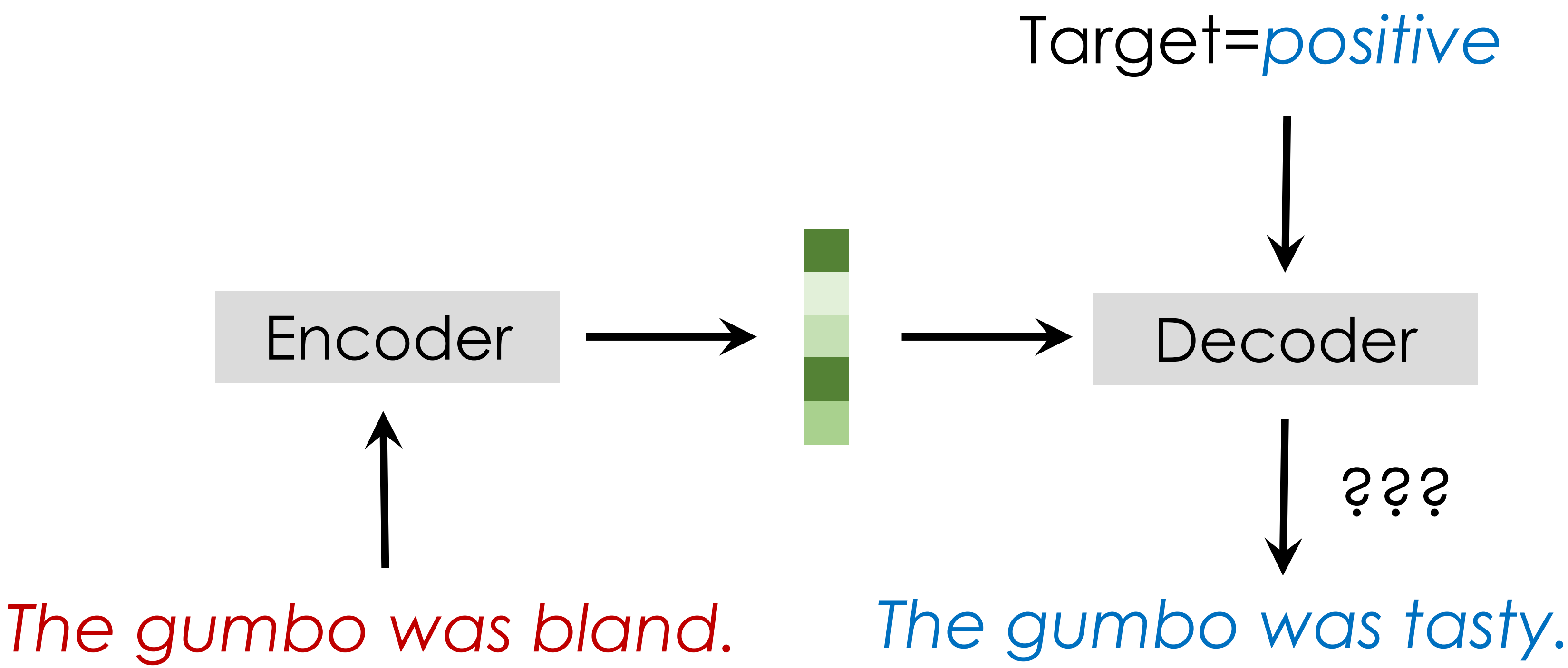
Shen et al. (2017); Fu et al. (2018)

Basic auto-encoder



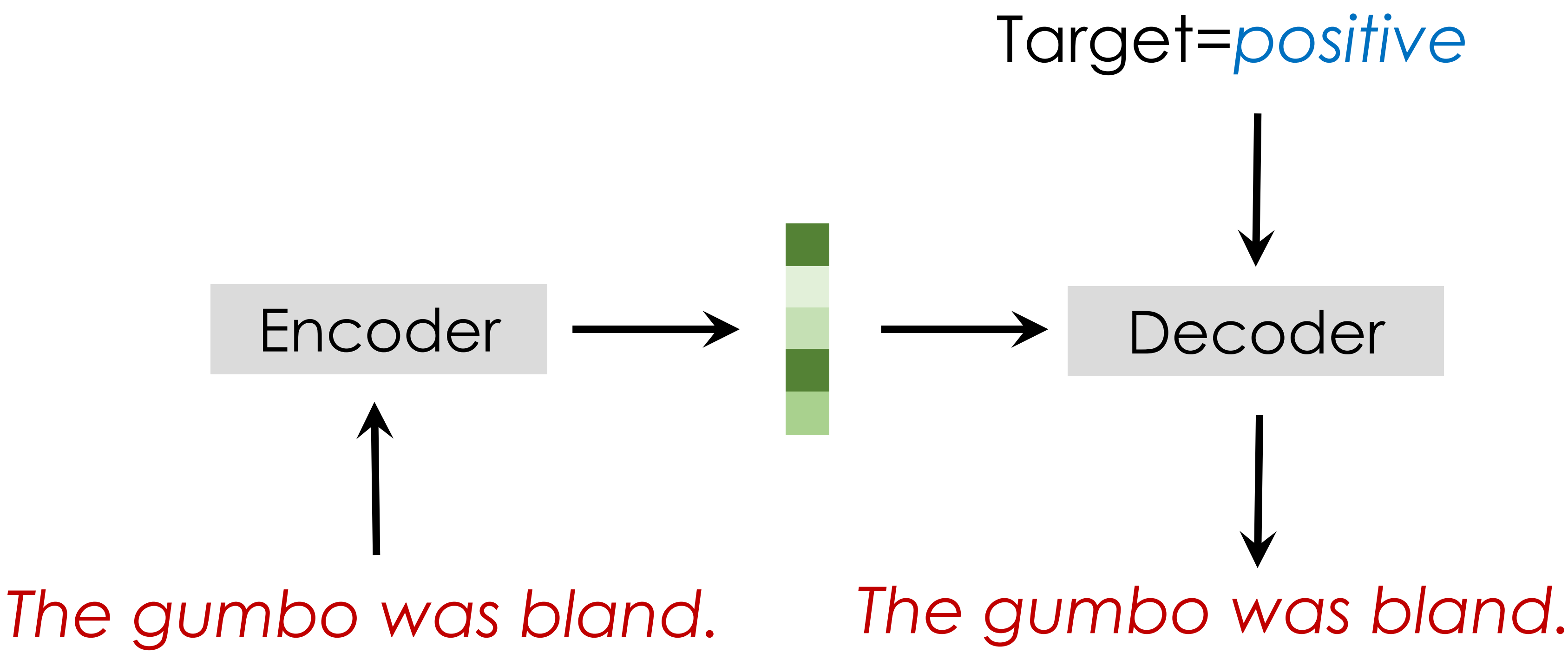
Shen et al. (2017); Fu et al. (2018)

