Stand-up Comedy: Quantifying Humor and Identity

Johnny Ma

January 27, 2020

Laughter Detection and Punchline Classification

1.1 Introduction

It's just me, the mic, and the audience.

— Richard Pyror, *Pyror Convictions*

Being a stand-up comedian is no joke. Very few can, with mere words, fill an arena or club basement with laughter for over an hour. Unlike TV or movie comedy with its canned laughter or studio editing, comedians kill or bomb in real time. Being funny is merely a prerequisite; you must be a charismatic and consistent writer, orator, and stage performer. In this highly personal environment, it is no surprise that identity plays a major role in not only what jokes are told, but what jokes are considered funny.

We can treat a stand-up comedian's routine as a multi-modal experience involving audio, visual, and textual cues that seek to entice an audience laughter response. There have been several attempts at computational humor detection from one-liners [20], puns [29], double-entendres [17], TV scripts [6], TV dialogue [23], and one-off jokes in TED talks [9]. As the primary goal of stand-up comedy is to elicit laughter in a live setting, the goals of a stand-up routine is most similar to that of an persuasive speech [5]. Just like campaign speeches, a stand-up routine first takes form as a long written collection of bits and jokes containing lexical and rhetorical strategies in the form of set-up lines and punchlines. The routines are then performed in a large social setting, where comedians employ acoustic techniques to signal and elicit a laughter response from an expectant audience.

In this paper, we computationally examine the textual, rhetorical, and audio markers that go into making a joke. We build a dataset of contemporary stand-up routines, identify punchlines with a laughter detection algorithm, then train a Logistic Regression punchline classifier on a number of these markers. We then use these markers to quantitatively examine an socio-cultural aspect integral to stand-up comedy: gender identity.

1.2 Data

1.2.1 Corpus Introduction

In recent years, Netflix has established itself as a tent-pole figure in stand-up comedy by aggressively signing million dollar contracts for exclusive streaming rights of live comedy performances [31]. While these hour long "comedy specials" acquisitions are mostly from some of the industry's biggest names, Netflix has diversified its stand-up portfolio by hosting routines from up-and-coming talent of many backgrounds. Streamed routines are far from "the trenches" of the live club performance, as they are a highly edited and curated experiences [28]. However, since they are distributed, subtitled, and marketed towards Netflix's global audience, we believe that they not only represent a wide pool of contemporary comedy talent, but also are geared towards a more homogeneous and generic "audience than would be found in a New York, Chicago, or Los Angeles club.

1.2.2 Corpus Acquisition

We focus exclusively on stand-up routines in the standard "special" format; that is, approximately 1-hour long, filmed in front of a live-audience, and primarily spoken in the English language. We record the audio from these performances using the following methodology:

1. Steam the special in Google Chrome from a registered Netflix account.
2. Use the Soundflower [1] open source software to capture the stream's 2ch audio output.
3. Route and record the audio steam into a 44100 Hz .wav format using Audacity (ver 2.3.2) [2].

This yielded 126 hours of comedy from 116 routines performed by 104 comedians, released globally on Netflix between the dates of August 16, 2012 and September 10, 2019. We extracted the subtitles for each routine by feeding an .xml taken from the Netflix data stream into a "Netflix-to-srt" python script [3]. This gave us time-stamped lines that are roughly aligned with the routine's spoken audio. This gives us an average of 1115.21 lines and 4562.11 lemmatized words per routine.

1.2.3 Laughter Detection

To detect laughter from our .wav audio tracks, we use Jon Gilick's laughter detection algorithm[4]. The algorithm was created for the paper "Capturing, Representing, and Interacting with Laughter" [?] and uses 3 layer feed-forward neural network trained on standard Mel-frequency cepstral coefficients (MFCC) and delta MFCC features from 10ms audio frames of the Switchboard dataset [13], which contains 260 hours of speech. The detection algorithm achieves a 88% per-frame accuracy at identifying laughter in an held out validation test. We use this same model to detect laughter for our audio data. Taking advantage of the musical quality of laughter, we pre-process our data with an audio source separator algorithm[5] from Chandra *et al.* (2017) [8], roughly separating the

[1] `https://github.com/mattingalls/Soundflower/releases`

[2] Audacity® software is copyright © 1999-2019 Audacity Team. The name Audacity® is a registered trademark of Dominic Mazzoni.

