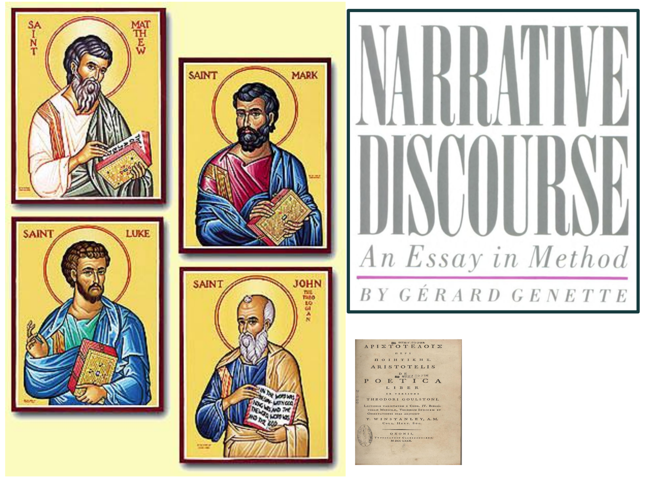
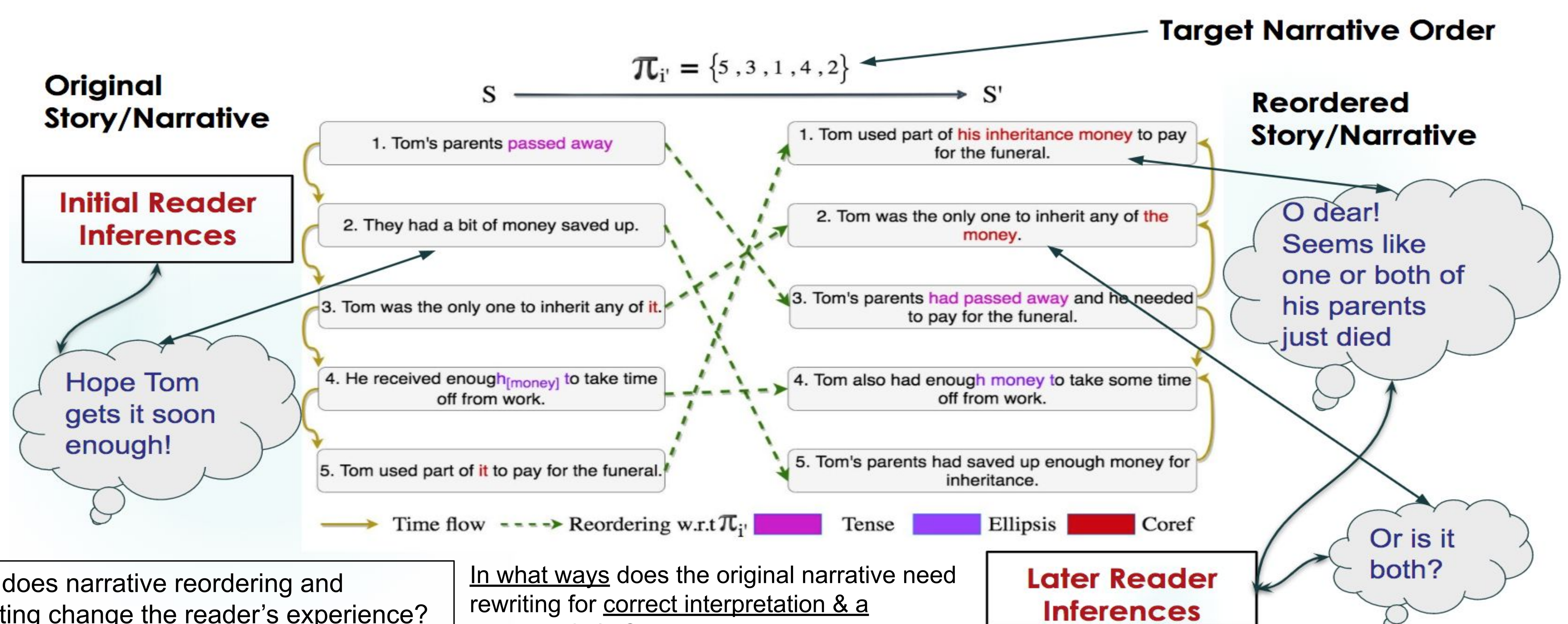