Basic auto-encoder



Shen et al. (2017); Fu et al. (2018)

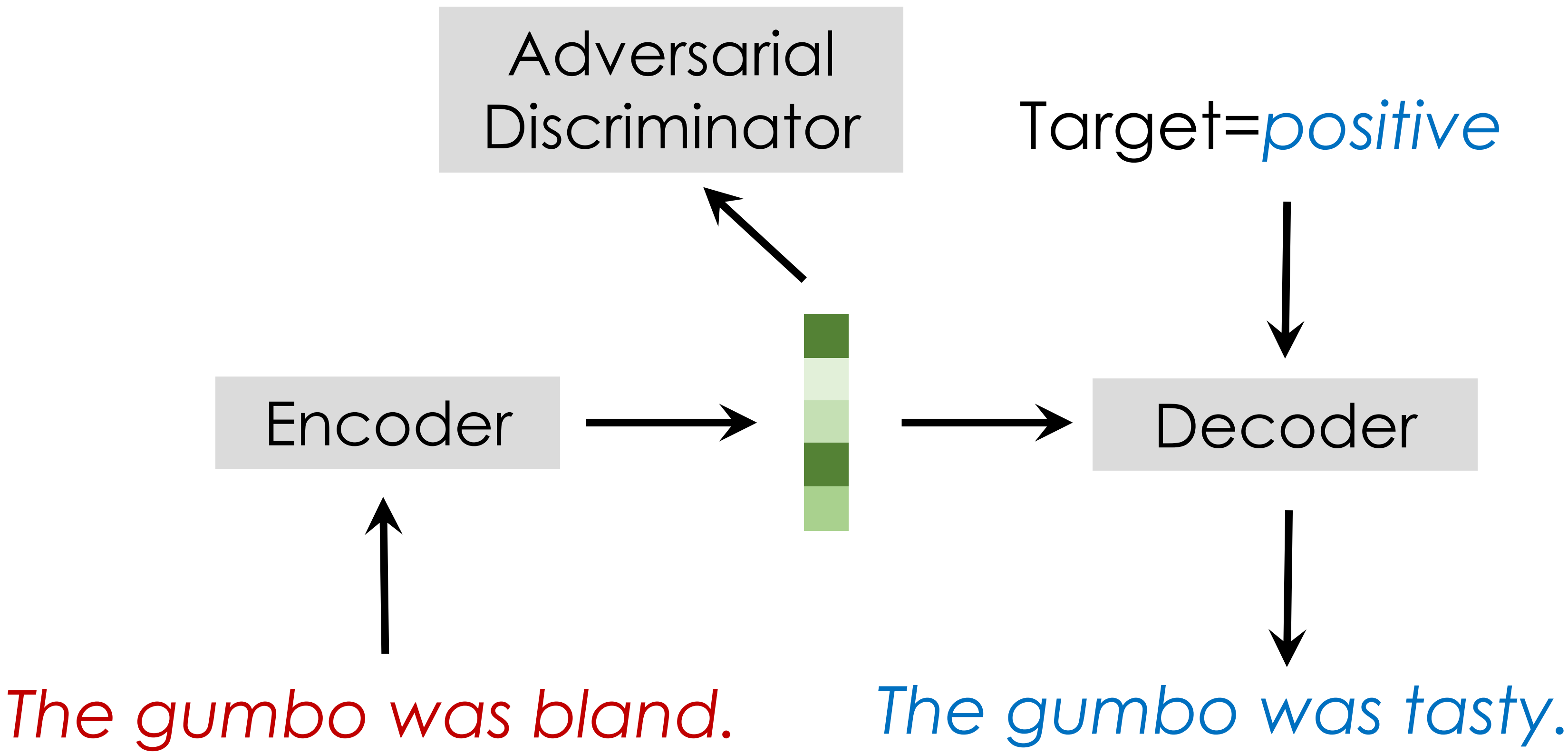
Basic auto-encoder



Can copy input and ignore target attribute

Shen et al. (2017); Fu et al. (2018)

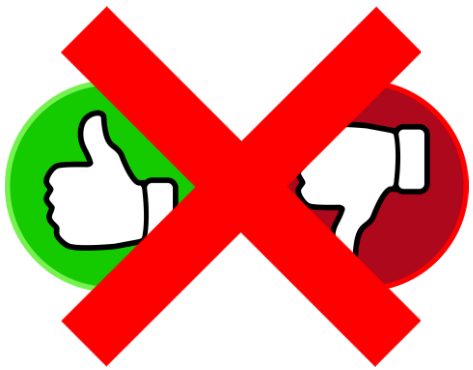
Adversarial content separation



Make discriminator unable to predict attribute

Shen et al. (2017); Fu et al. (2018)

