



Microsoft®  
**Research**

“President Vows to Cut ~~Taxes~~ Hair”:

## Dataset and Analysis for Creative Text Editing for Humorous Headlines

**Nabil Hossain**

PhD Candidate  
Dept. Computer Science  
University of Rochester, NY

**John Krumm**

Microsoft Research AI  
Microsoft Corporation  
Redmond, WA

**Michael Gamon**

Microsoft Research AI  
Microsoft Corporation  
Redmond, WA

**NAACL 2019**

# A Simple Humor Task

- News Headline:
  - Trump cancels meeting with North Korea, citing “tremendous anger and open hostility”

# A Simple Humor Task

- News Headline:
  - Trump cancels meeting with **North Korea**, citing “tremendous anger and open hostility”

# A Simple Humor Task

- News Headline:
  - Trump cancels meeting with ~~North Korea~~ wife, citing “tremendous anger and open hostility”

# A Simple Humor Task

- News Headline:
  - Trump cancels meeting with ~~North Korea~~ wife, citing “tremendous anger and open hostility”
  - Hillary Clinton staffers considered campaign slogan “Because it’s her turn”

# A Simple Humor Task

- News Headline:
  - Trump cancels meeting with **North-Korea** wife, citing “tremendous anger and open hostility”
  - Hillary Clinton staffers considered campaign slogan “Because it’s her **turn**”

# A Simple Humor Task

- News Headline:
  - Trump cancels meeting with **North-Korea wife**, citing “tremendous anger and open hostility”
  - Hillary Clinton staffers considered campaign slogan “Because it’s her **turn fault**”

# A Simple Humor Task

- News Headline:
  - Trump cancels meeting with ~~North Korea~~ wife, citing “tremendous anger and open hostility”
  - Hillary Clinton staffers considered campaign slogan “Because it’s her ~~turn~~ fault”
- **Task:** Apply a micro-edit in a news headline to generate humor

**Micro-edit:**

<u>Replaced</u>		<u>Replacement</u>
entity	→	noun
noun	→	noun
verb	→	verb



# A Simple Humor Task

- News Headline:
  - Trump cancels meeting with **North Korea** wife, citing “tremendous anger and open hostility”
  - Hillary Clinton staffers considered campaign slogan “Because it’s her **turn** fault”
- **Task:** Apply a micro-edit in a news headline to generate humor

**Micro-edit:**

<u>Replaced</u>		<u>Replacement</u>
entity	→	noun
noun	→	noun
verb	→	verb

- Headlines ripe for humor
- Micro-edits focus on atomic changes

# Contributions

- **Humicroedit:** 15k edited news headlines
  - 5k headlines x 3 edits per headline
  - Each graded by 5 judges
  - **Categories:** World News, Politics
  - **Time:** Jan 2017 - May 2018
- Analysis of humor generation techniques
- Baseline LSTM funny headline classifiers and analysis

<u>Replaced</u>		<u>Replacement</u>
entity	→	noun
noun	→	noun
verb	→	verb

# Examples

ID	Original Headline	Substitute
1	Kushner to visit <b>Mexico</b> following latest Trump tirades	therapist
2	Trump wants you to take his <b>tweets</b> seriously. His aides don't	hair
3	Essential Politics: California's hottest congressional <b>races</b> , ranked	mistresses
4	Hillary Clinton Staffers Considered Campaign Slogan 'Because It's Her <b>Turn</b> '	fault
5	Trump Vows North Korea Could be Met With 'Fire and <b>Fury</b> '	marshmallows
6	Here's how <b>Wall</b> Street is reacting to Trump's tax plan	sesame
7	Swedish prosecutor says <b>truck</b> attack suspect has not spoken	mime
8	<b>Steve Bannon</b> questioned by special counsel	kindergarteners
9	New survey shows majority of US troops has 'unfavorable' view of Obama's <b>years</b>	ears
10	The Latest: BBC cuts <b>ties</b> with Myanmar TV station	pies
11	Bill Maher: "I <b>doubt</b> that Trump will be president the full term"	hope
12	Malawi arrests 140 in clampdown after 'vampirism' <b>killings</b>	rumors
13	<b>Rising</b> Dem star announces engagement to same-sex partner	gay
14	<b>Taylor Swift</b> claims Denver DJ sexually assaulted her back in 2013	hen
15	4 <b>soldiers</b> killed in Nagorno-Karabakh fighting: Officials	rabbis

