Generating similes ~~effortlessly~~ *like a Pro* : A Style Transfer Approach for Simile Generation

Tuhin Chakrabarty

Smaranda Muresan

Nanyun (Violet) Peng

Natural Language Processing
Columbia University

PlusLab

UCLA NLP

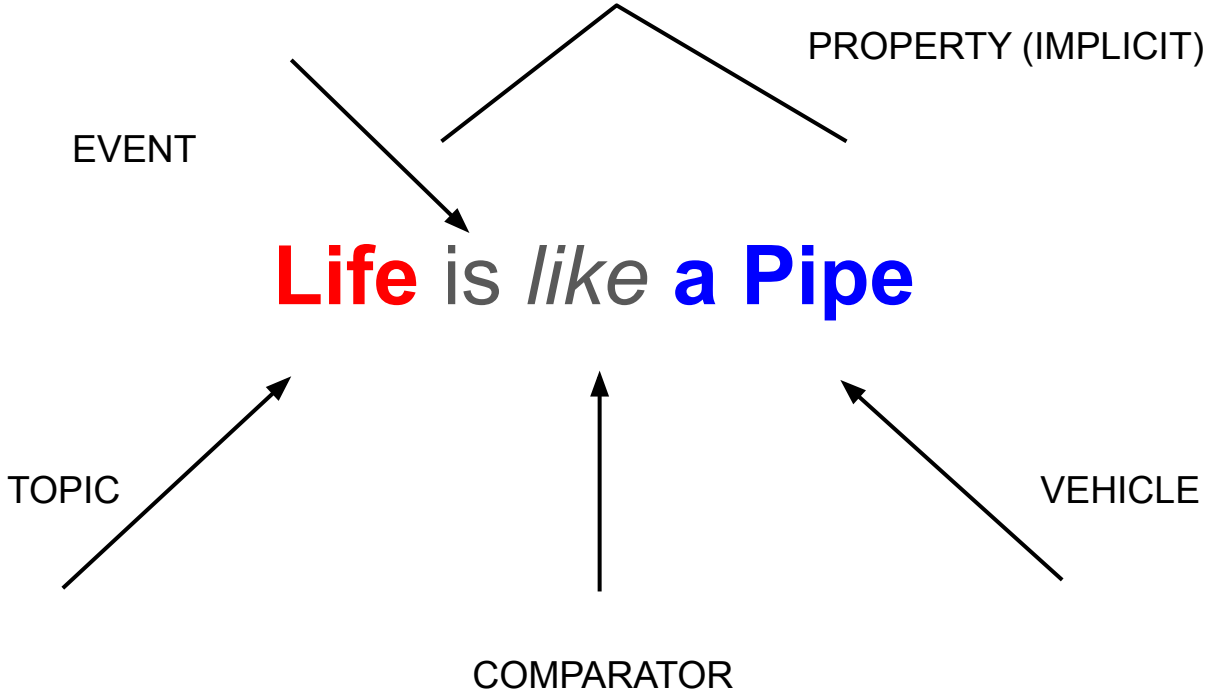
What is a Simile?

Simile is a figure of speech that compares two different kind of things, usually with the intent to make the description more emphatic or vivid

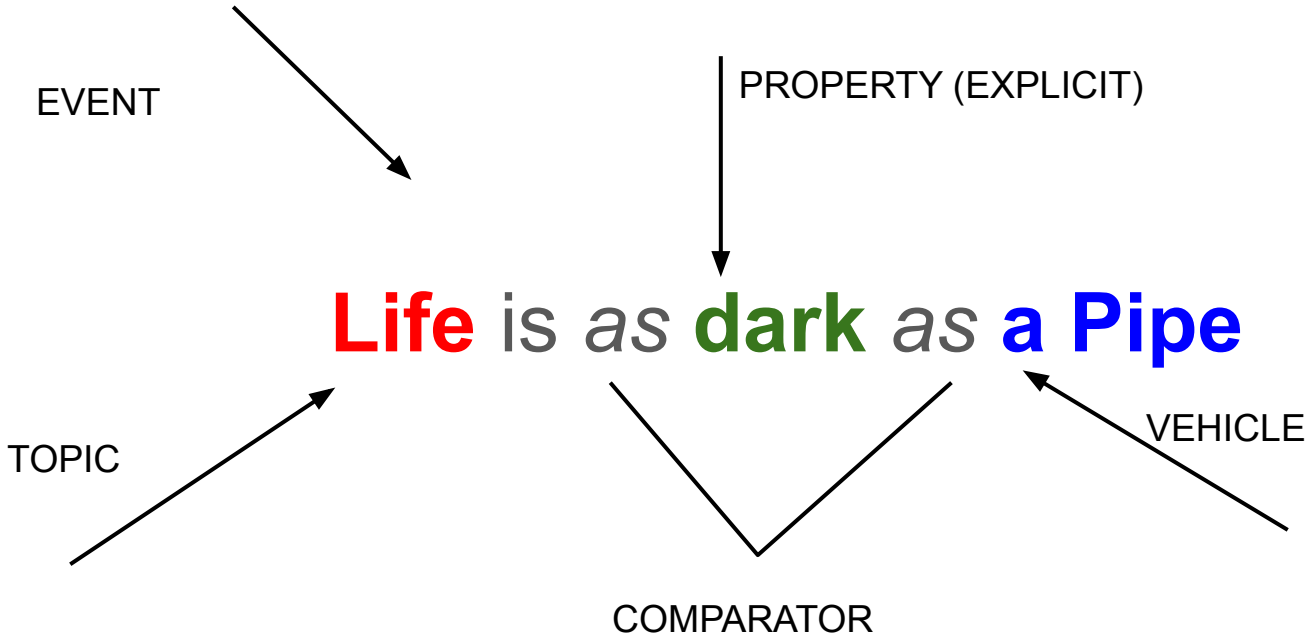
One of my favorite similes

Several interpretations on the Internet. What properties of a pipe can map to Life?
Long / Dark / Hollow typically all using commonsense knowledge

