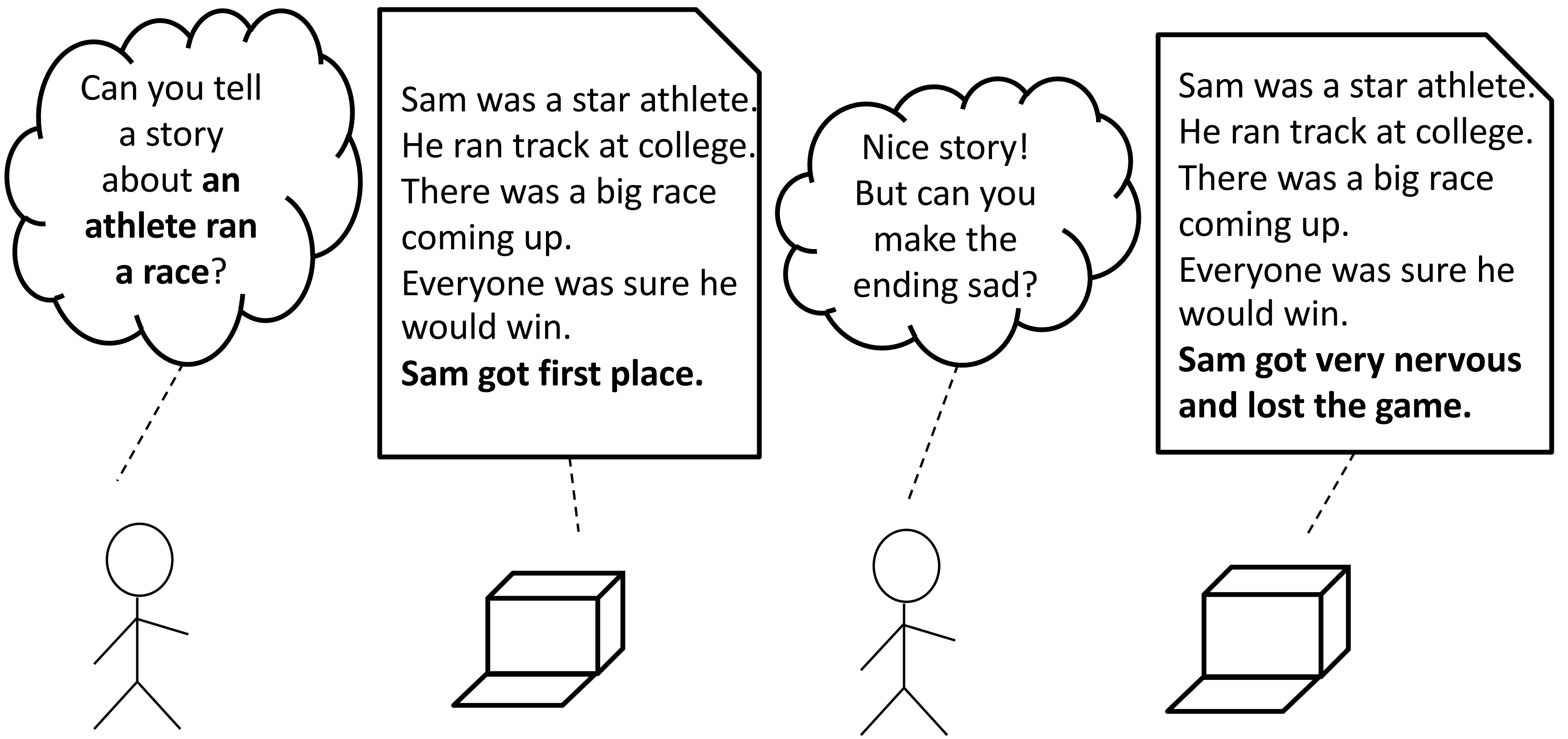
Analyze, Plan, and Compose: Towards Creative Text Generation

Nanyun (Violet) Peng

June 7th, NAACL Workshop on Narrative Understanding

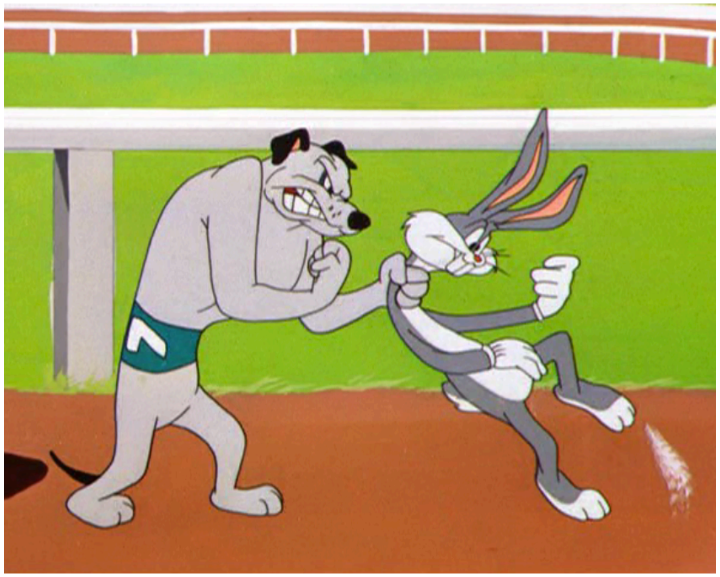
USC Viterbi
Information Sciences Institute

A Cartoon for Controllable Story Generation



Creative Pun Generation

- **Input:** hare, hair
- **Output:** The greyhound stopped to get a **hare** cut.



Challenges of (Neural) Text Generation

- **Title:** bicycle path accident
- **Generated Story:** sam bought a new bicycle. his bicycle was in an accident. his bicycle was in an accident. his bicycle was in an accident. his bicycle was totaled.
- **Title:** darth vader on earth
- **Generated Story:** it was a very windy day. i 've never been to it before. i do not know what to do. i do not know what to do. i think it is a good idea.

How About GPT-2?