# Annotation

# Qualifying Judges and Editors

**Orig:** EU says **summit** with Turkey provides no answers to concerns

**Edit:** EU says **gravy** with Turkey provides no answers to concerns

0 (Not Funny)

1 (Slightly Funny)

2 (Moderately Funny)

3 (Funny)

[Hossain et al. 2017. Filling the Blanks for Mad Libs Humor]

# Qualifying Judges and Editors

**Orig:** EU says **summit** with Turkey provides no answers to concerns

**Edit:** EU says **gravy** with Turkey provides no answers to concerns

0 (Not Funny)

1 (Slightly Funny)

2 (Moderately Funny)

3 (Funny)

- Ground truth: internal judges grade 20 edited headlines
- Qualify Mechanical Turk judges based on grade distance from ground truth

[Hossain et al. 2017. Filling the Blanks for Mad Libs Humor]

# Qualifying Judges and Editors

**Orig:** EU says **summit** with Turkey provides no answers to concerns

**Edit:** EU says **gravy** with Turkey provides no answers to concerns

0 (Not Funny)

1 (Slightly Funny)

2 (Moderately Funny)

3 (Funny)

- Ground truth: internal judges grade 20 edited headlines
- Qualify Mechanical Turk judges based on grade distance from ground truth

**ORIGINAL:** Eric Trump: Those Who **Oppose** My Dad Are ' Not Even People '

**EDITED:** Eric Trump: Those Who **support** My Dad Are ' Not Even People '

Substitute:

[Hossain et al. 2017. Filling the Blanks for Mad Libs Humor]

# Qualifying Judges and Editors

**Orig:** EU says **summit** with Turkey provides no answers to concerns

**Edit:** EU says **gravy** with Turkey provides no answers to concerns

0 (Not Funny)

1 (Slightly Funny)

2 (Moderately Funny)

3 (Funny)

- Ground truth: internal judges grade 20 edited headlines
- Qualify Mechanical Turk judges based on grade distance from ground truth

**ORIGINAL:** Eric Trump: Those Who **Oppose** My Dad Are ' Not Even People '

**EDITED:** Eric Trump: Those Who **support** My Dad Are ' Not Even People '

Substitute:

- Candidate editors micro-edit sets of 10 headlines to make them funny
- Qualify editors based on mean funniness grade from judges

[Hossain et al. 2017. Filling the Blanks for Mad Libs Humor]