Error Cases



No Attribute Transfer

Input: *“Think twice -- this place is a dump.”*

Output: *“Think twice -- this place is a dump.”*

Error Cases

Content changed

Input: *"The queen bed was horrible!"*

Output: *"The* ***seafood part*** *was wonderful!"*

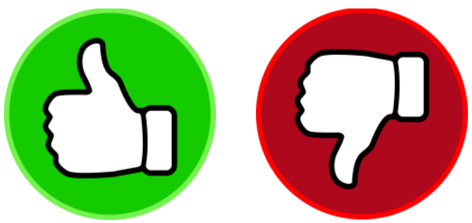
Error Cases

Poor grammar

Input: *"Simply, there are far superior places to go for sushi."*

Output: *"Simply, there are **far of vegan to go** for sushi."*

A balancing act



Attribute Transfer

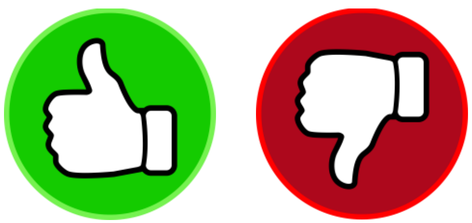
Content Preservation

Grammaticality

Outline

- Prior work with adversarial methods
- Simple baselines
- Simple neural methods

Pick two out of three



Attribute Transfer

Content Preservation

Grammaticality

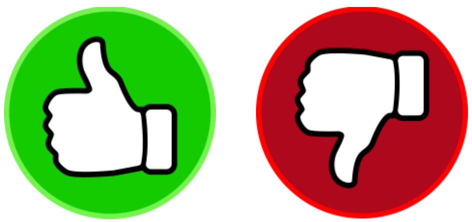
Content + Grammar

Content Preservation

Grammaticality

Just return the original sentence…

Attribute + Grammar



Attribute Transfer

Grammaticality

- Any sentence in the target corpus works!
- **Retrieve** one that has similar content as input

Retrieval Baseline

The gumbo was bland

The beignets were tasty
Great prices!
The gumbo was delicious
My wife loved the po'boy
…

Retrieval Baseline

The gumbo was bland

The beignets were tasty

Great prices!

The gumbo was delicious

My wife loved the po'boy

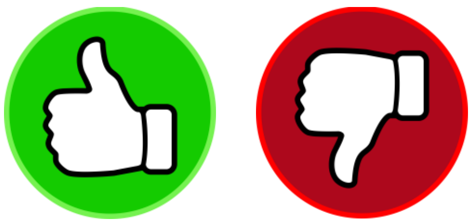
…

Retrieval Baseline

I hated the shrimp

The beignets were tasty
Great prices!
The gumbo was delicious
My wife loved the po'boy
…

Content + Attribute

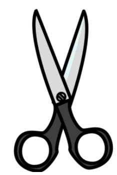


Attribute
Transfer

Content
Preservation

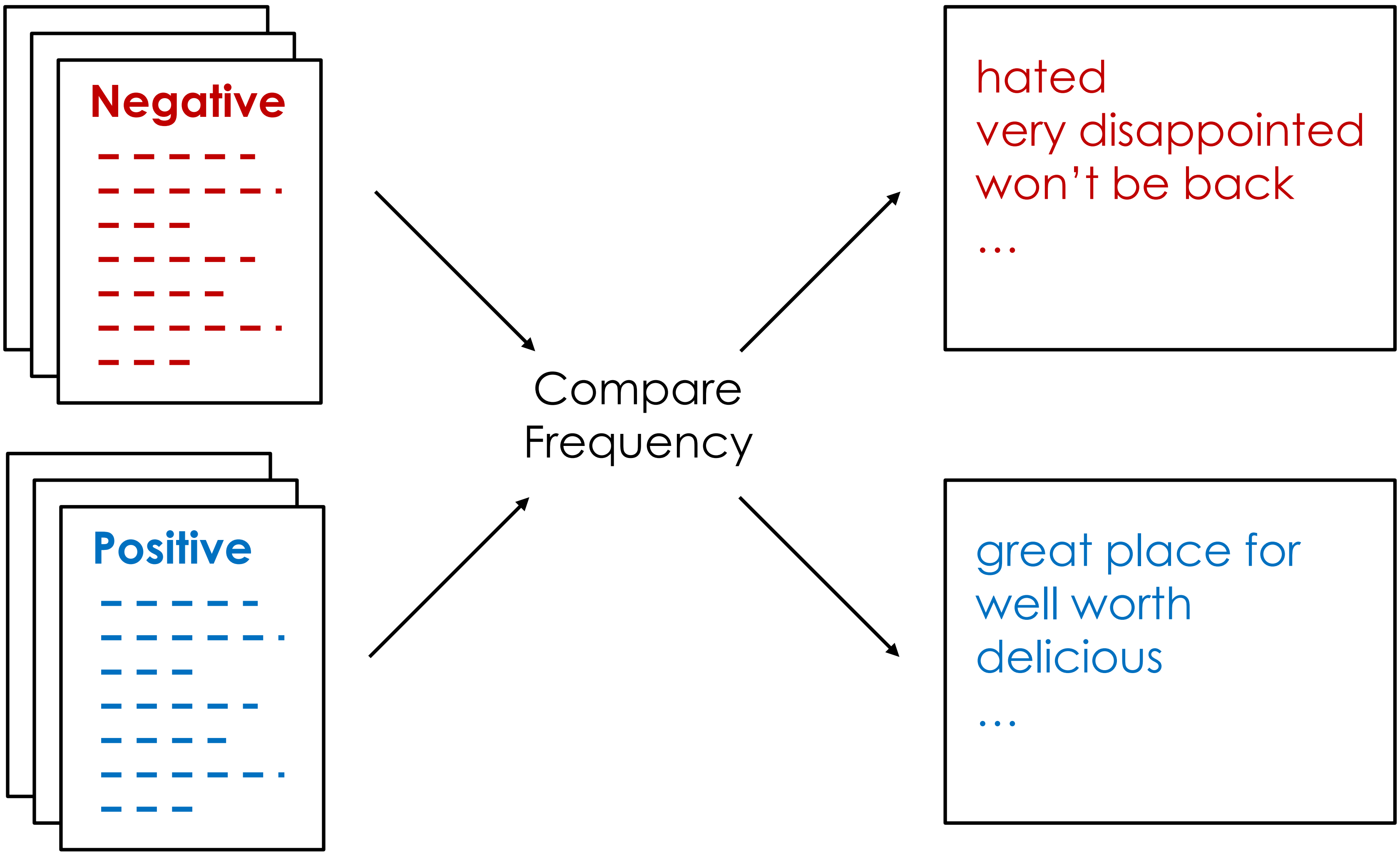
Content + Attribute

My wife hated the shrimp



- **Delete** markers of the source attribute
- Replace them with markers of the target attribute