- **bicycle path accident**, but it's still a pretty bad event. But, you know, they're also doing other events. ... I'm pretty excited, actually. I'm really excited. And I hope, you know, there's some other people out there feeling it. …
- **darth vader on earth** !!!
  HOLY HOP! HE JUST WOKE UP ON HIM!!! He did nothing wrong, he's a guy just like us. But it's funny how when someone says, "I'm so grateful he's alive. I'm so happy to see you're still alive."
  …

The Missing Ingredient for (Controllable) Creative Text Generation

- The Lack of Understanding and Analysis Tools
  - To be able to *control*, and generate *creative* elements.
- What's in a story?
  - Characters, genre, key events, morals, conflicts, ending spirit…
- What makes a pun funny/memorable?
  - Incongruity theory, ambiguity, distinctiveness…

From Analyzing to Generation

- Aspirational goal: analyze large corpus of creative contents (stories/poems/puns) to
  - Understand:
    - help humans to come up with concrete theories of story, humor in pun, etc.
  - Generate:
    - control the content
    - interact with human
    - generate novel contents that are coherent, creative, and interesting ….

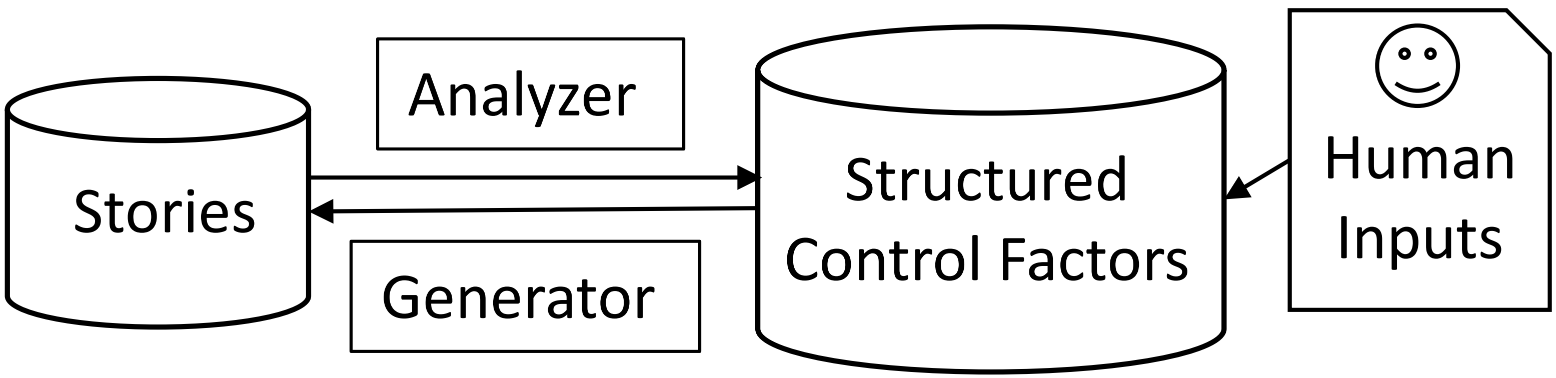
Outline

- Analyzing Story Structures
  - Plan-and-Write Hierarchical Story Generation

- Analyzing Humor in Puns
  - Pun Generation with Surprise

Analyzing to Generate Stories

- Analyzing stories to generate stories with minimal or no supervision.



Ending Valence: happy ending, sad ending

Storyline: tom → wanted → decided → practiced → won

Plan-and-Write Hierarchical Generation

- Can computer generate storylines automatically (given titles)?
  - Mimic human writers' common practice of writing a sketch/plot: have a big picture.
  - Equip our system with the ability to model "what happens next".
  - Computer and human can interactively modify the storylines, more fun interactions.

Yao et. al. (AAAI 2019). Demo: http://cwc-story.isi.edu/

Extracting Storylines

Title: christmas shopping

Story: <u>frankie</u> had christmas shopping to do.

she went to the <u>store</u>.

inside, she walked around looking for <u>gifts</u>.

soon her <u>cart</u> was full.

she <u>paid</u> and took her things home.

Storyline (unsupervised extraction): frankie -> store -> gifts -> cart -> paid

Title: farm

Story: bogart lived on a <u>farm</u>.

he loved <u>bacon</u>.

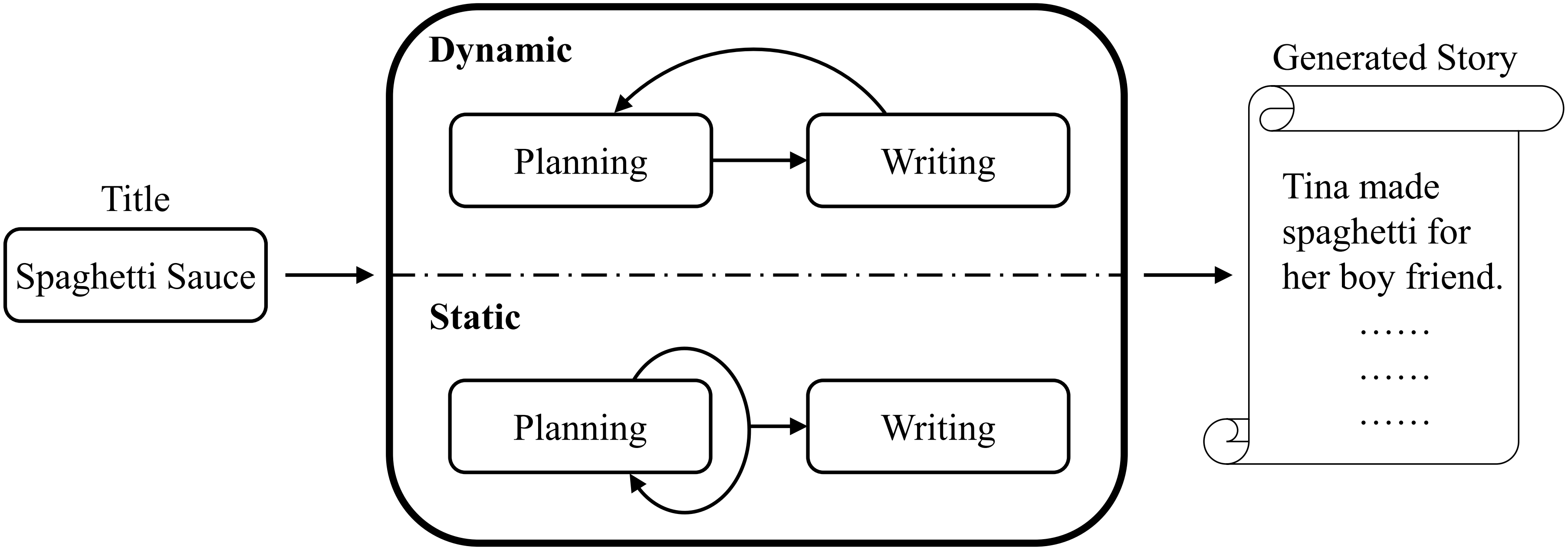
he <u>decided</u> to buy a pig. shortly after, he grew fond of the <u>pig</u>.