[3] `https://github.com/isaacbernat/netflix-to-srt`

[4] `https://github.com/jrgillick/laughter-detection.thresholdanddurationwerebothsetto0.2.Differentvaluesproducedsimilarresults.`

[5] `https://github.com/MTG/DeepConvSep`

crowd's laughter from comedian utterances. We then applied the laughter detection to the routine's music audio, with post-processing from a low-pass Butterworth from the sci-py package. This process returns a list of start and stop times for every instance of detected laughter in each routine.

1.2.4 Alignment and Validation

We aligned each laughter instance with its corresponding subtitle line text by comparing time stamps. If the laughter clip's start time is within 0.5 seconds of the end time of a subtitle line, we associate that laughter clip with that line. For each subtitle line, we measure the duration of all associated laughter clips, classifying that line as a punchline if the duration is non-zero and as a set-up line if the there is no associated laughter. We are able to assess the recall of the detection algorithm by using moments when audio descriptions such as [laughter] or [crowd laughs] appear in the subtitle text. For the 8595 such captions across 94 routines, 95.29% of the lines before [laughter] are correctly marked as punchlines. It is important to note that these instances of laughter are captioned because they are uninterrupted, and despite the algorithm's effectiveness there are likely several false positives. After this test, we removed all hypertext and unicode, all non-dialogue subtitles such as [laughter] or [with heavy accent], and all musical intros and outros from each routine in our corpus. Table 1 provides a number of summary statistics across our final corpus.

1.3 Punchline Classifier Features

To validate our dataset, we replicate the task of classifying punchlines and set-up lines from computational humor literature. We incorporate a number of lexical and audio features from previous literature. We calculate the following features for our dataset on the line level.

1.3.1 Euphony

Chimps in the dirt playing with sticks." What makes that joke is that out of seven words, four of them are funny. - Jerry Seinfeld

Guerini *et al.* (2015) [14] test the notion that certain words have an aesthetic quality that naturally produce a response. They use the CMU English pronunciation dictionary [6] to break down words into phonemes, which are used to calculate four scores by subtitle line. The *alliteration* and *rhyme* score are the number of repeated phoneme sounds at the beginning and end of each word respectively. The *plosive* score is the count of words starting with plosive sounds, defined as 'P', 'T', 'K', 'B', 'D', and 'G'. Finally, the *homogeneity* score is the count of unique phoneme sounds within the utterance. All scores are normalized by the total number of phonemes in the utterance, with all scores $s \in [0, 1]$.

[6] `http://www.speech.cs.cmu.edu/cgi-bin/cmudict`

1.3.2 Lexical Structure

Bertero and Fung (2016) show the success of various lexical features for predicting humorous lines from TV dialogue. For each subtitle line, we calculate a number of their high level features, namely average word length, part-of-speech proportion (tagged by spaCy[7]), and average positive, negative, and neutral sentiment (from SentiWordNet[8]).

1.3.3 Comedic Devices

Costello: Look, all I wanna know is when you sign up the first baseman, how does he sign his name? Abbott: Who. Costello: The guy. Abbott: Who. - Abbott and Costello, Who's on First?[9]

As the great success of Abbott and Costello and other "straight man - funny man" comedy double acts shows, linguistic incongruity and opposition is effective in eliciting comedic response. Similarly, puns and wordplay rely on linguistic ambiguity to produce a humorous doubling effect. Yang *et al.* (2015) quantify the latent structure of these comedic devices. They define semantic incongruity with *disconnection*, the maximum meaning distance of word pairs in an utterance, and *repetition*, the minimum meaning distance derived from pre-trained Word2Vec[10] word embeddings. *Ambiguity* uses WordNet's path similarity function to [22] calculate ambiguity with sense farmost and sense closest, the largest and smallest path similarity between words in a sentence.

1.3.4 Acoustic Performance

"I talk very quietly in a monotone voice where there's almost zero performance in there, to see if the material holds up," - Ali Wong[11]

Some say that comedy is all timing and delivery. Litman *et al.*(2006) use acoustic software to capture summary statistics for punchline detection in a TV situational comedy context. We also calculate the mean, max, min, range, and standard deviation pitch (F0) and energy (RMS), as well as internal silence and tempo (syllables per second) using Librosa [19] and REAPER[12]. Our unit of analysis is the audio clipped with each subtitle start and stop timestamp.

1.3.5 Bi-grams

Bamman *et al.*(2018) find evidence that bi-grams capture stock phrases that often occur across speakers in a similar performance context. Similarly, we capture all bi-grams that occur more than 25 times in our entire corpus, yielding a sparse matrix of 4421 bi-gram columns (e.g., 1 if "please_welcome" appears in a line, 0 otherwise) for each routine.