NAREOR: The Narrative Reordering Problem

Varun Gangal*[1], Steven Y. Feng*[1], Malihe Alikhani[2], Teruko Mitamura[1], and Eduard Hovy[2]
Carnegie Mellon University[1]   University of Pittsburgh[2]

Notion of Narrative: How a story is told/presented in the text → Studied since times of the *Poetics*
Narrator Dependent! → Many elements to vary → Character Focus, Omniscience, Narrative Order → 1 Story, many narratives! → That's how we end up with 4 Gospels, 300 Ramayanas, and many Bibles ..
Element We Focus On : Narrative Order → In what order are the story events presented in the text?
Story Order: Order in which events actually happened in the story → For linear narratives, equals narrative order
NAREOR: Given a story written in linear narrative order; reorder rewrite it to match a given, target narrative order while ensuring A) Plot Preservation B) Clear and Coherent Reader Interpretation
Our contributions: 1) No prior data → Collect *NAREORC* from 1500 ROCStories 2) What kind of rewrites do humans do? 3) How to finetune SOTA pretrained NLG models e.g BART? 4) How well do they do? Where do they fall short?

Illustrative Example



<u>How</u> does narrative reordering and rewriting <u>change the reader's experience</u>?

1. Readers make *different intermediate inferences* !
2. This can now alter the *interestingness* and *suspense* experienced …

<u>In what ways</u> does the original narrative need rewriting for <u>correct interpretation & a preserved plot</u>?

How do humans rewrite narratives?

Change Type	Story Examples with Changes Highlighted
Ellipsis (Sent: 5.7%) (Stor: 27.5%)	**S**: 1. All of the Ross family has red hair, except Henry. 2. Henry has blonde hair that is very curly. 3. Henry's father often teases Henry's mother about the mailman. 4. The mailman has blonde, curly hair, but he is very ugly. *5. His dad's teasing makes Henry feel bad.* ; $\pi_{i'}$: {1, 5, 4, 2, 3} **S'**: 1. All of the Ross family has red hair, except Henry. *2. His dad's teasing* ***about the mailman*** *makes Henry feel very bad.* 3. This is because the mailman has blonde, curly hair, but he is very ugly. 4. Henry also has blonde hair that is very curly. 5. Henry's father often teases Henry's mother about the mailman.
Tense (Sent: 19.1%) (Stor: 64.0%)	**S**: 1. Sam bought a new SUV. 2. It was all wheel drive. 3. He figured he would take it off road. *4. He* ***hit*** *a few hard bumps and broke his suspension.* 5. Sheepishly, he brought it to the dealership for repair. ; $\pi_{i'}$: {2, 3, 5, 1, 4} **S'**: 1. Sam's SUV was an all wheel drive. 2. He thought he could take it for a spin off road. 3. Embarrassed by the outcome of his drive, Sam took the car to the dealership for repair. 4. He had just bought the SUV. *5. The car* ***had hit*** *a few hard bumps and the suspension broke when Sam took it off road.*
Timexes (Sent: 34.0%) (Stor: 85.5%)	**S**: 1. There was once a kitten that did not have a home. *2. The poor kitten walked around cold and hungry.* 3. One day, a nice lady let the kitten into her home. 4. The woman gave the kitten food and a bed. 5. The kitten was happy to be adopted. ; $\pi_{i'}$: {4, 2, 5, 1, 3} **S'**: 1. A woman gave a home to a cat. *2.* ***Before that*** *it was cold and hungry.* 3. It made the cat happy to have a home. 4. The little cat originally was homeless. 5. But in the end, it met the nice woman and she let it in.
Coreference (Sent: 20.7%) (Stor: 71.5%)	**S**: 1. Jimmy wandered around the city looking for a place for a soda. 2. Before he knew it, he was in an unfamiliar area. 3. He was scared of strangers and didn't want to ask anyone. 4. Soon a policeman came by and asked if he was lost. *5.* ***He*** *told him that he was lost.* ; $\pi_{i'}$: {5, 4, 2, 1, 3} **S'**: *1.* ***Jimmy*** *told a police officer that he was lost.* 2. He was lucky the police showed up in the first place. 3. He had no idea where he was. 4. He had wandered off when trying to find somewhere to buy a soda. 5. It was pretty terrifying being all alone in a mysterious area with strangers.