<u>bogart</u> stopped eating bacon.

Storyline (unsupervised extraction): farm -> bacon -> decided -> pig -> bogart

- Dataset: ROCStories – 90k turker generated five-line stories.
- Extraction tool: A modification to RAKE (rapid automatic keyword extraction), Rose et. al. 2010.
- Storyline composition: extracting one word/phrase per sentence, and order them according to the narrative order.

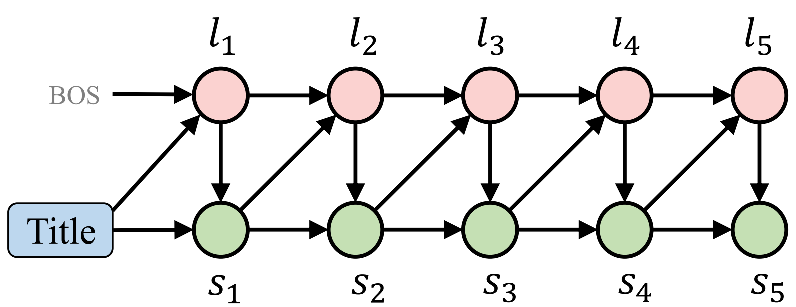
Plan-and-Write Overview



The *planning* component generates storylines from titles.
The *writing* component generates stories from storylines and titles.

Dynamic and Static Schemas

Dynamic Schema



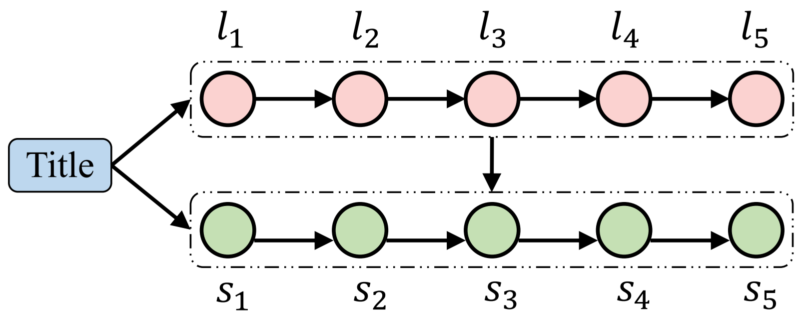
We define context as: $\boldsymbol{ctx} = [\boldsymbol{t}; \boldsymbol{s}_{1:i-1}]$

At the plan step, we model: $P(li|\mathbf{ctx}, l_{1:i-1})$

At the write step, we model: $P(s_i|\mathbf{ctx}, l_{1:i})$

The probabilities are computed by some specifically designed fusion-RNN cells.

Static Schema



At the plan step, we model: $P(li|\mathbf{t}, li_{-1})$

At the write step, we model: $P(s_i|\mathbf{ctx},l_{1:5})$

The probabilities are computed by standard language models and sequence to sequence with attention models.

Some Observations

- Plan-and-Write strategies generate more interesting, less repetitive stories.
- Plan-and-Write strategies generate more on-topic stories.
- Static strategy works better than dynamic strategy.

Generation Results

Title: gymnastics

Without Storyline Planning	With Storyline Planning
Story (generated): i wanted to learn how to draw. so, i decided to go to the gym. i went to the local gym. i got a lot of good grades. i was very happy.	Storyline (generated): wanted -> decided -> class -> practiced -> well Story (generated): i wanted to be a gymnast. i decided to learn how to do gymnastics. i decided to take a class. i practiced every day. i was able to do well on the class.

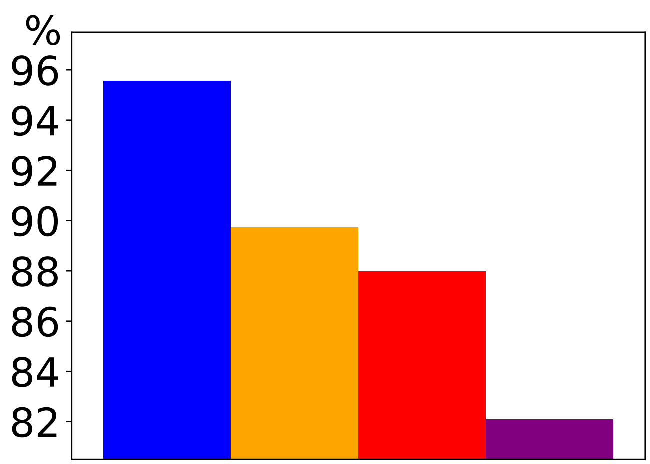
Generation Results (Cont.)

Title: build a house

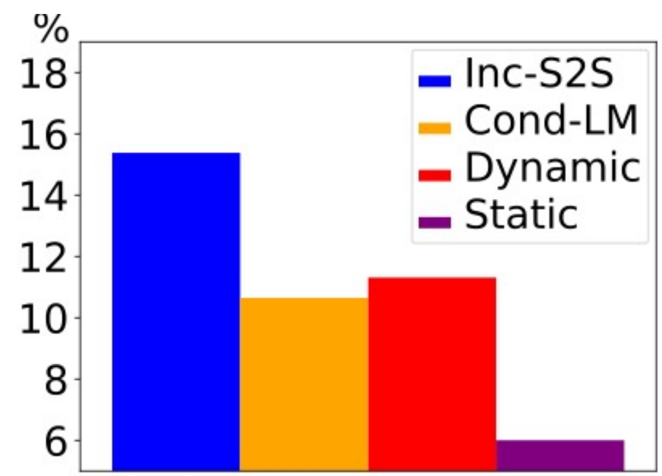
Without Storyline Planning	With Storyline Planning
Story (generated): When I was young, I wanted to build a house. I went to the store and bought all the supplies I needed. I went to the store and bought all the supplies I needed. I went to the store and bought all the supplies I needed. I bought the supplies and went home .	Storyline (generated): build house -> decided -> built -> finished -> proud Story (generated): john wanted to build a house. he decided to build a house. he built a small house. he finished the house. he was proud of himself.