[7] `https://spacy.io`
[8] `https://github.com/aesuli/sentiwordnet`
[9] `https://www.youtube.com/watch?v=kTcRRaXV-fg`
[10] `5https://code.google.com/p/word2vec/`
[11] `https://www.nytimes.com/2019/10/09/books/ali-wong-dear-girls-interview.html`
[12] `https://github.com/google/REAPER`

1.4 Punchline Classifier

Using these features, we train a punchline classifier on lines from our inter-comedian corpus. For each routine, we randomly pair a punchline and set-up line together without replacement until either all punchlines or all set-up lines are paired. This gives us balanced dataset of 45,000 punchlines and 45,000 set-up lines drawn from 116 routines. For our classifier we use logistic regression with l_2 regularization and hyper-parameters chosen through 10-fold cross-validation, a placeholder for leave one out cross-validation due to limited access to computing resources. We test each lexical and acoustic feature separately and in combination. The average scores of our system are displayed in Table 2.

Punchline Classifier	**Acc.**	**Prec.**	**Recall**	**F1**	**Top Feats**
Guerini	0.57	0.56	0.57	0.57	Homogeneity, Rhyme
Bertero & Fung	0.50	0.50	0.54	0.52	avg_neg, avg_pos
Yang	0.56	0.55	0.68	0.61	Repetition, Disconnection
Litman Acoustic	0.55	0.53	0.71	0.61	max_pitch, max_energy
Bigrams >25	0.53	0.53	0.48	0.50	just_fucking, holy_shit
Combined	0.61	0.60	0.64	0.62	Repetition, max_energy
Combined + Bigrams	0.61	0.69	0.64	0.62	N/A

Table 2: Logistic Regression Punchline Classification Results

1.4.1 Discussion

Almost all of our features perform slightly above baseline for the task of binary punchline classification. The comedy specific markers and delivery features perform the best, with relatively high recall scores compared to precision. Yet, the poor performance of lexical structures, bi-grams, and across the board low precision score show us that punchlines are still relatively indistinguishable from set-up lines. Notably, easily observable features, such as high pitch & energy, repetition, and rhyme, perform quite well. This lends support to the notion that stand-up comedy audiences are playing a game of coordination [2], where laughing at an line not intended to be fun is a risky action (and may even be signs of mental disorder [1]). Thus, punchlines involving more visceral features send a strong signal to audiences that the line is meant to be humorous, and audience members may laugh even if they don't "get" the joke.

To capture the punchlines that are less signaled and improve our classifier's performance, we can incorporate more contextually dependent features. As jokes in stand-up comedy are often part of narratives and chains of information, our within line calculations and bag-of-words bi-gram approach may be insufficient. To represent meaning in longer form bits, we can incorporate features from a contextual language model such as BERT, or delta features that can capture large shifts between features at line index i and $i + 1$. Despite these shortcomings, we have identified a number of lexical and acoustic features that are computationally associated with laughter responses. We use these features to inform a regression analyzing the relationship between comedian identity and audience response.

Identity and Stand-up

2.1 Introduction

Identity has always been important to performance comedy, from the crude black-face vaudeville shows of the 1800's to the borderline confessional style of the 2010's. As comedians started to make a living telling jokes in the 1950's, they realized that routines based on gimmicks, universal bits, and joke books were no longer sufficient. To protect their jokes and carve out their niche in the growing industry of comedy clubs, comedians looked inward and incorporated humor derived from their lived experience. The usage of identity has been especially profound for comedians that identify as minorities. White male comedians have largely dominated the comedy stages of America. Other races were relegated to the Borscht Belt or Chitlin' Circuit [18]. Females were excluded almost entirely through claims that "women are not funny;" claims historically and philosophically rooted in body politics [15], societal and biological norms [11], and psychoanalysis [12]. While we adopt the theoretical perspective that identity and other social categories are performed rather than fixed [7], we believe that the first impression in-person performances draw out identity based social associations even before the first joke.

While insult comedians such as Don Rickles was beloved for maligning all identities with impunity [?], the industry is under intense scrutiny in the metoo era. Bill Cosby, Louis C.K., Aziz Ansari, and Kevin Hart have all been affected by statements and actions made in their personal and professional life. A recent example is the case of Shane Gillis, who had his offer to join Saturday Night Live rescinded as a result of "offensive, hurtful, and unacceptable" jokes he made about New York's Chinatown and minority comedians on his podcast[1]. Many believe his firing was justified as he is a non-minority "punching down," while others believe that even the self-censoring of comedy is detrimental to its purpose, an evocation of Hobbes' theory of humor as "triumph over the weak" [16] and Chesterton's theory of humor as "escape into a world...not fixed horribly in an eternal appropriateness" [10] respectively.