Simile Structure



Simile Structure



Task Definition

We focus on the task of generating a simile starting from a literal utterance that contains the **TOPIC, EVENT** and **PROPERTY.**

Literal

Having a thin figure, he looked **ugly**

Simile

Having a thin figure, he looked like **a walking corpse**

A Few Examples

Literal

Having a thin figure, he looked **ugly**

Simile

Having a thin figure, he looked like **a walking corpse**

Task Definition

Literal

It was obscene, but she was drawn to it, **fascinated**

Simile

It was obscene, but she was drawn to it, like a **moth to a flame**

What Does GPT-2 Say?

It was obscene, but she was drawn to it, **like a magnet**
It was obscene, but she was drawn to it **fascinated**, **like a bloodhound circling a blood trail**

Having a thin figure, he looked like **a well built young man**
Having a thin figure, he looked **ugly** like **a volcano**

Generating Simile is Challenging

- Lack of training data that consists of pairs of literal utterances and their equivalent simile in order to train a supervised generative model

- Ensuring that the generated simile makes a meaningful comparison between the TOPIC and the VEHICLE via the shared PROPERTY explicitly or implicitly expressed

Our approach

We frame this task as a style-transfer problem where the author's intent is to make the description of the **TOPIC** more emphatic by introducing a comparison with the **VEHICLE** via a shared **PROPERTY**

Given the lack of training data for simile generation, we propose **SCOPE** (Style transfer through COmmonsense PropErty).

- Automatic creation of a parallel corpus of [literal sentence, simile] pairs using Common sense Property
- Transfer learning from a pre-trained model for generating high quality similes.