Quantitative Results on Repetition

Inter- and intra-story tri-grams repetition rates, the lower the better. We also conduct the same computation for four and five-grams and observed the same trends. As reference points, the whole story repetition rates on the human-written training data are 34% and 0.3% for the inter- and intra-story measurements respectively.



Inter-story repetition rates



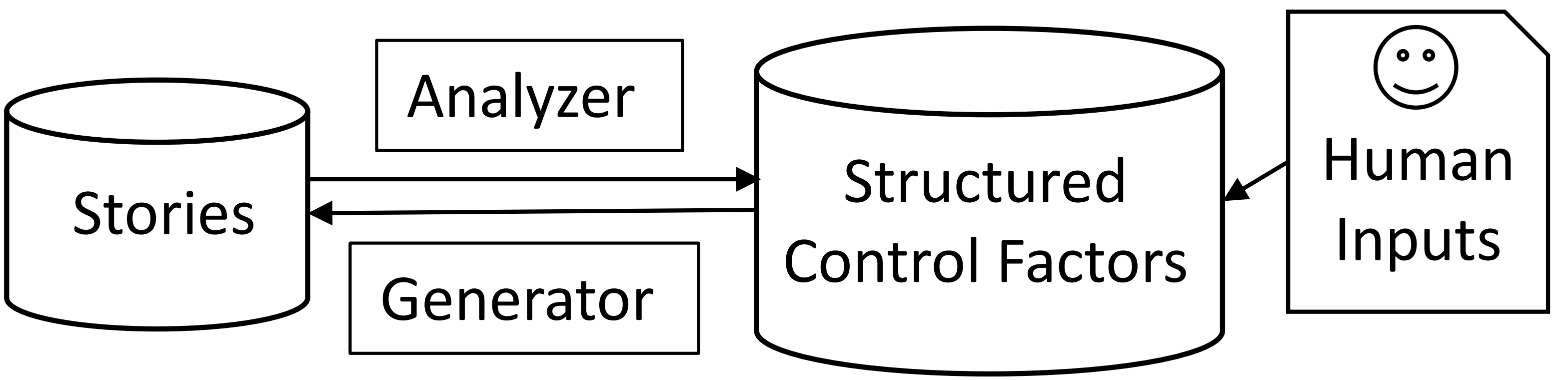
Intra-story repetition rates

User Preferences

Aspect	Dynamic v.s. Inc-S2S		Static v.s. Cond-LM		Static v.s. Dynamic	
	Dyna.	Inc.	Static	Cond.	Static	Dyna.
Fidelity	**35.8%**	12.9%	**38.5%**	16.3%	**38.0%**	21.5%
Coherence	**37.2%**	28.6%	**39.4%**	32.3%	**49.5%**	28.3%
Interesting	**43.5%**	26.7%	**39.5%**	35.7%	**43.6%**	34.4%
Overall	**42.9%**	27.0%	**40.9%**	34.2%	**50.1%**	30.1%

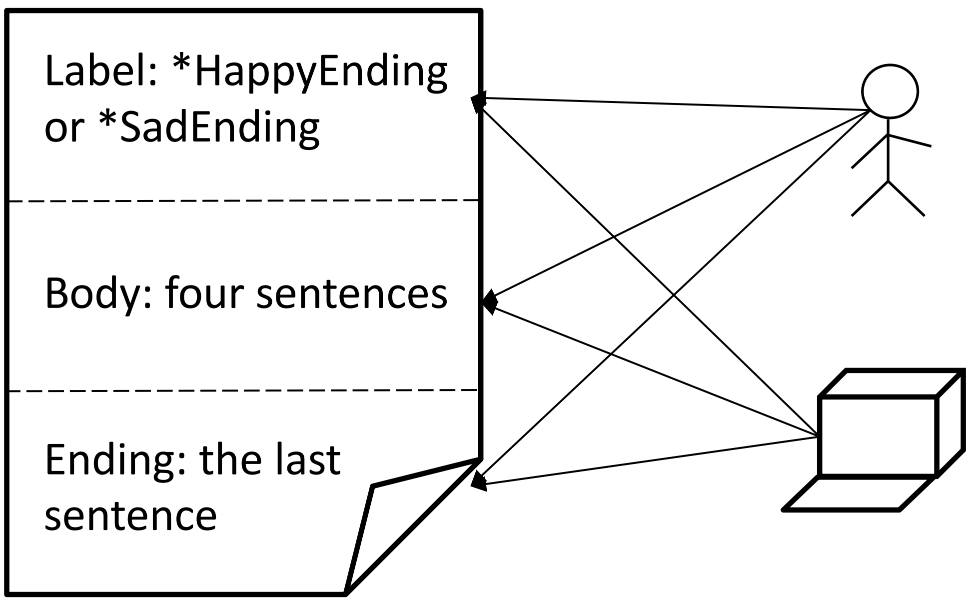
The human study is conducted on Amazon Mechanical Turk (AMT). 233 users were participated in the study.