This has prompted us to ask: what role does the relationship between comedian identity and identity-based jokes play in eliciting laughter? We empirically evaluate this problem in a contemporary setting using our dataset of features and punchline classifications from Netflix stand-up routines.

[1] [?]

2.2 Empirical Design

We seek to evaluate whether or not, all else being equal, an identity-based joke told by a comedian of the same identity class will elicit a stronger laughter response than that elicited by a comedian of a different identity class. This analysis attempts to control for joke structure and delivery by including the lexical and acoustic features from our punchline classification in our regression. We employ an empirical design similar to that of So *et al.* (2019) [27]'s analysis of the relationship between authorial race, gender, and the contextual "sociality" of biblical citations in 20^{th} century literature. We take advantage of the ability of the laughter detection algorithm to extract the duration of each instance of laughter to measure the intensity of the audience's response. We run the following regression specification:

$$DurationLaughter_{ij} = \alpha + \beta_1 * identity\, joke_i$$
$$+\beta_2 * comedian\, identity_j + \beta_3 * (identity\, joke_i \times comedian\, identity_j)$$
$$+\boldsymbol{X}^{'}\gamma_i + \boldsymbol{X}^{'}\gamma_j + \epsilon_{ij}$$

$$identity\, joke_i = \begin{cases} 1, & \text{if } i \in identity\, topic, \\ 0, & \text{otherwise.} \end{cases}$$

$$identity\, comedian_j = \begin{cases} 1, & \text{if } j \in identity\, category, \\ 0, & \text{otherwise.} \end{cases}$$

$$(identity\, joke_i \times comedian\, identity_j) = \begin{cases} 1, & \text{if } identity\, i = identity\, j, \\ 0, & \text{otherwise.} \end{cases}$$

Where i are jokes and j are comedians, with $i * j$ total number of jokes. Duration of laughter is in (s). γ_i are audio and text controls, γ_j are comedian controls. We are interested in the β_3 coefficient of the interaction term between comedian identity and identity-based joke. If this coefficient is positive, this suggests that comedians receive a greater response for being in the same identity category that a joke they tell is contextually reliant on. If we assume that our textual and acoustic features capture the quality and delivery of a joke and essentially "match" jokes irrespective of identity, then a positive coefficient means that

2.3 Data and Identity Classification

To collect the performed identity of our 104 comedians, we use a simple counting scheme on each comedian's wikipedia.org page. Inspired by Bamman (2015) [3], we scrape each comedian's wikipedia page and use regex word searches for terms related to identity. We count the number of female, male, and non-binary pronouns to classify comedians into gender categories: male, female and other. We manually add these categories for those without wikipedia pages and reviewed all comedians. This process yielded 84 male comedians, 19 female comedians, and 1 non-binary comedian who we remove from our dataset.[2]

To gather a sample of identity-based jokes, we manually tag 4,000 punchlines randomly sampled from a balanced dataset of female and male comedians. To be tagged as a "female" or "male" joke, the punchline must explicitly require knowledge of social categories, objects, and actions surrounding that gender, in the author's judgment. We use these binary gender tags to train a logistic regression classifier, with l_2 regularization and 10-fold cross-validation to maximize *precision score* on a tf-idf weighted bag-of-words representation of each line with no stop-words removed. Results of the classifier are in Table 3.

Manual Class	cases	Precision	F1	Top Words
Non-Gendered	3729	0.95	0.97	say, okay, just, did, fuck
Female	127	0.67	0.15	provide, feminism, panties, single, girls
Male	92	0.50	0.09	father, penis, begged, dick, man

Table 3: Logistic Regression Multi-Class Identity Joke Classifier.

While this approach is far from rigorous and impartial, attempts to computationally learn the gendered words and contexts for tagging, such as conditioning on author identity in Bamman *et al.* (2014) [4], topic modeling methods from Mimno *et al.* (2012) [21], or using pronouns and proper nouns from text in Underwood *et al.* (2018) [30], are met with data availability issues and additional assumptions that could confound the possible causal analysis. Despite the poor scores from our classifier, as long as the type I errors from our assignments are randomly distributed across our treatment and outcome variables, our regression results should not be significantly affected.