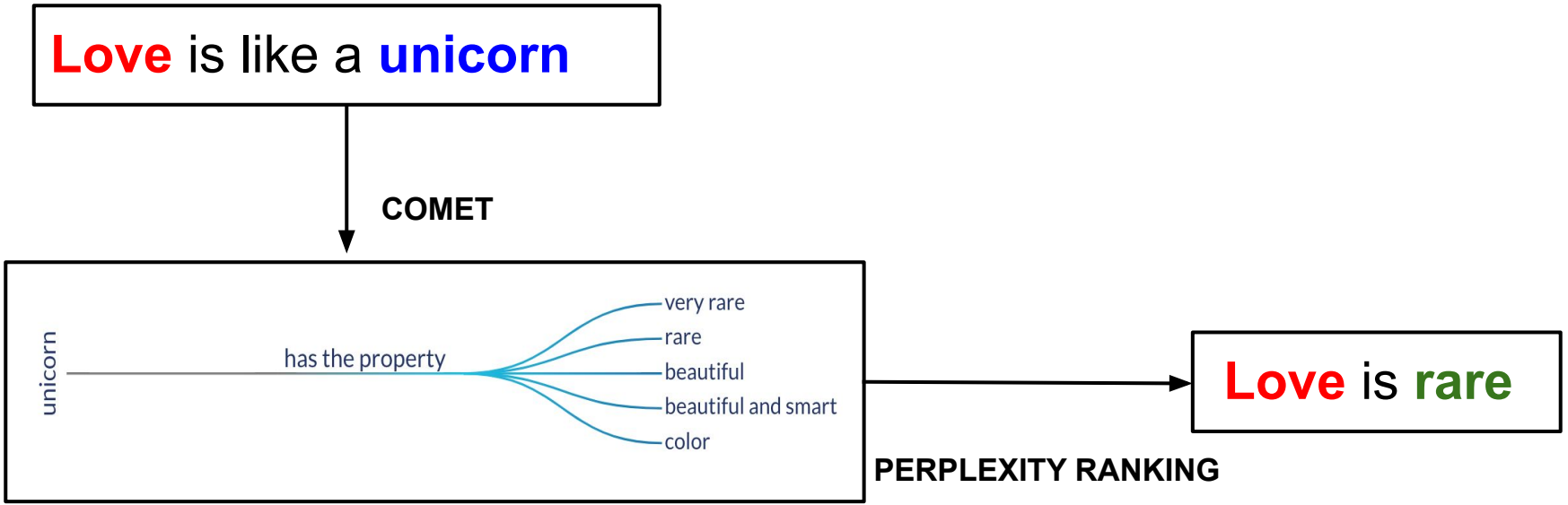
Parallel Data Creation

Crawl self labeled similes from Reddit

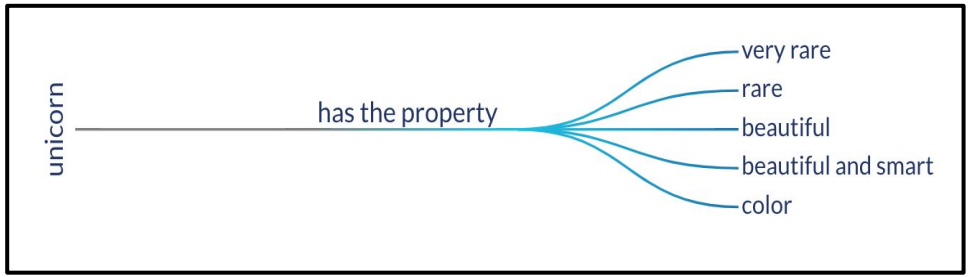
I'm at work laughing like a crazy person
Now the food I eat, tastes like a divine cuisine
It looked like a massacre
It almost sounded like a roar

Parallel Data Creation

Convert them to literal sentences using Common Sense

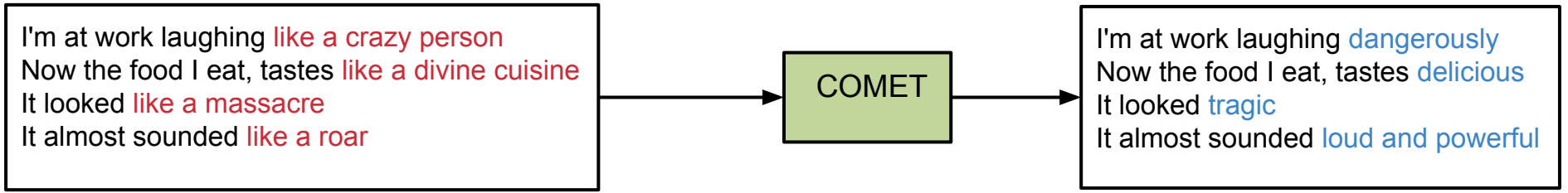


What is COMET ?

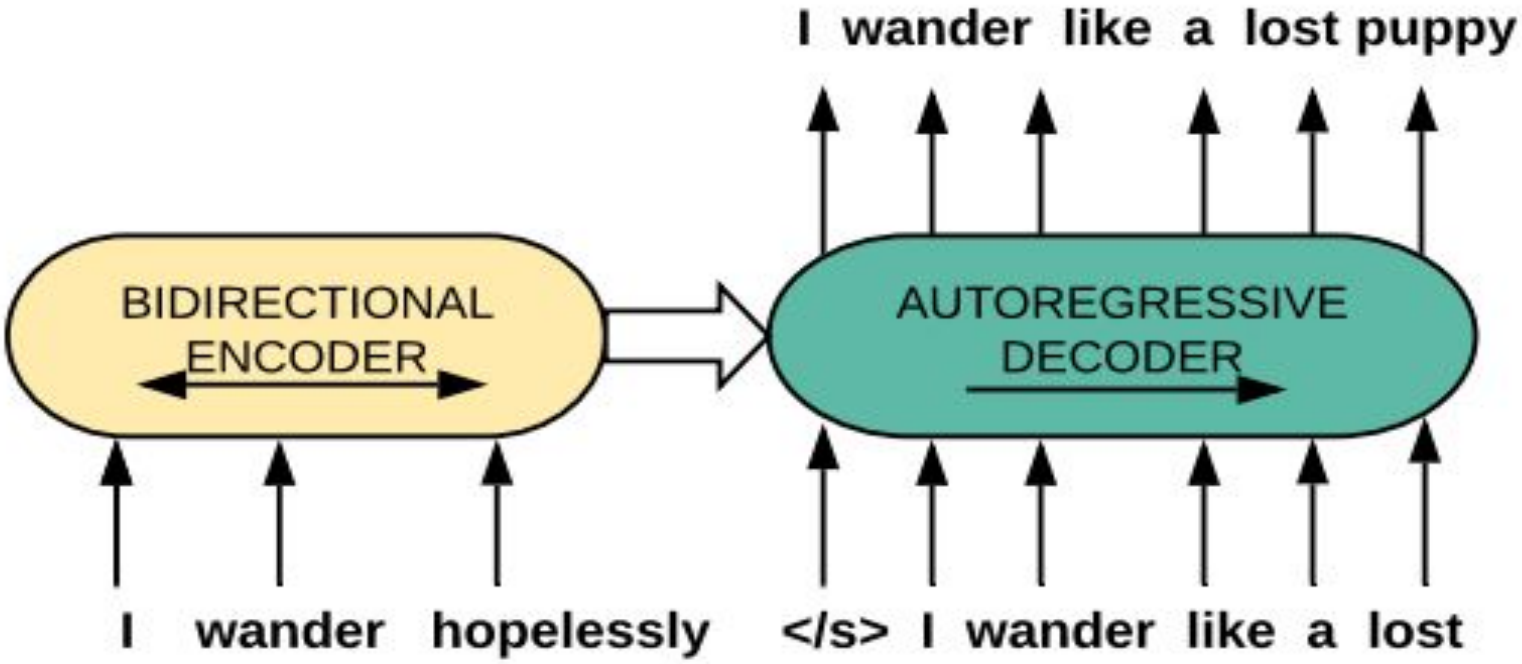


COMET is a transformer based GPT model fine-tuned on ConceptNet that can produce common sense knowledge on demand for any head entity that can be expressed through language