Attribute Markers



Template Baseline

*My wife **hated** the shrimp*

Template Baseline

My wife ______ *the shrimp*

loved
tasty
polite
…

Template Baseline

My wife ______ *the shrimp*

loved
tasty
polite
…

Template Baseline

My wife ______ *the shrimp*

loved

tasty

polite

…

Template Baseline

My wife ______ the shrimp

loved

tasty

polite

…

Template Baseline

My wife _____ the shrimp

Retrieve attribute markers from similar contexts

The beignets were tasty
Great prices!
The gumbo was delicious
My wife loved the po'boy
...

Template Baseline

My wife ______ *the shrimp*

The beignets were tasty

Great prices!

The gumbo was delicious

My wife loved the po'boy

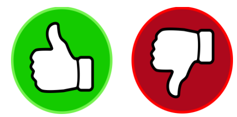
…

Retrieve attribute markers from similar contexts

Experiments

- Average over 3 datasets
  - Sentiment for Yelp reviews (Shen et al., 2017)
  - Sentiment for Amazon reviews (He and McAuley, 2016; Fu et al., 2018)
  - Factual to Romantic/Humorous style for image captions (Gan et al., 2017)

Experiments

- Human Evaluation
  - Likert scale from 1-5 for
    - Attribute transfer 
    - Content preservation 
    - Grammaticality 
  - Overall success: get ≥ 4 on each category

Results

Model	Attribute	Content	Grammar	Success
StyleEmbedding (Fu et al., 2018)				12%
MultiDecoder (Fu et al., 2018)				11%
CrossAligned (Shen et al., 2017)				12%

Results

Model	Attribute	Content	Grammar	Success
StyleEmbedding (Fu et al., 2018)				12%
MultiDecoder (Fu et al., 2018)				11%
CrossAligned (Shen et al., 2017)				12%
Retrieval Baseline				23%
Template Baseline				**24%**

Results

Model	Attribute	Content	Grammar	Success
StyleEmbedding (Fu et al., 2018)	2.6	3.2	3.3	12%
MultiDecoder (Fu et al., 2018)	3.0	2.8	3.1	11%
CrossAligned (Shen et al., 2017)	3.2	2.4	3.3	12%
Retrieval Baseline	**3.7**	2.7	**4.1**	23%
Template Baseline				**24%**

Results

Model	Attribute	Content	Grammar	Success
StyleEmbedding (Fu et al., 2018)	2.6	3.2	3.3	12%
MultiDecoder (Fu et al., 2018)	3.0	2.8	3.1	11%
CrossAligned (Shen et al., 2017)	3.2	2.4	3.3	12%
Retrieval Baseline	**3.7**	2.7	**4.1**	23%
Template Baseline	3.5	**3.9**	3.2	**24%**