# Quality Control

- Avoid exhaustion: launch mini-batches, at least 24 hours apart
- Diversity

# Quality Control

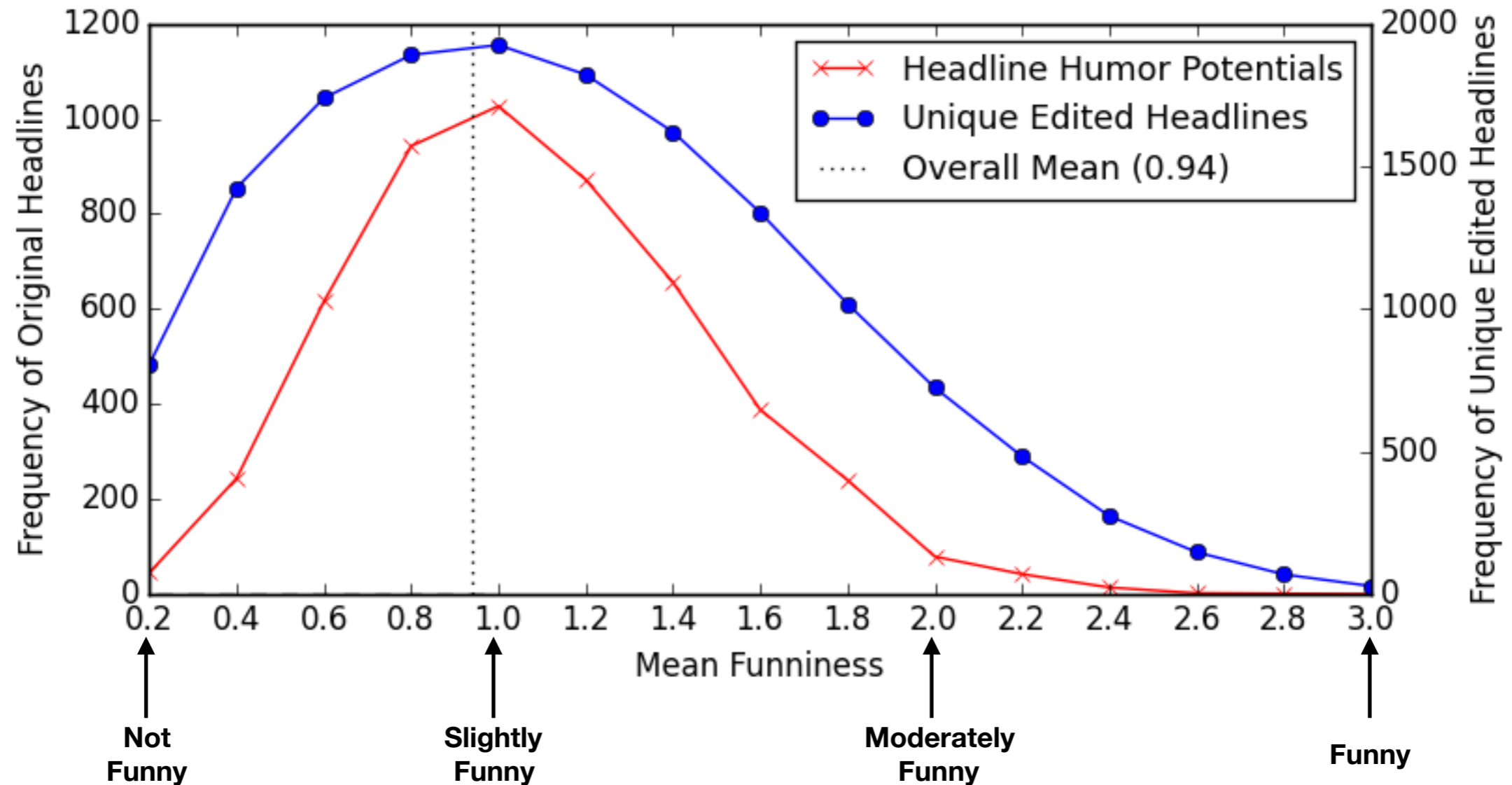
- Avoid exhaustion: launch mini-batches, at least 24 hours apart
- Diversity
  - Sample subset of annotators per batch
  - Apply annotator max HIT limit per batch
  - Temporarily suspend annotators who have done many HITs

# Quality Control

- Avoid exhaustion: launch mini-batches, at least 24 hours apart
- Diversity
  - Sample subset of annotators per batch
  - Apply annotator max HIT limit per batch
  - Temporarily suspend annotators who have done many HITs
- Get quality humor: remove editors if historical funniness  $< 0.7$
- Manual correction: fix spelling mistakes, remove multi-words