Evaluating NAREOR

1. **Do generated stories actually stick to the Target Narrative Order?**
   a. Target Order Fidelity Metrics → Check against aligned original sentences, as per $\Pi_{i'}$.
   b. Only a sanity check metric → Gameable by "no edits"
2. **Fluency**: Use a LM to score, removing unigram frequency effects ()
3. **Reference Matching**: BLEU, METEOR, BERTScore
4. **Plot Preservation**: in part by 3 , but fully automatic metric an open challenge

How do we adapt SOTA NLG models for the task?

<u>The **Denoise** approach</u>

- ***Reorder*** input naively as per target narrative order
- Model has to learn **in-place edits** to ensure coherence and plot preservation.
- **Unsupervised Training:** Reconstruct original stories from randomly reordered +noised versions*
- **Supervised Training**: Construct reference target narrative from noised original story
- Noise with delete + insert

<u>Automatic Evaluation Results</u>

Method\Metric	BERTScore	BLEU	METEOR	TOF-BERTScore	TOF-METEOR
Human rewritings	N/A	N/A	N/A	66.85	56.79
GPT2-d-2S	60.75	37.01	45.20	79.23	74.23
GPT2-r-2S	58.03	32.57	40.85	73.04	63.00
BART-d-1S	67.14	44.73	49.88	95.61	93.43
BART-d-2S	<u>67.93</u>	<u>46.03</u>	<u>50.54</u>	93.55	90.81
BART-r-2S	67.16	44.63	49.16	91.32	86.43
T5-d-2S	**67.99**	**46.95**	**51.12**	94.20	91.83
T5-r-1S	66.24	43.40	48.20	89.85	84.26
T5-r-2S	66.62	44.30	49.00	91.61	86.16

<u>Human Evaluation Results</u>

Method\Metric	Fluency	Coherence	Logic	Plot-pres
Original stories	4.209	4.0	3.851	N/A
Human rewritings	3.797	3.723	3.784	3.972
GPT2-d-2S	3.635	3.399	3.399	3.708
GPT2-r-2S	3.595	3.378	3.291	3.375
BART-d-1S	3.628	3.412	3.318	3.847
BART-d-2S	**3.818**	<u>3.507</u>	3.493	3.722
BART-r-2S	3.757	3.439	3.493	<u>3.861</u>
T5-d-2S	3.764	3.419	<u>3.5</u>	**3.889**
T5-r-1S	3.655	3.378	3.486	3.847
T5-r-2S	<u>3.784</u>	**3.595**	**3.520**	<u>3.861</u>

Human Study I (Fluency, Coherence, Logic)   Human Study II (Plot-pres)

<u>Automatic & Human Eval</u>

- ***Denoise variants > Reorder*** on auto
- T5-reorder does the best overall on human, but BART-denoise best for preserving plot
- BART, T5 >> GPT2,stemming from reasons discussed below

<u>The **Reorder** approach</u>

- **Postfix tags sequence** indicating target narrative order
- Assign <X1>, <X2> … to mentions, add a portion of input with a list of their named mention text
- **Unsupervised Training**: Given naively reordered story, $\Pi_{i'}^{-1}$ as target narrative, rebuild original
- **Supervised Training**: Build reference ← input

What do the NLG models generated narratives look like?