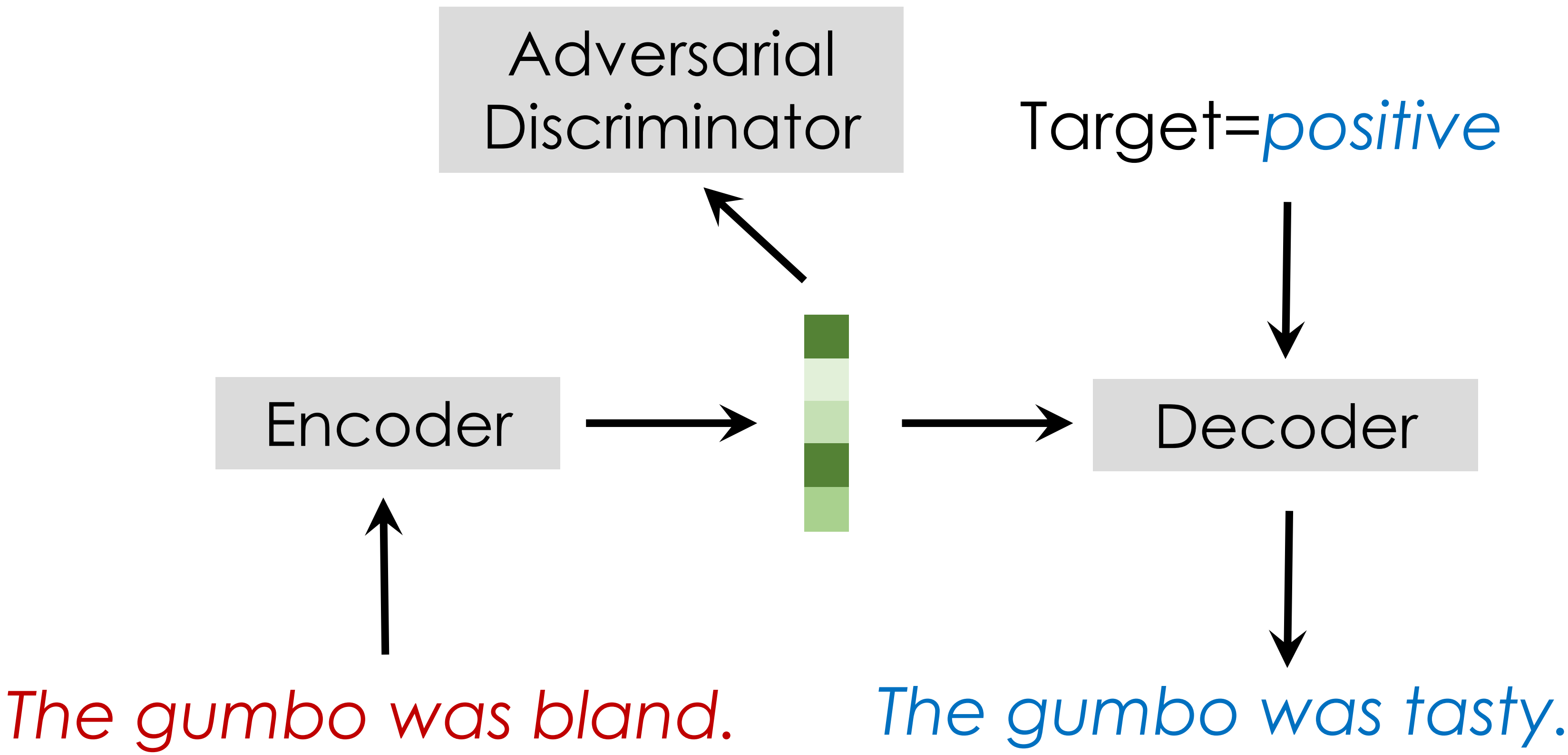
Results

Model	Attribute	Content	Grammar	Success
StyleEmbedding (Fu et al., 2018)	2.6	3.2	3.3	12%
MultiDecoder (Fu et al., 2018)	3.0	2.8	3.1	11%
CrossAligned (Shen et al., 2017)	3.2	2.4	3.3	12%
Retrieval Baseline	**3.7**	2.7	**4.1**	23%
Template Baseline	3.5	**3.9**	3.2	**24%**
Human	4.1	4.1	4.4	58%