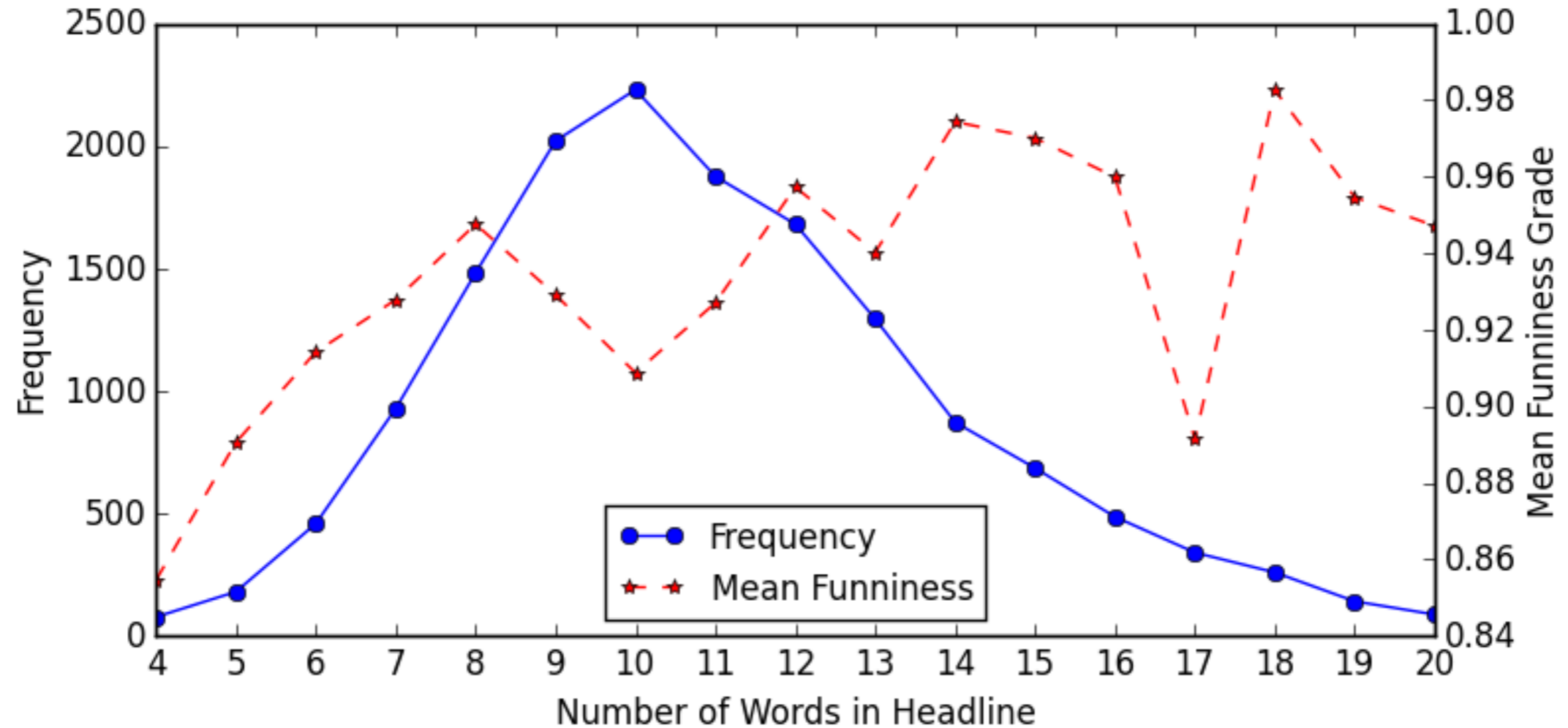
# Humor Analysis

# Funniness Distribution



- **Blue:** Histogram of mean funniness of each distinct edited headline
- **Red:** Humor *potential* of a headline (mean funniness of it's edits)

# Word Length



- Short headlines harder to make funny
- Headlines with higher proportion of replaceable words were funnier

# Humor Generation Strategies

## Mean Grades:

0 - Not Funny

1 - Slightly Funny

2 - Moderately Funny

3 - Funny

ID	Original Headline	Substitute	Grade
1	Kushner to visit <b>Mexico</b> following latest Trump tirades	therapist	2.8
2	Trump wants you to take his <b>tweets</b> seriously. His aides don't	hair	2.8
3	Essential Politics: California's hottest congressional <b>races</b> , ranked	mistresses	2.8
4	Hillary Clinton Staffers Considered Campaign Slogan 'Because It's Her <b>Turn</b> '	fault	2.8
5	Trump Vows North Korea Could be Met With 'Fire and <b>Fury</b> '	marshmallows	2.6
6	Here's how <b>Wall</b> Street is reacting to Trump's tax plan	sesame	2.4
7	Swedish prosecutor says <b>truck</b> attack suspect has not spoken	mime	2.4
8	<b>Steve Bannon</b> questioned by special counsel	kindergarteners	2.4
9	New survey shows majority of US troops has 'unfavorable' view of Obama's <b>years</b>	ears	2.2
10	The Latest: BBC cuts <b>ties</b> with Myanmar TV station	pies	1.8
11	Bill Maher: "I <b>doubt</b> that Trump will be president the full term"	hope	0.2
12	Malawi arrests 140 in clampdown after 'vampirism' <b>killings</b>	rumors	0.2
13	<b>Rising</b> Dem star announces engagement to same-sex partner	gay	0.0
14	<b>Taylor Swift</b> claims Denver DJ sexually assaulted her back in 2013	hen	0.0
15	4 <b>soldiers</b> killed in Nagorno-Karabakh fighting: Officials	rabbits	0.0