2.4 Regression Results

We apply our identity-based joke classifier on the 45,000 punchlines and 45,000 set-up lines from our previous exercise. The results of this classification are reported in Table 4.

[2] Because of lack of non-white comedians and few white comedians making non-white racial jokes, we restrict our analysis to gender identity.

Gender	Identity Class	Punchline %	Example Punchline
Female	female	2.28% (231)	We only allow women to post pictures when they are a work in progress right
	male	1.33% (116)	They are like they will always use the mic as their dick like this
	not-gendered	96.63% (9695)	I literally watch my Netflix special on my sister in laws login
Male	female	1.05% (366)	Tiger Woods wife is a babysitter worth a quarter of a billion fucking dollars
	male	0.86% (296)	as a father you just have this need to protect your family from dicks
	not-gendered	98.31% (44120)	You can get infamous but you can not get unfamous

Table 4: Statistics and Examples of Identity Classification Applied to Lines.

We run our logistic linear regression specification on the 1,009 punchlines that are tagged as gendered, with controls for all lexical and acoustic features (bi-grams are omitted to prevent overfitting), as well as a control for which comedian uttered each line. We report the lexical and acoustic variables that are statistically significant ($p < 0.05$) in Table 5.

	Log Duration of Laughter:
mean_pitch	-0.028^{***} (0.009)
max_pitch	0.002^{**} (0.001)
internal_silence	-1.252^{**} (0.541)
mean_energy	26.600^{***} (6.980)
min_energy	-42.377^{***} (15.142)
max_energy	3.546 (4.453)
tempo	-0.113^{*} (0.060)
female_joke	-0.163^{**} (0.078)
gender_interaction	0.358^{***} (0.135)
Observations	1,009
Comedians Controls?	Yes
Adjusted R^2	0.231
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 5: Comedian and Joke Gender Identity Interaction Regression.

Several of the significant predictors of punchlines, such as pitch, energy, and internal silence, are also statistically significant in predicting laughter duration. While the negative coefficient on "female_joke" suggests that audiences laugh more at male identity jokes than female identity jokes, we find a positive and significant coefficient on our gender interaction variable. This supports our assertion that female jokes told by females (or a male jokes told by males) elicits more audience laughter, with lexical and acoustic controls.

2.5 Discussion and Future Work

While we have found a relationship between comedian gender and gender jokes as measured by audience response, this regression does not causally distinguish channels. This relationship can be driven by two phenomena; audiences may truly reward unity between textual and performed identity, or it could simply be that women (or men) are better at writing female (or male) jokes. While we attempt to control for the latter by using lexical and acoustic controls, given that we find that they can hardly distinguish between punchlines and set-ups, these features may be insufficient to control for joke "content." A matching process, such as the conditional topic matching algorithm developed by Roberts *et al.* (2019) [25], can decrease confounding by statistically finding nearly identical gender jokes that are told by both male and female comedians. Of course, we would also need to increase the accuracy of our identity classifier through BERT contextual features or more predictive features.

While we cannot manipulate comedian gender or a comedian's routine, we may be able to gather data where audience compositions vary. Our Netflix data only collects a single performance of a stand-up routine; we can follow a comedian as they tour various cities and venues with different gender compositions. If an entirely female crowd reacts strongly to a female joke told by a female comedian, while the same joke gets few laughs from an all male audience, gender identity becomes more clearly related to audience response. After all, while comedians navigate the boundaries of societal acceptance, it is the spectator that defines the boundaries of funny. While explaining jokes may kill the humor, statistical and computational models of stand-up comedy as writing, delivery, and social performance offer much more to discretize and dissect.

Bibliography

[1] Ahmed, Aiesha, and Zachary Simmons. "Pseudobulbar affect: prevalence and management." Therapeutics and Clinical Risk Management 9 (2013): 483.

[2] Asch, Solomon E., and Harold Guetzkow. "Effects of group pressure upon the modification and distortion of judgments." Documents of gestalt psychology (1951): 222-236.

[3] Bamman, David. "People-Centric Natural Language Processing." PhD diss., PhD thesis, Carnegie Mellon University, 2015.

[4] Bamman, David, Jacob Eisenstein, and Tyler Schnoebelen. "Gender identity and lexical variation in social media." Journal of Sociolinguistics 18, no. 2 (2014): 135-160.

[5] Gillick, Jon, and David Bamman. "Please clap: Modeling applause in campaign speeches." In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), pp. 92-102. 2018.

[6] Bertero, Dario, and Pascale Fung. "A long short-term memory framework for predicting humor in dialogues." In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 130-135. 2016.