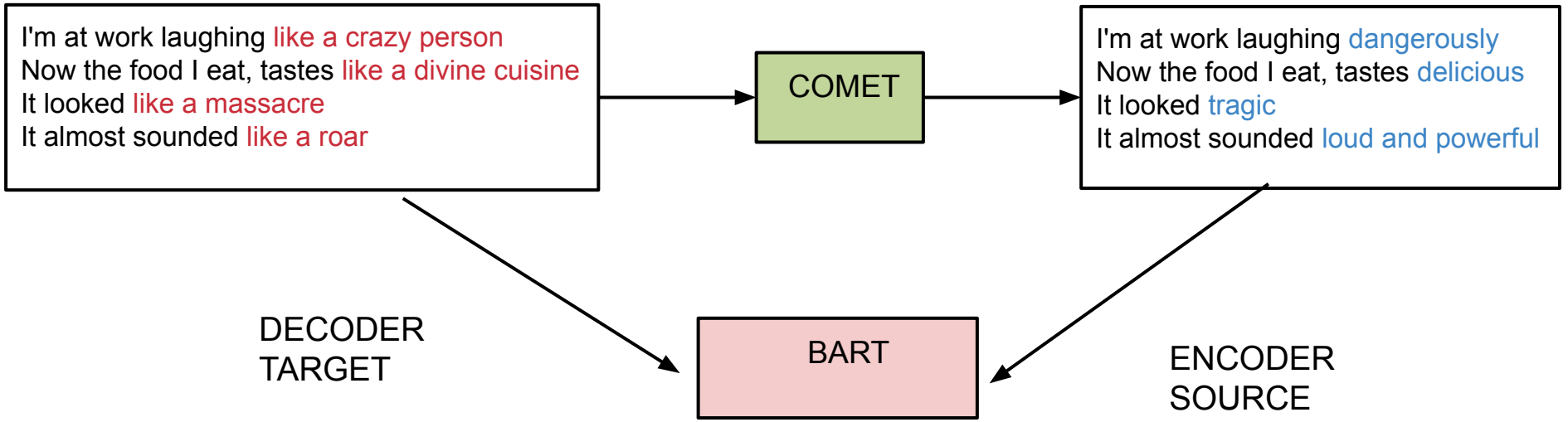
Parallel Data Creation



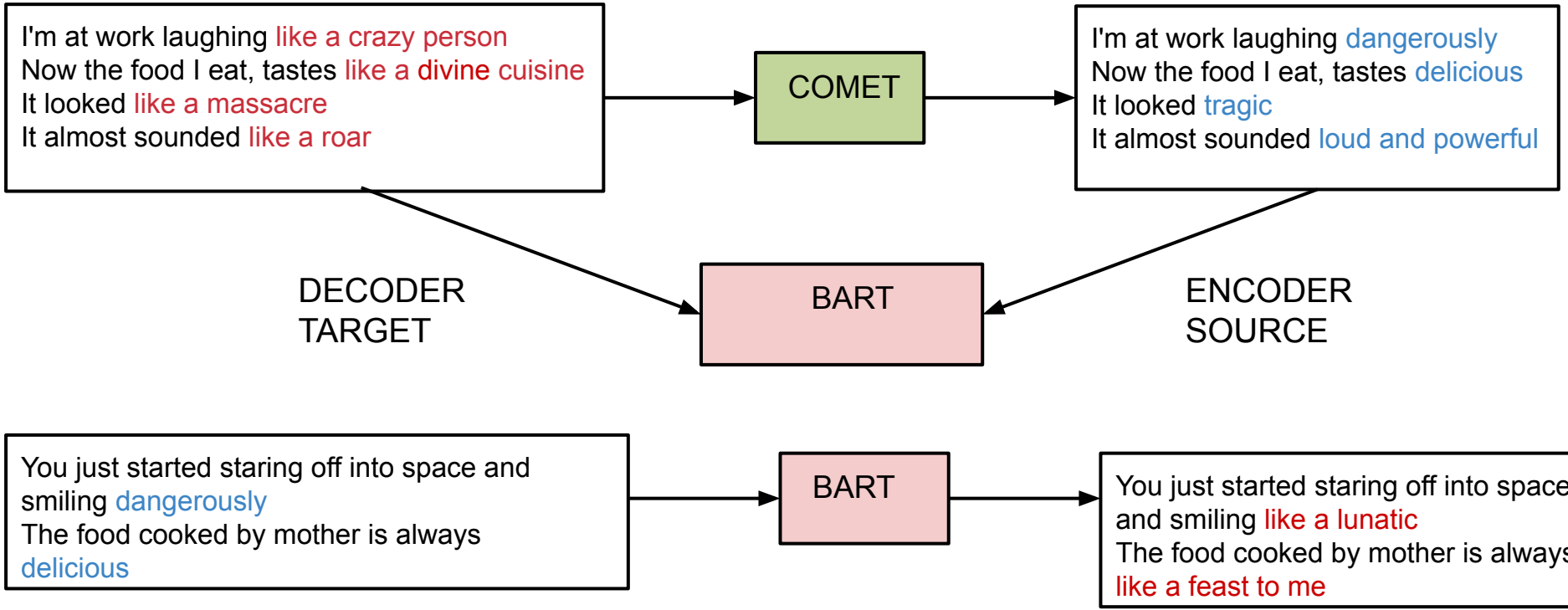
Finetune BART(Lewis et al 2019) A pre-trained seq2seq model