Outline

- Prior work with adversarial methods
- Simple baselines
- Simple neural methods

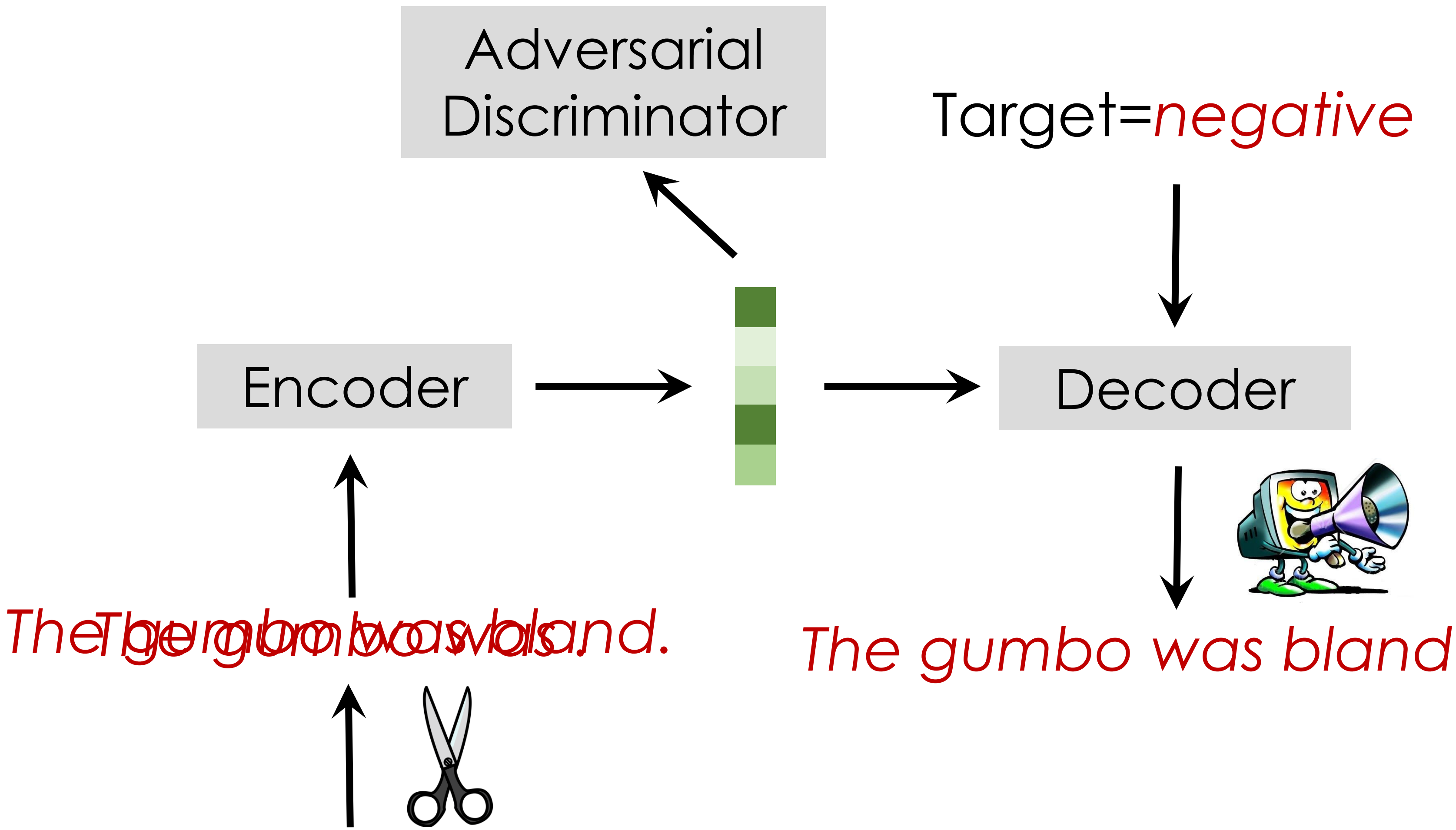
Content separation revisited



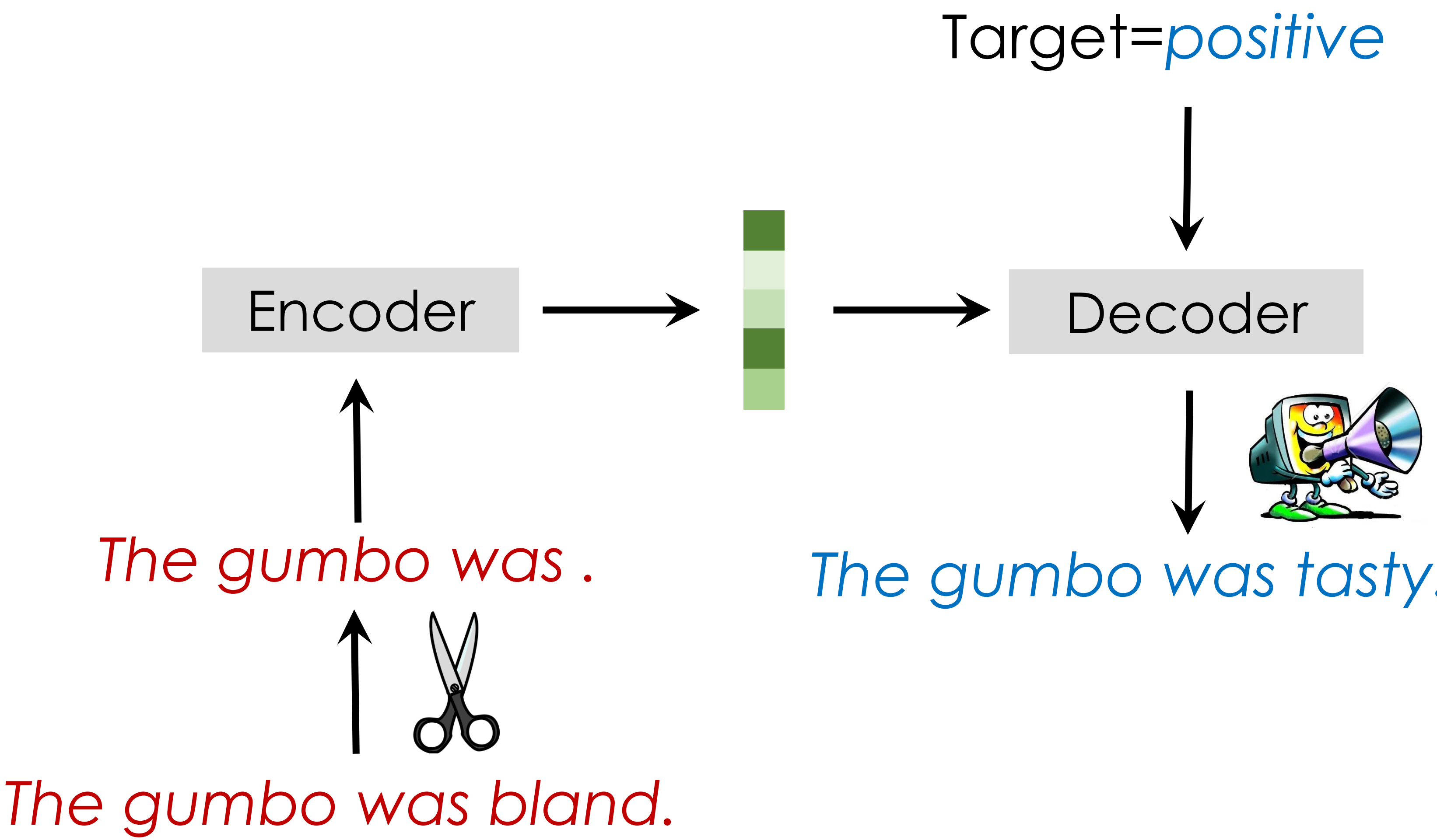
Make discriminator unable to predict attribute

Shen et al. (2017); Fu et al. (2018)