# Clusters of Replacement Words

---

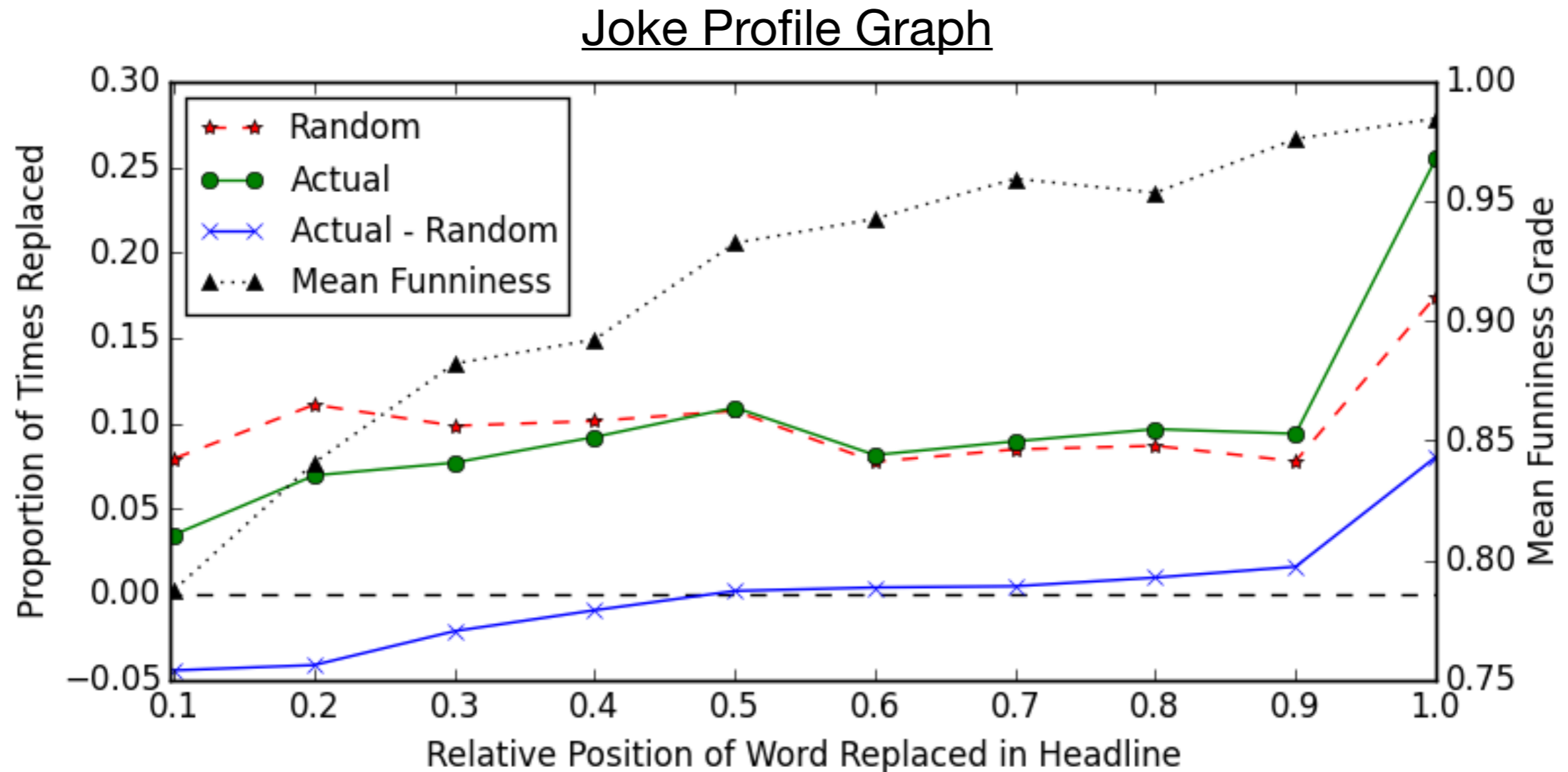
Class Label	Funniness	Frequency %	5 Frequent Sample Words
clothes and fashion	1.213	5.77	hair, pants, haircut, fashion, underwear
sex	1.209	3.52	orgy, spank, mistress, porn, striptease
food (savory)	1.113	3.89	cheese, sandwich, chicken, potato, tacos
eating	1.048	4.18	food, pizza, dinner, restaurant, eat
fantasy characters	1.036	4.94	clown, aliens, penguin, robot, ghost
music and shows	1.023	3.85	dance, circus, music, sings, mime
bodily functions	1.021	3.11	diet, brain, odor, dandruff, pimple
food (snacks)	1.017	5.36	pumpkin, cake, vodka, cookies, candy
animals	1.000	9.22	dog, monkey, puppy, cats, duck
various nouns (1)	0.935	5.82	toupee, hoedown, jaywalking, barbers, seance
jobs and roles	0.890	4.95	children, wife, baby, mother, barber
insults	0.876	3.36	clowns, tantrum, trolls, racist, whining
emotional	0.868	7.74	love, hug, jokes, fights, cry
leisure	0.862	6.01	vacation, shopping, tanning, hotel, pool
sports	0.856	3.03	horse, game, bowling, golf, wrestling
various nouns (2)	0.850	4.90	water, nose, balloon, gas, smoke
human deficiencies	0.803	4.21	lies, ignorance, humor, boredom, stupidity
media	0.787	4.20	tweet, movie, book, video, television
aspirations	0.726	7.28	party, date, people, money, model
corrupted	0.712	4.48	president, bribes, politicians, destroy, prison