Finetune BART (A pre-trained seq2seq model)



Finetune BART (A pre-trained seq2seq model)



Evaluation Setup

Test Set

Test on 150 randomly selected literal sentences and compare with gold similes written by literary experts

Evaluation Criterion

1. Automated
2. Human
3. Task Based

Automated evaluation

- **BLEU-1 and BLEU-2**: ngram-overlap between reference and candidate

- **BERTScore** (Zhang et al., 2019) : a similarity score using contextual embeddings between reference & candidate

- **Novelty**: proportion of generated VEHICLE conditioned on an literal PROPERTY that does not appears in the training set.

Human evaluation

- **CREATIVITY** (C) “How creative are the utterances?”
- **OVERALL QUALITY** (OQ) “How good is the simile overall? (Turk guidelines was to score based on how creative, well formed, meaningful and relevant it is with respect to the literal utterance))
- **RELEVANCE1** (R1) “How relevant is the generated **VEHICLE** in terms of portraying the **PROPERTY**?”
- **RELEVANCE2** (R2) “How relevant is the **VEHICLE** to the **TOPIC** in the generation?”

Comparisons

- **RTRVL :** Use ConceptNet to retrieve vehicle using the property
- **BART :** Use pre-trained BART
- **META_M** [Stowe et al. (2020)] : Metaphor Masking model where we mask the literal property but keep the context

  The city was <MASK> The city was like a painting
- **SCOPE :** Style Transfer through Common Sense Property
- **HUMAN :** Human Written or Gold Similes

Sample Generation Examples

INPUT : Having a thin figure, he looked **ugly**

RTRVL : Having a thin figure, he looked ***like a pain***

META_M : Having a thin figure, he looked ***like a child***

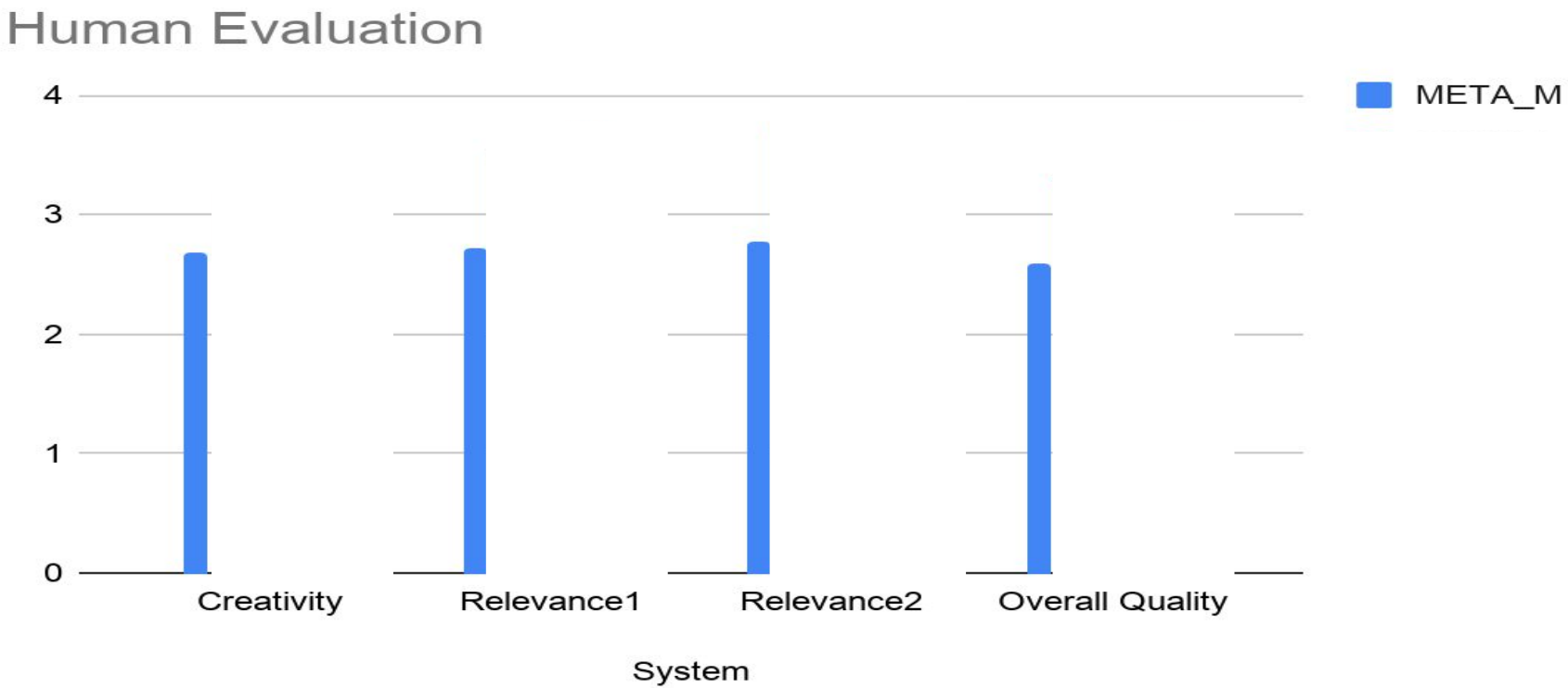
BART : Having a thin figure, he looked ***like a man***