Delete and Generate



Delete and Generate



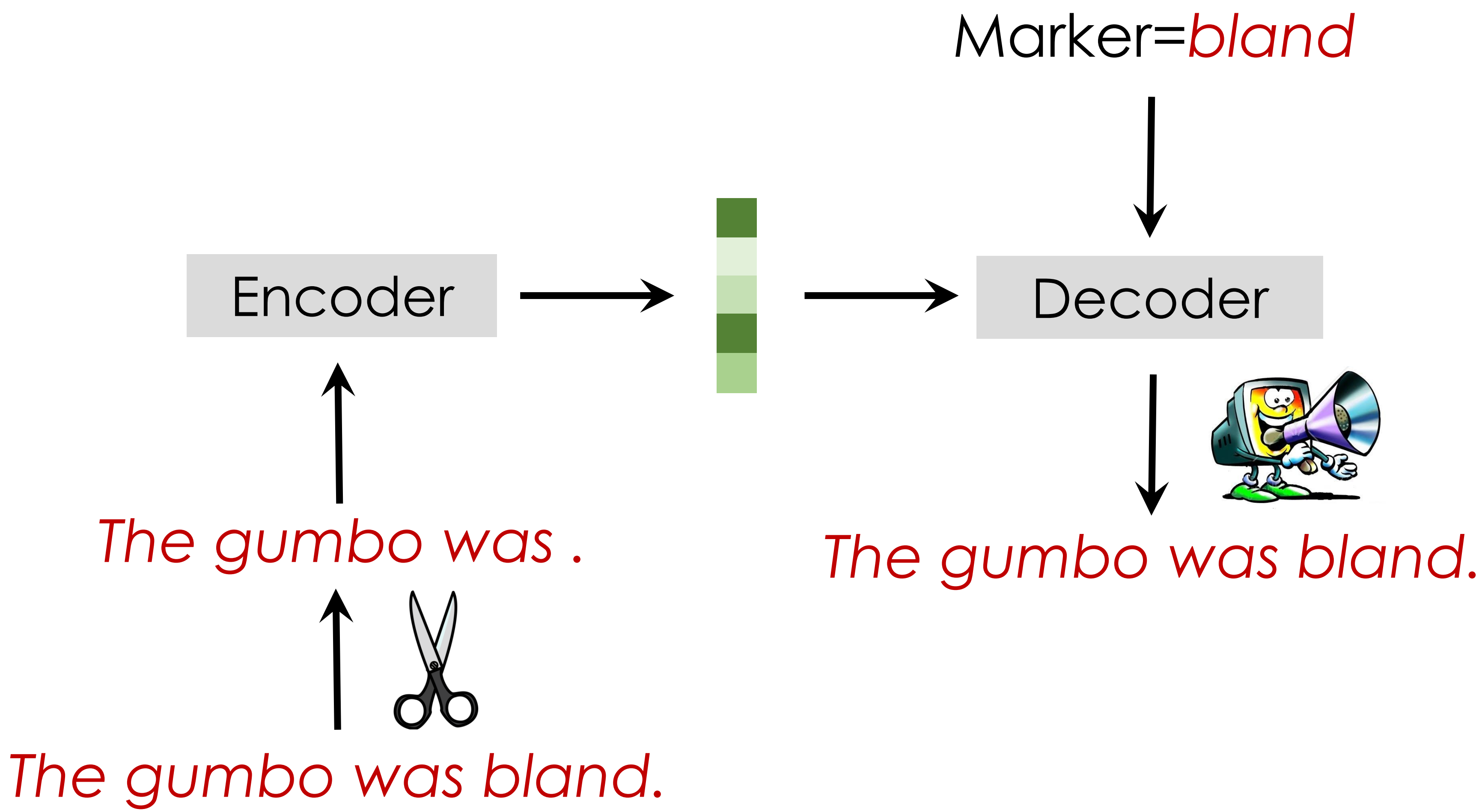
Results

Model	Attribute	Content	Grammar	Success
StyleEmbedding (Fu et al., 2018)	2.6	3.2	3.3	12%
MultiDecoder (Fu et al., 2018)	3.0	2.8	3.1	11%
CrossAligned (Shen et al., 2017)	3.2	2.4	3.3	12%
Retrieval Baseline	**3.7**	2.7	**4.1**	23%
Template Baseline	3.5	**3.9**	3.2	24%
Delete and Generate	3.6	3.6	3.4	**27%**
Human	4.1	4.1	4.4	58%

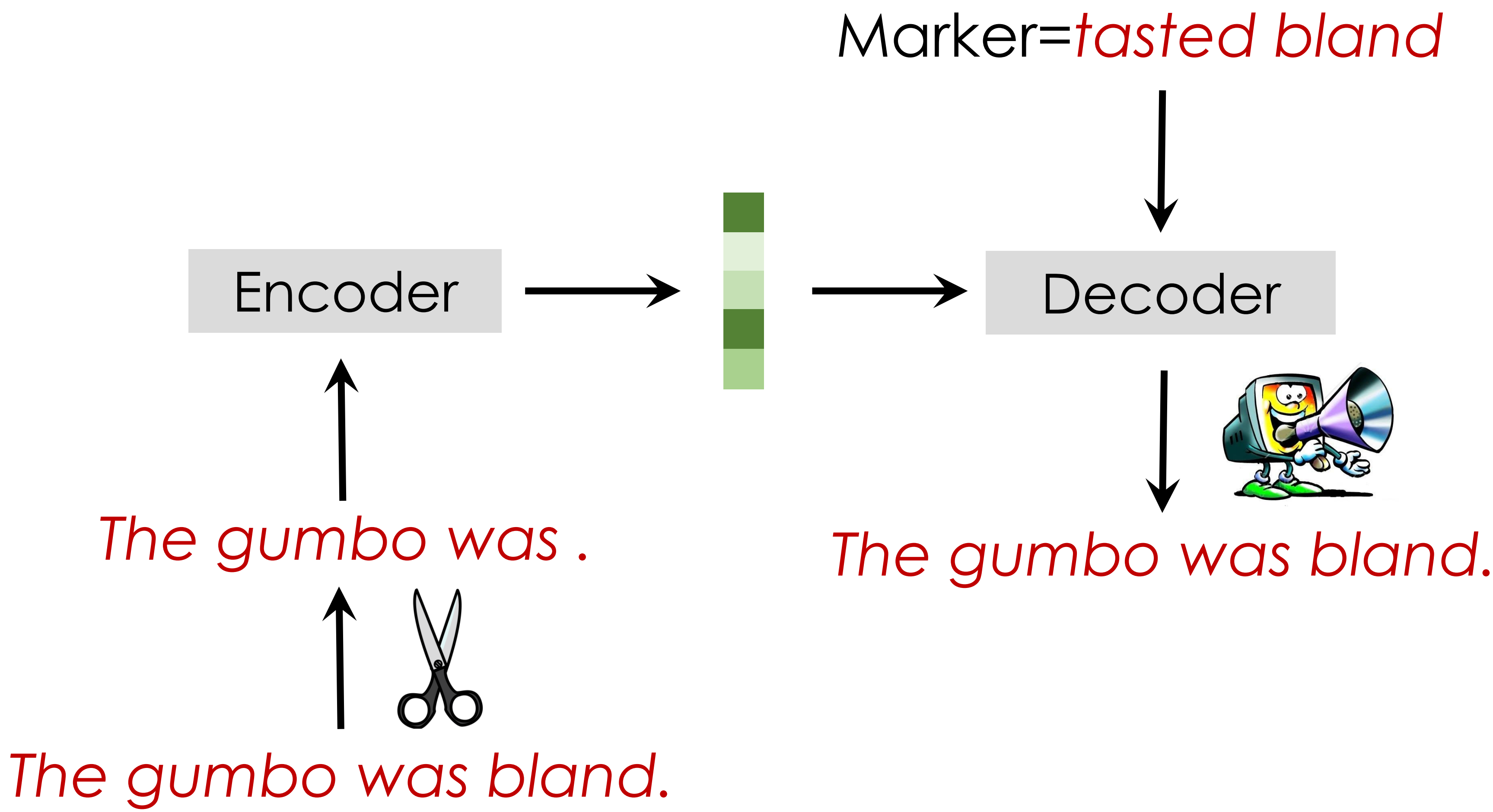
Context Cues

- Can retrieved attribute markers help the model?