---

- Words clustered based on GloVe Word Embeddings (300d)

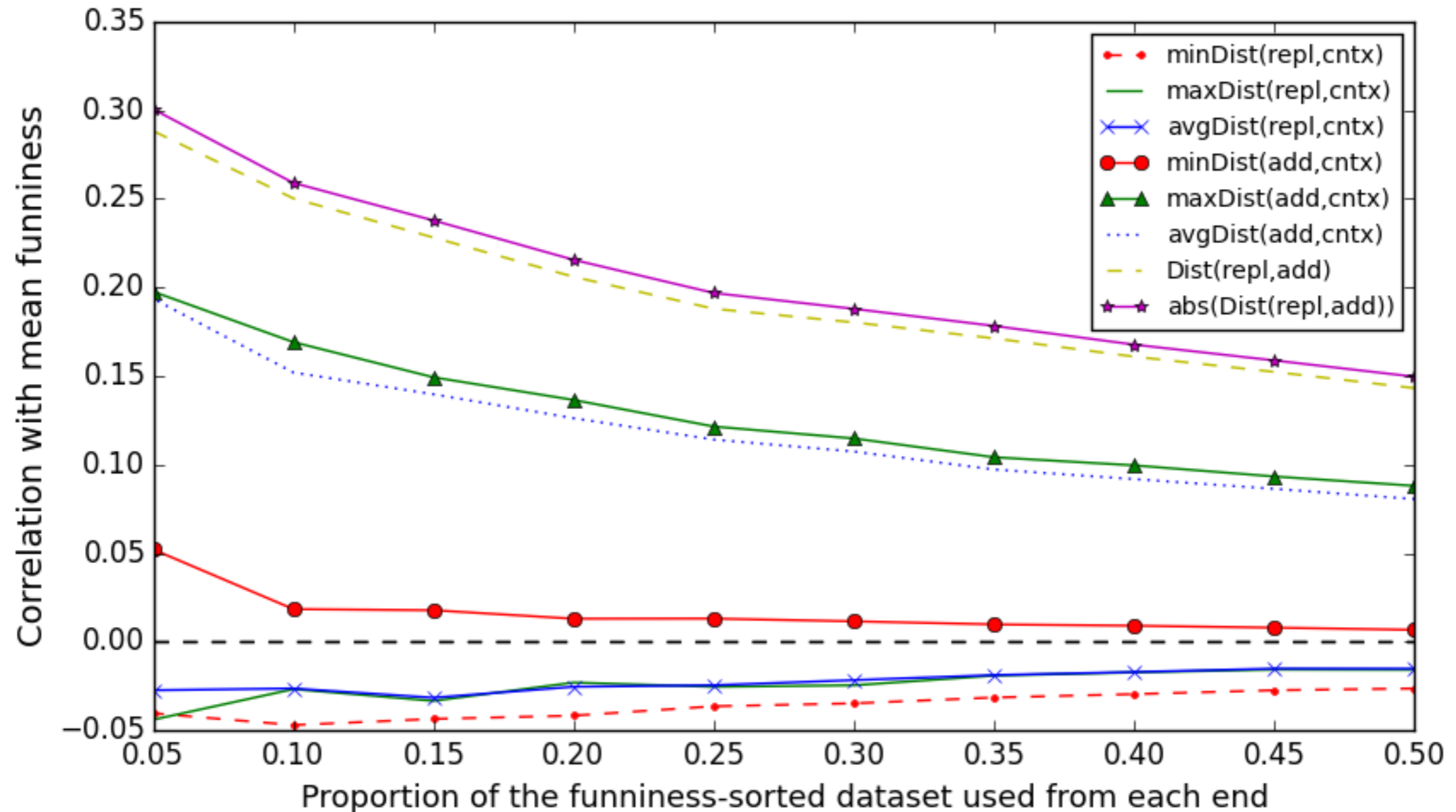


# Setup and Punchline



- Turkers generally chose to replace more words towards the end of headlines
- More humor achieved when words are removed towards the end of headlines

# Incongruity



- Higher word vector distance between the added word and the headline's words implies more humor
- Incongruity effect stronger at the extremes of the dataset

# Humor Classification

# Humorous Headline Classification

- Use only the edited headlines