Beyond Hierarchical Generation…



Controllable Generation

Creative Story Generation with Ending Valence Control

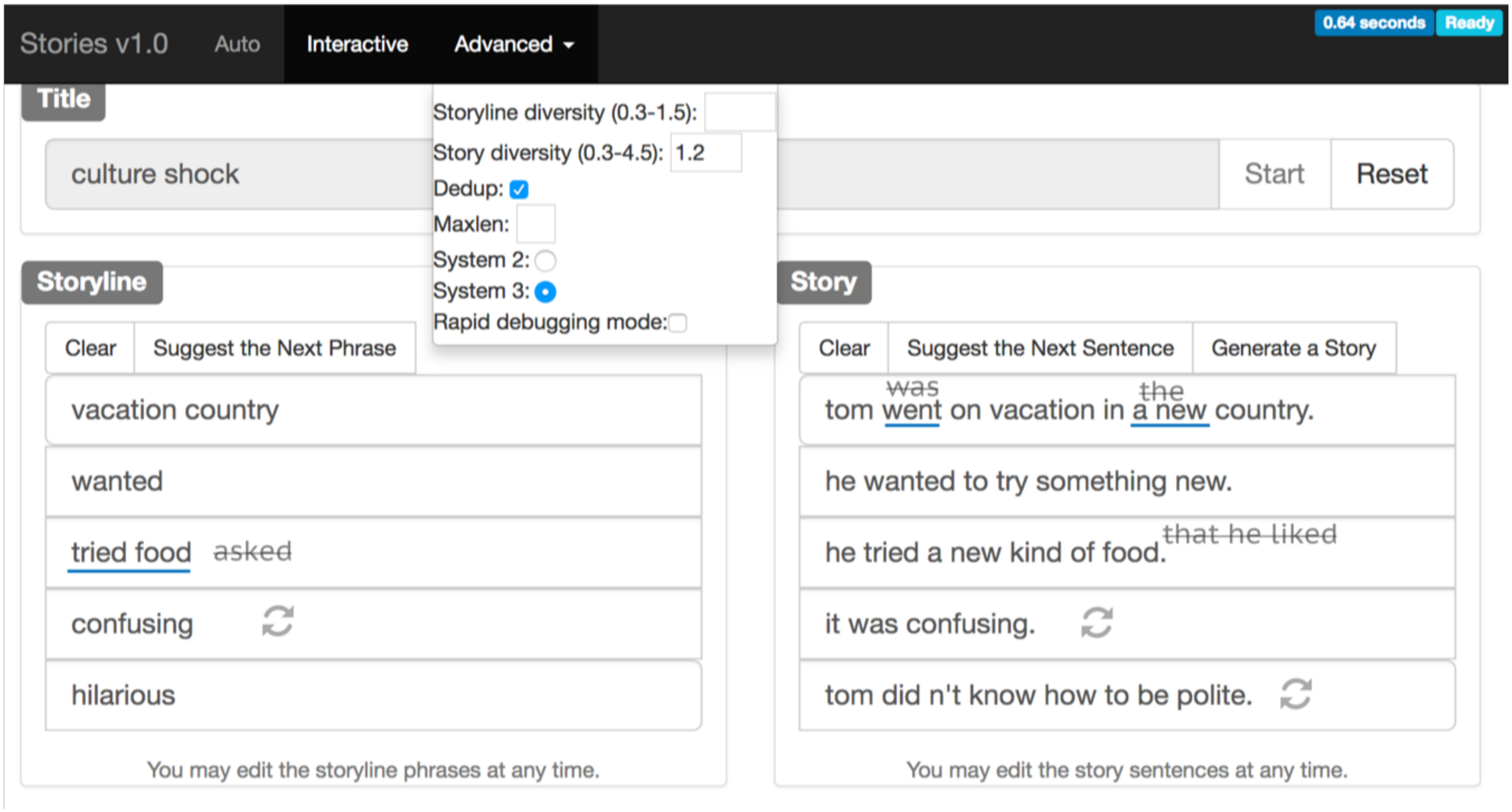


Peng et. al. (NAACL 2018, Storytelling Workshop).

Generation Samples

Computer generated	**Human flip the *EndingSpirit**
*HappyEnding: i was playing soccer with my friends. we were playing in the park. i was a bit nervous. i never played. *i ended up winning the game.*	*HappyEnding -> *SadEnding: i was playing soccer with my friends. we were playing in the park. i was a bit nervous. i never played. *i was n't very good at it.*
*SadEnding: tom was on vacation. he was going to the beach. he had to go to the beach. he was on the beach. *he was feeling sick.*	*SadEnding -> *HappyEnding: tom was on vacation. he was going to the beach. he had to go to the beach. he was on the beach. *he was so excited to go back.*

Human-Computer Collaborative Generation



Goldfarb-Tarrant et. al. (NAACL 2019, Demo).

Event Temporal and Causal Relation

- Kelly and her friends (**e1:decided**) to have a hot dog contest. The girls (**e2:competed**) against each other.
- Relation between <e1, e2>: **before**

Han et. al. (NAACL 2019, WNU).

Outline

Generation by …

- Analyzing Story Structures
  - Plan-and-Write Hierarchical Story Generation

- Analyzing Humor in Puns
  - Pun Generation with Surprise

Generating Puns is Challenging

Creative Composition

- No large corpus of puns (poems, jokes, stories) to train a generative model.
- Even if a large pun corpus exists, learning the distribution of existing data and sampling from it will likely to just *mimic/memorize*, rather than generate truly *novel*, *creative* sentences.

He et. al. (NAACL 2019)

Surprise in Puns

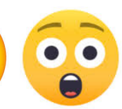
Yesterday I accidentally swallowed some food coloring.
The doctor says I'm OK, but I feel
like I've a little inside.

Local context

Global context

Pun word: dyed.

Alternative word: died

- In the local context:
  - *died a little inside.*  *dyed a little inside.*  
- In the global context: *swallowed some food coloring*
  - *dyed a little inside.*  *died a little inside.* 

Analyzing for Generation

- The surprisal principle for humor in pun.
  - Quantitative instantiate of the surprisal principle to measure funniness in pun.
  - Instantiate the principle in pun generation
    - Retrieve and edit

Prior Theories - Funniness in Puns

- Kao et. al. 2015 proposes two principles to quantify funniness in puns.
  - **Ambiguity:** The sentence has two meanings (necessary but insufficient condition)
    - the person died/dyed.
  - **Distinctiveness:** The two sentence meanings are supported by *distinct* subsets of words in the sentence.
- The incongruity theory.

Quantifying Local-Global Surprisal

- We quantify surprise using a pre-trained language model.