[7] Butler, Judith, and Gender Trouble. "Feminism and the Subversion of Identity." Gender trouble 3 (1990): 1-25.

[8] Chandna, Pritish, Marius Miron, Jordi Janer, and Emilia Gómez. "Monoaural audio source separation using deep convolutional neural networks." In International conference on latent variable analysis and signal separation, pp. 258-266. Springer, Cham, 2017.

[9] Chen, Lei, and Chong Min Lee. "Predicting Audience's Laughter During Presentations Using Convolutional Neural Network." In Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications, pp. 86-90. 2017.

[10] Chesterton, Gilbert Keith. A Defence of Nonsense: And Other Essays. Dodd, Mead, 1911. 5.

[11] Elias, Norbert, and Anca Parvulescu. "Essay on laughter." Critical Inquiry 43, no. 2 (2017): 281-304.

[12] Freud, Sigmund. Jokes and their relation to the unconscious. WW Norton & Company, 1960. 117, 128.

[13] Godfrey, John J., and Edward Holliman. "Switchboard-1 Release 2." Linguistic Data Consortium, Philadelphia 926 (1997): 927.

[14] Guerini, Marco, Gözde Özbal, and Carlo Strapparava. "Echoes of persuasion: The effect of euphony in persuasive communication." arXiv preprint arXiv:1508.05817 (2015).

[15] Kotthoff, Helga. "Gender and humor: The state of the art." Journal of pragmatics 38, no. 1 (2006): 4-25.

[16] Hobbes, Thomas. "orig. 1651. Leviathan." (1982).

[17] Kiddon, Chloe, and Yuriy Brun. "That's what she said: double entendre identification." In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2, pp. 89-94. Association for Computational Linguistics, 2011.

[18] Nesteroff, Kliph. The Comedians: Drunks, Thieves, Scoundrels, and the History of American Comedy. Open Road+ Grove/Atlantic, 2015. p 181-186, 332, 353.

[19] McFee, Brian, Colin Raffel, Dawen Liang, Daniel PW Ellis, Matt McVicar, Eric Battenberg, and Oriol Nieto. "librosa: Audio and music signal analysis in python." In Proceedings of the 14th python in science conference, vol. 8. 2015.

[20] Mihalcea, Rada, and Carlo Strapparava. "Making computers laugh: Investigations in automatic humor recognition." In Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing, pp. 531-538. Association for Computational Linguistics, 2005.

[21] Mimno, David, and Andrew McCallum. "Topic models conditioned on arbitrary features with dirichlet-multinomial regression." arXiv preprint arXiv:1206.3278 (2012).

[22] Princeton University "About WordNet." WordNet. Princeton University. 2010.

[23] Purandare, Amruta, and Diane Litman. "Humor: Prosody analysis and automatic recognition for f* r* i* e* n* d* s." In Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, pp. 208-215. Association for Computational Linguistics, 2006.

[24] Rickles, Don. Hello Dummy! - Recorded Live at the Sahara Hotel, Las Vegas, Nevada, 1995. Accessed Sept. 19, 2019. https://open.spotify.com/track/3SrghhJzMkyzGao1b4IvhH?si=8G4qSh3wSzWhNFds_PPvIQ

[25] Roberts, Margaret E., Brandon M. Stewart, and Richard A. Nielsen. "Adjusting for confounding with text matching." American Journal of Political Science (2018).

[26] Ryokai, Kimiko, Elena Durán López, Noura Howell, Jon Gillick, and David Bamman. "Capturing, Representing, and Interacting with Laughter." In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, p. 358. ACM, 2018.

[27] So, Richard Jean, Hoyt Long, and Yuancheng Zhu. "Race, Writing, and Computation: Racial Difference and the US Novel, 1880-2000." (2019).

[28] Thurm, Eric. "How a Comedy Special Gets Made." The Outline. The Outline, June 8, 2017. https://theoutline.com/post/1648/how-comedy-specials-get-made-rory-scovel.

[29] Yang, Diyi, Alon Lavie, Chris Dyer, and Eduard Hovy. "Humor recognition and humor anchor extraction." In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pp. 2367-2376. 2015.

[30] Underwood, Ted, David Bamman, and Sabrina Lee. "The transformation of gender in English-language fiction." (2018).

[31] Zinoman, Jason. "The Netflix Executives Who Bent Comedy to Their Will." The New York Times. The New York Times, 9 Sept. 2018. www.nytimes.com/2018/09/09/arts/television/netflix-comedy-strategy-exclusive.