SCOPE : Having a thin figure, he looked ***like a walking corpse***

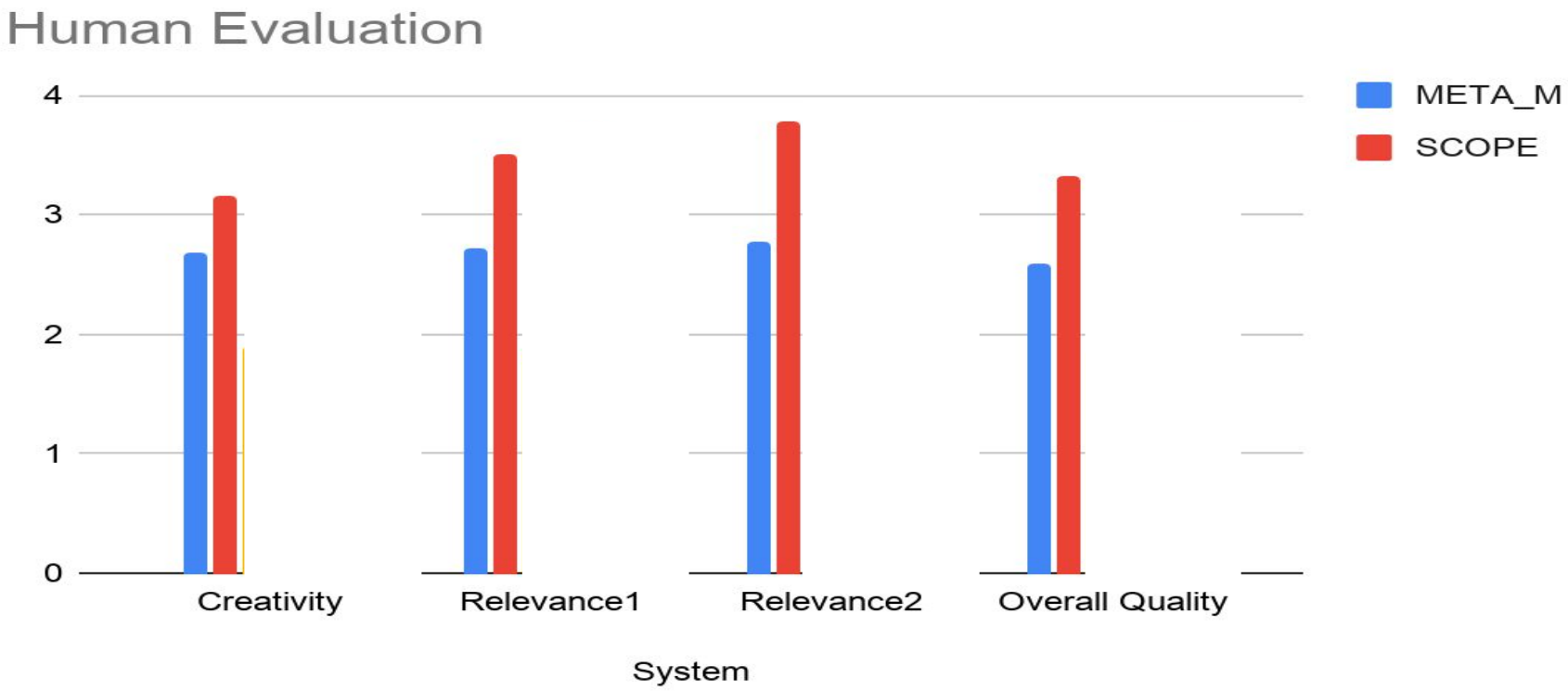
Auto Eval Result

	BLEU-1	BLEU-2	BERT SCORE	NOVELTY
RTRVL	0.0	0.0	0.13	92.6
BART	3.25	0.32	0.12	92.6
META_M	3.73	0.96	0.15	**93.3**
SCOPE	**8.03**	**3.59**	**0.18**	88.6