  Inspired by Levy, 2015

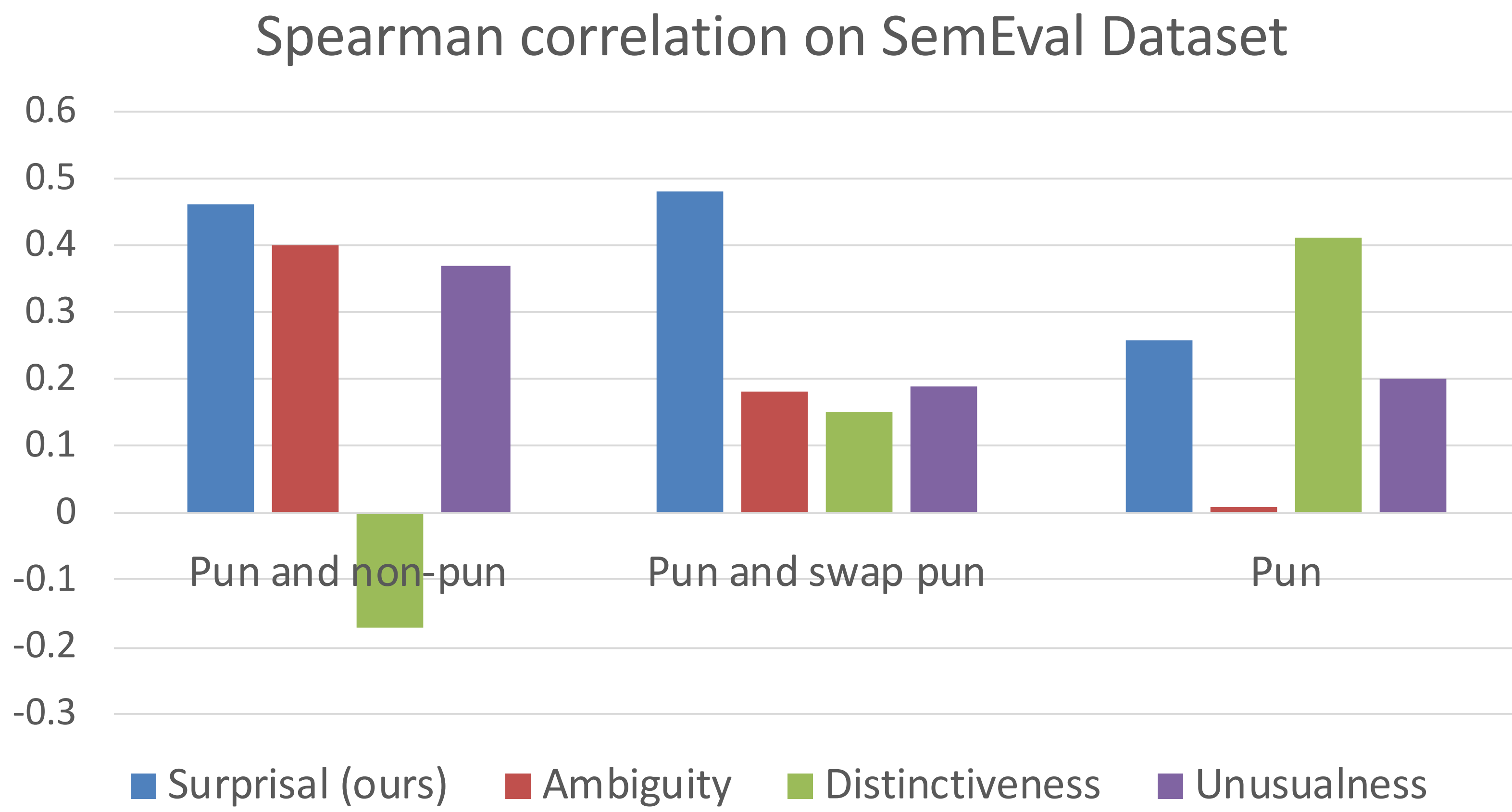
  - Local surprisal S_{local}
    - $-\log \frac{p(like\ I've\ dyed\ a\ little\ inside)}{p(like\ I've\ died\ a\ little\ inside)}$
  - Global surprisal S_{global}
    - $-\log \frac{p(Yesterday\ I\ \ldots\ like\ I've\ dyed\ a\ little\ inside.)}{p(Yesterday\ I\ \ldots\ like\ I've\ died\ a\ little\ inside.)}$
  - Local-global surprisal ratio: $S_{ratio} \stackrel{\text{def}}{=} \frac{s_{local}}{s_{global}}$ (the larger the better)

    Humor theory about incongruity resolution, Tony, 2004

Evaluating The Surprisal Principle

- Goal: compute the *correlation* between human ratings of funniness and the scores produced by our principle.
- Three types of sentences:
  - **Pun:** The magician got so mad he pulled his **hare** out.
  - **Swap pun:** The magician got so mad he pulled his **hair** out.
  - **Non-pun:** Look at that **hare**.
- Datasets
  - Derived from SemEval 2017 (130 sentences including the three cases)
  - Human rating of funniness from 1 (not at all) to 7 (extremely)

No Single Metric Works Well Across the Board



Directly optimizing the surprisal score cannot work well for generation.

A Retrieve-and-Edit Framework for Pun Generation

$w^p = \textit{hare}$

$w^a = \textit{hair}$

↓ Retrieve using *hair*

the man stopped to get a hair cut.

↓ Swap *hair* → *hare*

the man stopped to get a hare cut.

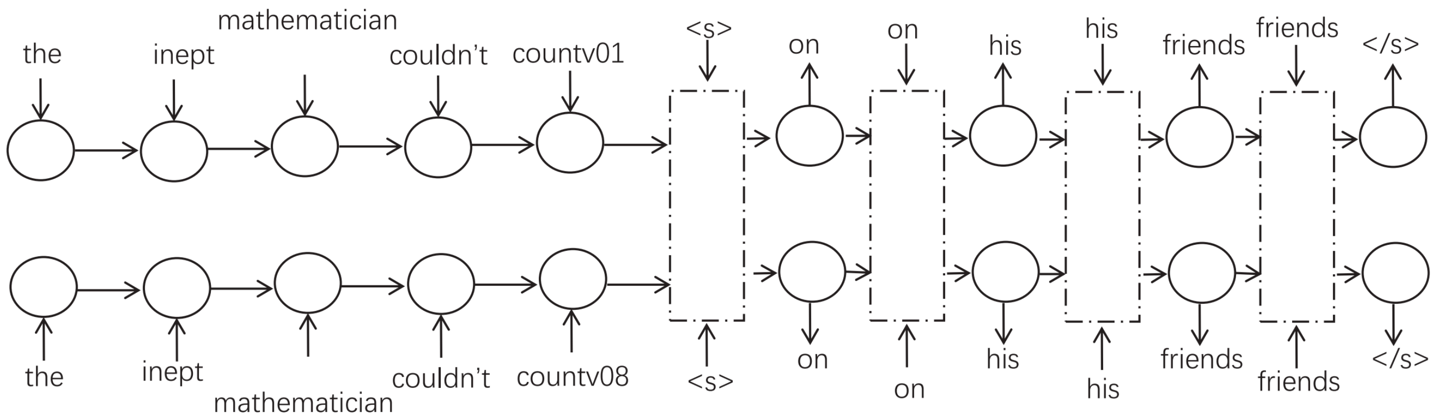
↓ Insert topic *man* → *greyhound*

the greyhound stopped to get a hare cut.

- Generating **local-surprisal**: retrieve and swap
- Generating **global-local contrast**: inserting a *topic word* at the beginning.
  - ➢ Relativeness measured by a distance skip-gram.