- Select top X% most funny and least funny headlines into 2 classes
  - 64% train, 16% dev, 20% test

# Humorous Headline Classification

- Use only the edited headlines
- Select top X% most funny and least funny headlines into 2 classes
  - 64% train, 16% dev, 20% test

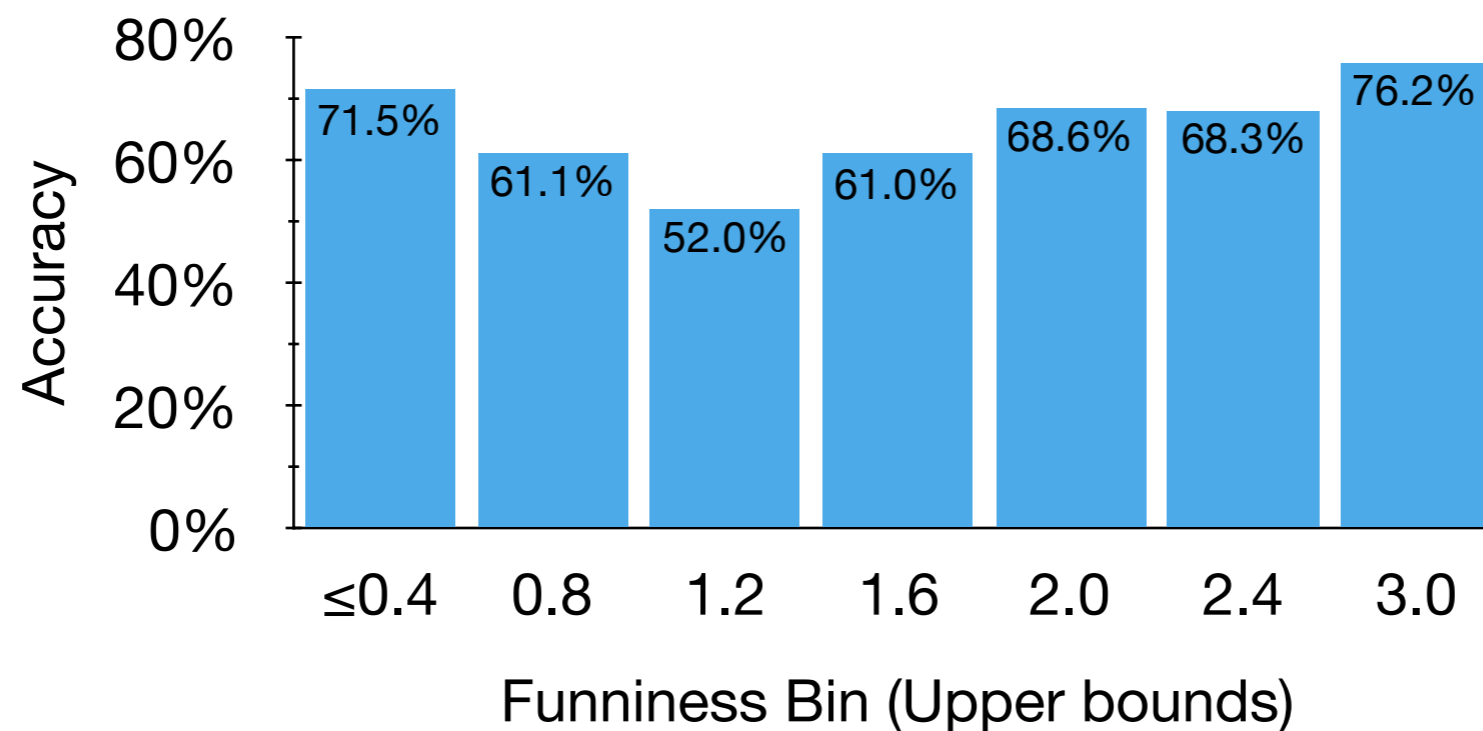
$X$	MaxUF	MinF	$\alpha$	Rf-Glv	Rf-ngram	Lstm-Glv
10	0.2	1.8	0.66	60.27	65.56	68.54
20	0.4	1.4	0.49	61.67	63.41	67.21
30	0.6	1.2	0.37	59.43	62.96	66.11
40	0.8	1.0	0.27	57.45	59.35	64.07
50	0.8	0.8	0.20	55.63	56.52	60.63