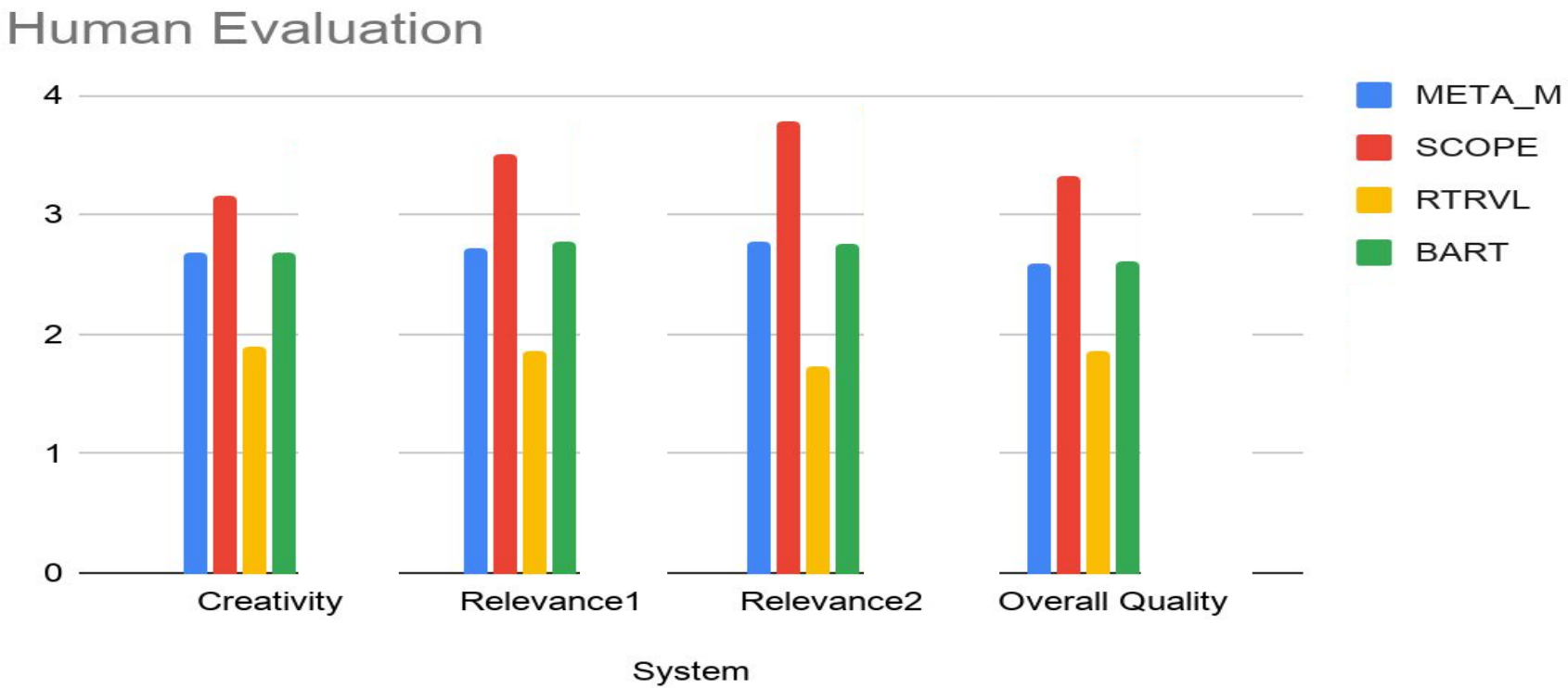
Human Eval Results



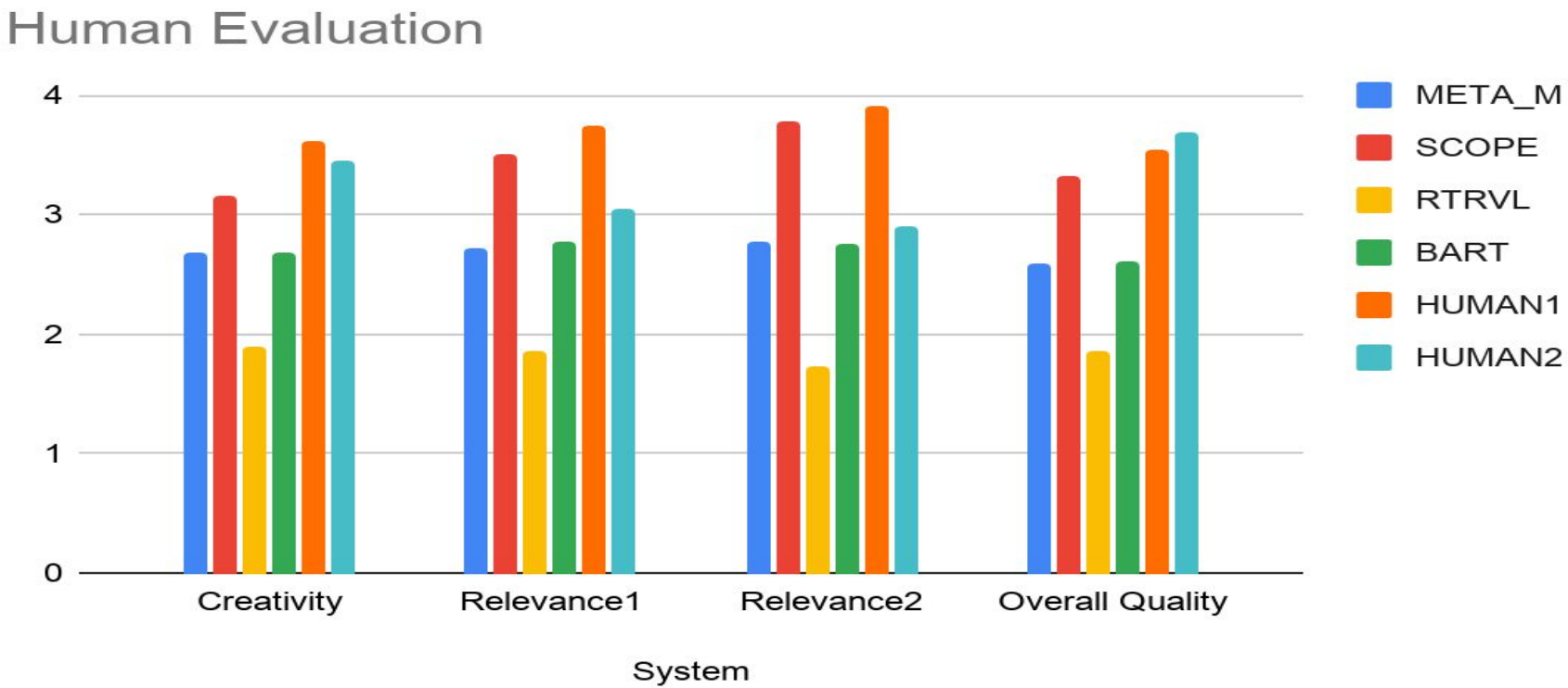
Human Eval Results



Human Eval Results



Human Eval Results



Task Based Eval

- Generate 5 sentence stories by fine-tuning GPT2

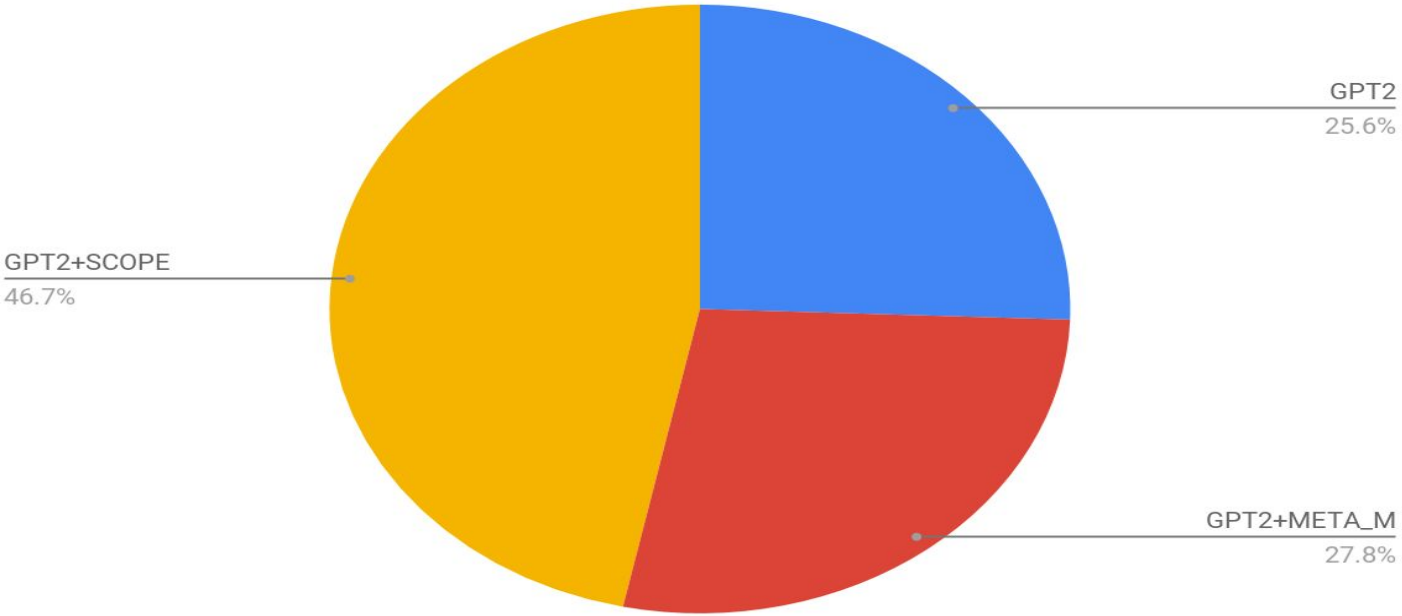
- Rewrite the stories where we replace one random literal sentence ending in an adjective / adverb with a generated simile from 2 best systems

- Ask Turkers if similes generated from SCOPE/META_M make the rewritten stories more creative and evocative ,

Task Based Eval

PROMPT : **Sunset**
STORYLINE: **sky → sunset → walk → walked → beautiful**
STORY: The sky was ~~**beautiful**~~ **[like a blue canvas]**. Jane wanted to see the sunset. She decided to go for a walk.She walked for a long time.When she was done she saw the sunset was beautiful.

Task Based Eval



Win percentage of each system when compared together. 3 systems presented to Turkers and asked to choose best in terms of creativity and evocativeness

Summary

- Big LMs even though good at literal text generation fail to be good at generating creative text rich in imagery or vividness
- We propose a *theoretically inspired approach* and employ *common sense knowledge*, showing how modeling the property correctly leads to better similes
- Our system outperforms existing baselines, but it is **still behind human performance**
- Generating creative and meaningful similes or metaphors opens new avenues for digital humanities research
- Creative figurative language generation is hard and interesting! Come work on it!

Code and Data

https://github.com/tuhinjubcse/SimileGeneration-EMNLP2020

https://github.com/PlusLabNLP/SimileGen

Contact

tuhin.chakr@cs.columbia.edu
smara@cs.columbia.edu
violetpeng@cs.ucla.edu