Baselines

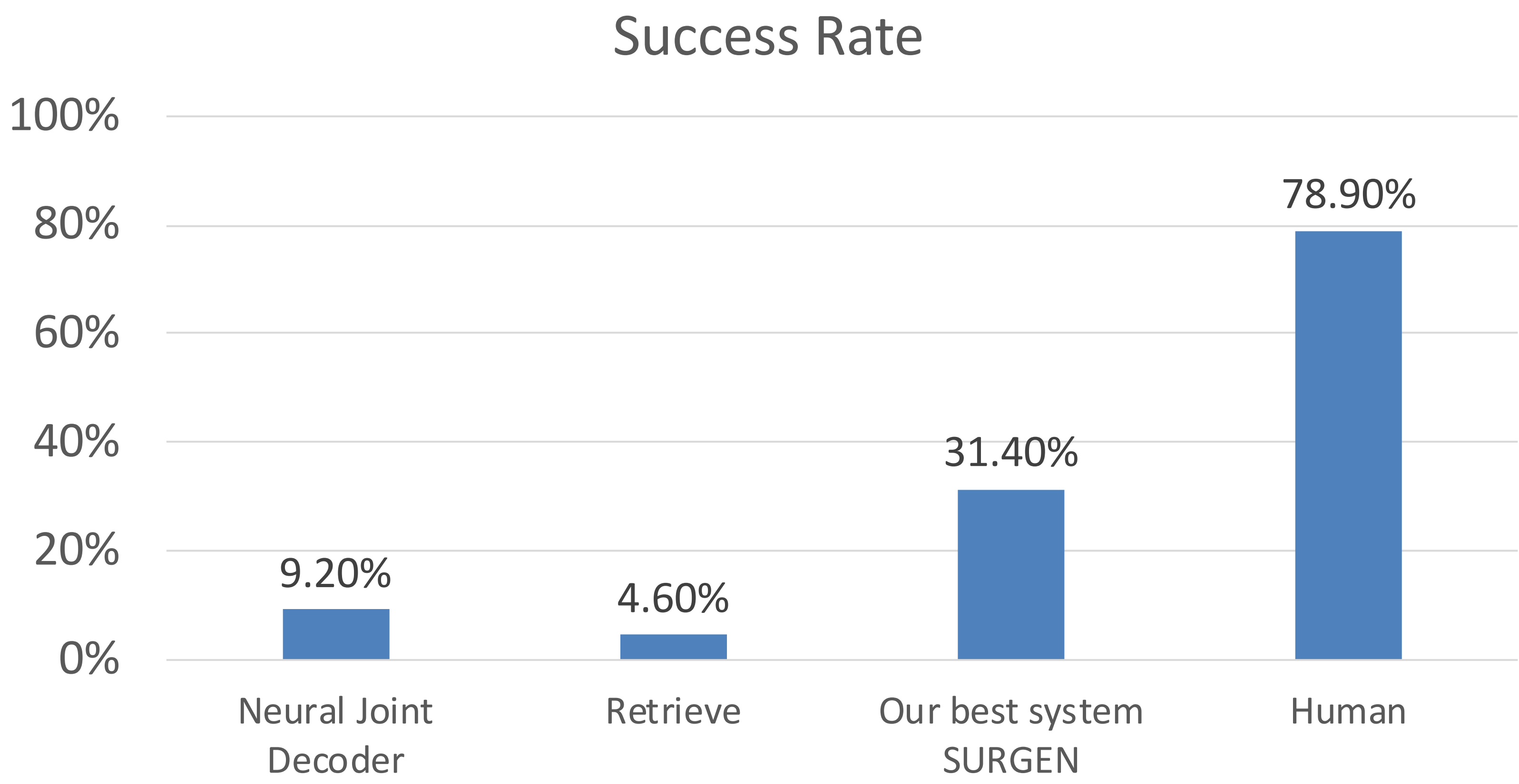
- Retrieve: just retrieve a sentence that contains the pun word.
  - Look at that **hare**.
  - I **dyed** my hair.
- Neural Joint Decoder (Yu et al., 2018)
  - An implementation to the ambiguity principle