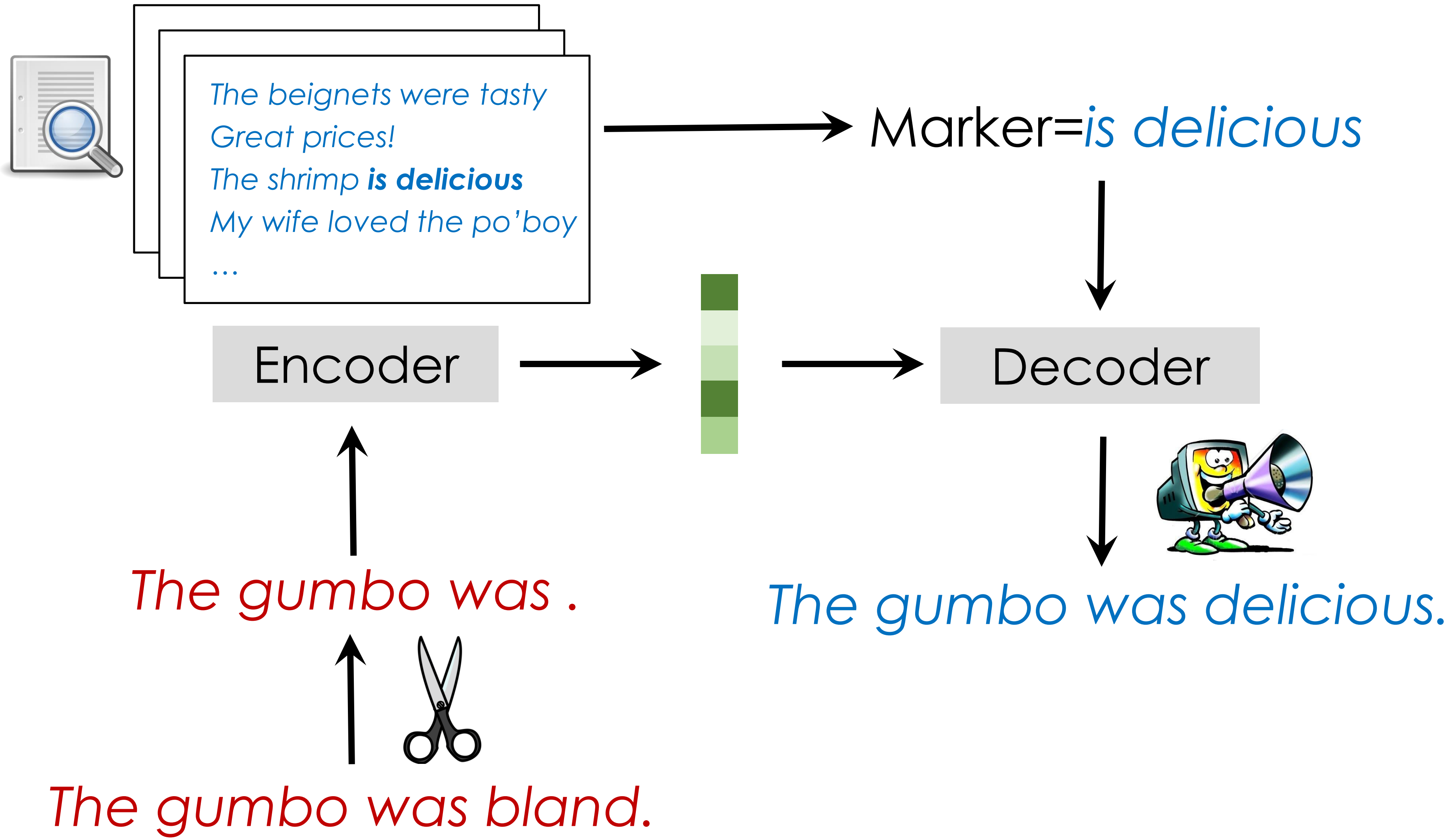
Delete, Retrieve, Generate



Delete, Retrieve, Generate



Delete, Retrieve, Generate



Results

Model	Attribute	Content	Grammar	Success
StyleEmbedding (Fu et al., 2018)	2.6	3.2	3.3	12%
MultiDecoder (Fu et al., 2018)	3.0	2.8	3.1	11%
CrossAligned (Shen et al., 2017)	3.2	2.4	3.3	12%
Retrieval Baseline	**3.7**	2.7	**4.1**	23%
Template Baseline	3.5	**3.9**	3.2	24%
Delete and Generate	3.6	3.6	3.4	27%
Delete, Retrieve, Generate	**3.7**	3.6	3.7	**34%**
Human	4.1	4.1	4.4	58%

Deleting too much…

Input: *“**Worst customer service** I have ever had.”*

Output: *“Possibly the best chicken I have ever had.”*

Deleting too little…

Input: *"I am actually afraid to open the remaining jars."*

Output: *"I am actually afraid to open the remaining jars **this is great.**"*

Thank you!

		
I ~~don’t like~~ NLP	*love it*	*I love NLP*
Delete	Retrieve	Generate

CodaLab http://tiny.cc/naacl2018-drg

GitHub https://github.com/lijuncen/Sentiment-and-Style-Transfer