Qualitative Examples with Narratives by the Different Models

Original (ex. 1)	Jimmy wandered around town looking for a soda. Before he knew it, he was in a strange area. He was scared of strangers and didn't want to ask anyone. Soon a policeman came by and asked if he was lost. He told him that he was lost. {5,4,2,1,3}
BART-d	**Jimmy told a policeman** that he was lost. The policeman **had come** by and asked if he **had been** lost. Before he knew it, he was in a strange area. Jimmy had wandered around town looking for a soda. He was scared of strangers and didn't want to ask anyone.
BART-r	**Jimmy told the policeman** that he was lost. The policeman came by and asked if he had lost **his wallet**. Before Jimmy knew it, he was in a strange area. He had wandered around town looking for a soda. He was scared of strangers and didn't want to ask anyone.
T5-d	**Jimmy told a policeman** that he was lost. **Soon after**, a policeman came by and asked if he was lost. Before he knew it, he was in a strange area. He had wandered around town looking for a soda. He was scared of strangers and didn't want to ask anyone.
T5-r	**Jimmy told a policeman** that he was lost. A policeman came by and asked **if he had a soda**. Before he knew it, he was in a strange area. He had wandered around town looking for a soda. He was scared of strangers and didn't want to ask anyone.

1. All models able to resolve the two "he"s correctly to their named mention as the sentence moves to first position → He told him that he was lost to Jimmy told a/the policeman that he was lost
2. BART-d adept at tense changes → Introduces "had come", "had been"
3. Hallucinations a concern → BART-r dreams up "his wallet", T5-r dreams up "if he had a soda". This alters the underlying plot

Original (ex. 2)	Every morning, a bird woke Fred up for work at 6 AM. Fred began to like the bird. And he enjoyed the sound of its singing. One day, the bird no longer sang. He hoped nothing bad had happened to the bird. {4,1,3,5,2}
Human	Fred didn't hear the bird singing this morning which is unusual. The bird had been waking him up every single day at 6 AM for quite some time. He had **grown accustomed to the bird being his alarm clock**. Now he's worried that something might have happened to the poor bird. He almost considers the bird a friend at this point.
GPT2-r	One day, Fred's bird began to sing. Every morning, **Fred woke him up** for work at 6 AM. **This was because he liked Fred**. He hoped nothing bad had happened to Fred. **This bird, however, did not like Fred**.
BART-r	**Fred** no longer sang. Every morning, he woke Fred up for work at 6 AM. He enjoyed the sound of his singing. He hoped nothing bad had happened to him. He **had begun** to like the bird.
T5-d	One day, the bird no longer sang. Every morning, a bird woke Fred up for work at 6 AM. He **had enjoyed** the sound of its singing. He had hoped nothing bad had happened to the bird. **After a while**, Fred began to like the bird.

1. GPT2-* more aggressive with rewriting than T5-* and BART-* - but hallucinates (the bird liked Fred, and then did not like him) & misattributes roles (Fred woke up the bird for work) considerably more
2. T5-d: enjoyed → had enjoyed ✔ , Timex "After a while" to beginning of last output sentence ✔

Appln I: How Interesting are the Generated Stories?

- *What's the "human ratings" used here?*
- How interesting is it vs original story?
- Rate 1-5 → 3=equivalent

Method	Interest
Human	3.75
BART-d	3.37
BART-r	<u>3.48</u>
T5-d	3.53
T5-r	3.30

- Both Human & BART-* / T5-* models generate interesting stories than original.

Appln II: Can generated stories act as challenge sets for temporal tasks?

Yes! For sentence ordering, sharp hits to performance on them

Conclusions & Future Work

- Explore control of ***other narratological variables*** such as focus [key character], narrator person.
- Bridge the gap between NLG models and Human.
- Devise a ***plot preservation metric*** for auto evaluation using event-event temporal and causal subgraphs.

Code+Data: https://tinyurl.com/yyhz4ehe

Paper: https://arxiv.org/abs/2106.02833