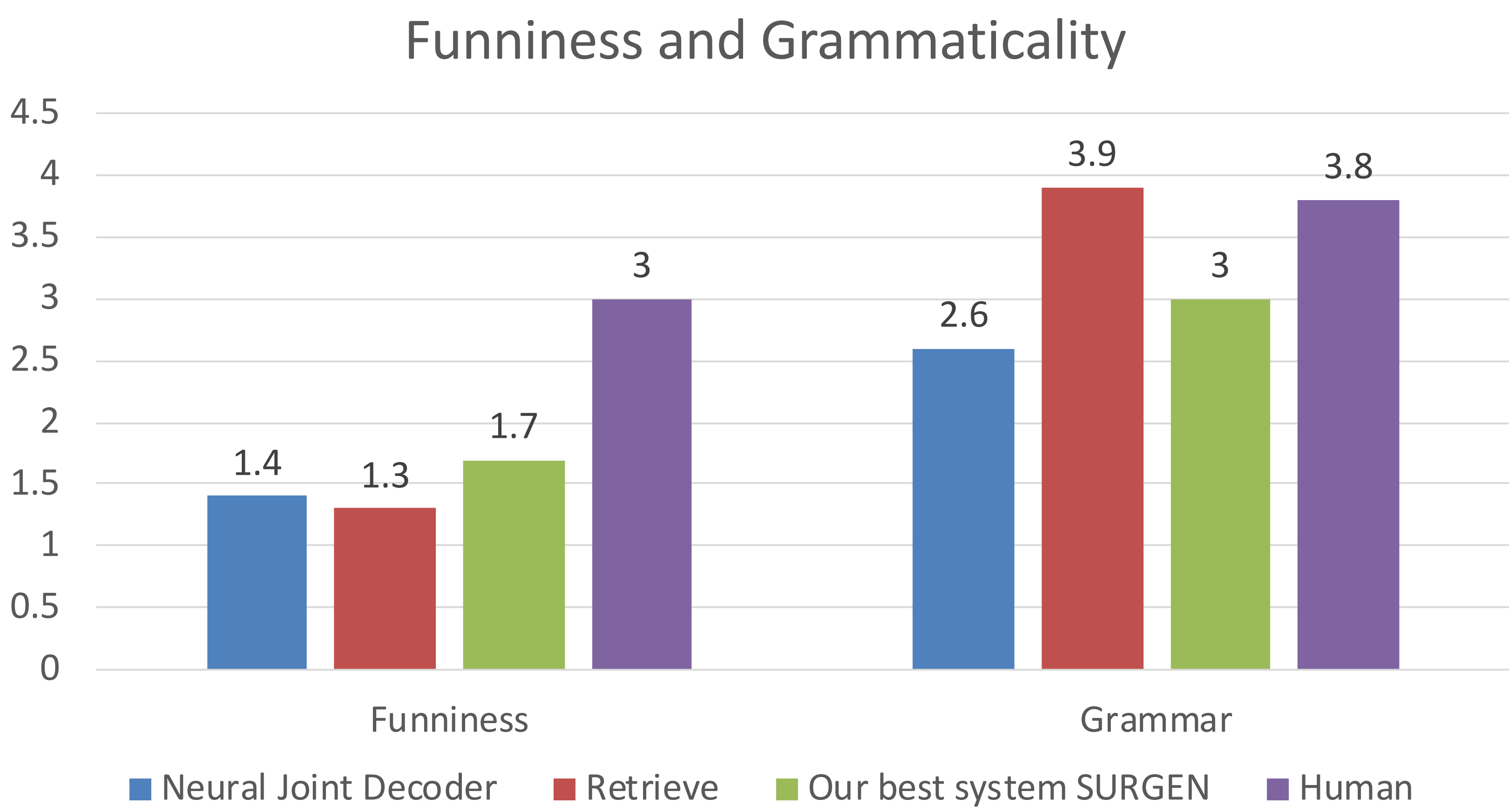
Evaluating the Generation System

- Selected 150 pun-alternative pairs.
- Each system generates puns from these words.
- Human ratings:
  - funniness (1-5),
  - grammaticality (1-5)
  - success (yes/no), given the formal definition in D. Aarons. 2017

Evaluating Generated Puns



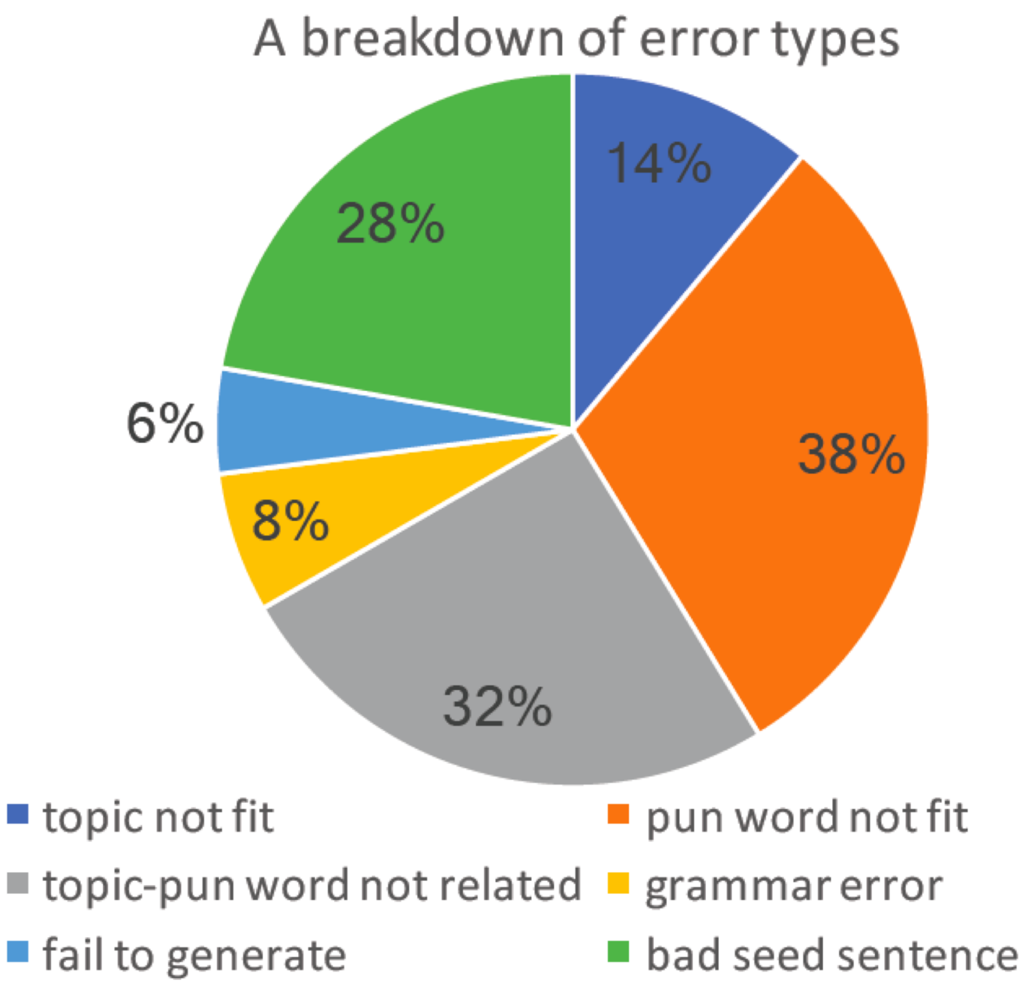
Evaluating Generated Puns



Case Study

Method	Example	Rating
Pun/alternative word pair: butter – better		
NJD	He is going to come up with the butter a ‘very good’ approach to the world’s economic crisis, the world’s biggest economic climate.	1
SURGEN	Well, gourmet did it, he thought, it’d butter be right.	2
Human	Why did the dairy churn? The less said, the butter...	1.5
Pun/alternative word pair: peace – piece		
NJD	Further, he said, at the end of the peace, it’s not clear that it will be a good example.	1
SURGEN	That’s because negotiator got my car back to me in one peace.	1.5
Human	Life is a puzzle; look here for the missing peace.	3

Error Analysis



- The main challenge is to find good seed sentences.
- The pun word needs to fit.
- Strong local association for the alternative word.

Conclusion

- Big LMs and mimicking existing materials are not likely to generate truly *novel*, *creative* sentences.
  - Structures and inductive bias are needed
- Narrative understanding (NLU) is necessary for controllable, creative generation.
- Code and demo can be found at:
  - https://violetpeng.github.io//

Thanks!
Questions?