- MaxUF = max grade for not-funny class
- MinF = min grade for funny class
- $\alpha$  = agreement (Krippendorff)
- Rf-Glv: 8 GloVe distance features

# LSTM Classification Analysis

$X$	MaxUF	MinF	$\alpha$	Rf-Glv	Rf-ngram	Lstm-Glv
40	0.8	1.0	0.27	57.45	59.35	64.07

- Classification accuracy on distinct mean-grade bins:



- Lower accuracy when entity replaced (61.8%) compared to when noun (64.5%) or verb (65.5%) replaced

# Qualitative LSTM Analysis

ID	Original Headline	Substitute	Grade	Prediction
1	Kushner to visit <b>Mexico</b> following latest Trump tirades	therapist	2.8	✗
2	Trump wants you to take his <b>tweets</b> seriously. His aides don't	hair	2.8	✓
3	Essential Politics: California's hottest congressional <b>races</b> , ranked	mistresses	2.8	✗
4	Hillary Clinton Staffers Considered Campaign Slogan 'Because It's Her <b>Turn</b> '	fault	2.8	✗
5	Trump Vows North Korea Could be Met With 'Fire and <b>Fury</b> '	marshmallows	2.6	✓
6	Here's how <b>Wall</b> Street is reacting to Trump's tax plan	sesame	2.4	✓
7	Swedish prosecutor says <b>truck</b> attack suspect has not spoken	mime	2.4	✗
8	<b>Steve Bannon</b> questioned by special counsel	kindergarteners	2.4	✗
9	New survey shows majority of US troops has 'unfavorable' view of Obama's <b>years</b>	ears	2.2	✗
10	The Latest: BBC cuts <b>ties</b> with Myanmar TV station	pies	1.8	N/A (train)
11	Bill Maher: "I <b>doubt</b> that Trump will be president the full term"	hope	0.2	✗
12	Malawi arrests 140 in clampdown after 'vampirism' <b>killings</b>	rumors	0.2	✗
13	<b>Rising</b> Dem star announces engagement to same-sex partner	gay	0.0	✗
14	<b>Taylor Swift</b> claims Denver DJ sexually assaulted her back in 2013	hen	0.0	✗
15	4 <b>soldiers</b> killed in Nagorno-Karabakh fighting: Officials	rabbits	0.0	✓

# SemEval-2020 Task Launched

- Sub-tasks:
  - Estimate Headline Funniness (RMSE)
  - Predict “Funnier of the Two” edit for the same original headline (Accuracy)
- Task (#7): [competitions.codalab.org/competitions/23212](https://competitions.codalab.org/competitions/23212)
- Dataset: [cs.rochester.edu/u/nhossain/humicroedit.html](https://cs.rochester.edu/u/nhossain/humicroedit.html)
- Quickstart: [github.com/n-hossain/semEval-2020-task-7-humicroedit](https://github.com/n-hossain/semEval-2020-task-7-humicroedit)
- SemEval: [alt.qcri.org/semEval2020](https://alt.qcri.org/semEval2020)
- Participate!

Checkpoints	Important Dates
Train; Dev w/o labels	released June 1, 2019
Dev labels; Test w/o labels	release on Dec 3, 2019
Test Set Evaluations	Jan 10 - Jan 31, 2020