MUCH: A Multimodal Corpus Construction for Conversational Humor Recognition Based on Chinese Sitcom

Hongyu Guo[1], Wenbo Shang[2], Xueyao Zhang[1], Binyang Li[1,*]
[1]University of International Relations, [2]Hong Kong Baptist University
[1]Beijing, China , [2] Hong Kong, China
[1]{chloe_guo, Zhang_X_Y, byLi}@uir.edu.cn
[2]cswbshang@comp.hkbu.edu.hk

Abstract

Conversational humor is the key to capturing dialogue semantics and dialogue comprehension, which is usually generated in multiple modalities, such as linguistic rhetoric (textual modality), exaggerated facial expressions or movements (visual modality), and quirky intonation (acoustic modality). However, existing multimodal corpora for conversation humor are coarse-grained, and the modality is insufficient to support the conversational humor recognition task. This paper designed an annotation scheme for multimodal humor datasets, and constructed a corpus based on a Chinese sitcom for conversational humor recognition, named MUCH. The MUCH corpus consists of 34,804 utterances in total, and 7,079 of them are humorous. We employed both unimodal and multimodal methods to test our MUCH corpus. Experimental results showed that the multimodal approach could achieve 75.94% in terms of F1-score and surpassed the performance of most unimodal methods, which demonstrated that the MUCH corpus was effective for multimodal humor recognition tasks.

Keywords: Conversational Humor, Annotation Scheme, Chinese Sitcom

1. Introduction

Humor, as the nature of experiences to induce laughter and provide amusement (Warren et al., 2018), plays an important role in machine translation (Hutchins, 1995), reading comprehension (Mckee, 2012), and sentiment analysis (Taboada, 2016), etc.. There are two main forms of humorous expression (Attardo et al., 2013): one-liners and conversational humor. One-liners are concise, non-narrative sentences in a joke, while conversational humor are linguistic narrative sentences that express humor in dialogues. Different from one-liners, conversational humor is generated based on the context of the dialogue and expressed more flexibly in conversations. Therefore, conversational humor recognition is significant for capturing the humorous semantics and dialogue comprehension.

In the real world, a conversation between individuals is usually adopted in a face-to-face way. Therefore, in addition to text, other modalities will be used to generate humor, such as funny facial expressions and quirky intonations. In Figure 1, actors frequently employ humorous language and quirky tones, together with exaggerated expressions and actions to evoke humor. Similar situations widely occur in the real world. Therefore, more information in different modalities should be taken into consideration to recognize conversational humor.

(Caption) Actors: Special Events.

Figure 1: Quirky tones and exaggerated expressions are the key to provoke laughter.

However, most of the existing research mainly used textual modality to recognize conversational humor (Taylor, 2004; Yang et al., 2015), which was difficult to identify conversational humor from unimodality. In conversations, to enhance humorous expressions, people may combine the acoustic rhythm (quirky intonation, etc.) or the visual features (comical expressions or movements, etc.) interacting with the textual contents. Some research has attempted to employ the visual modality as a supplement to the textual modality (Purandare and Litman, 2006). More recent studies have built several datasets that accounted for other modal information rather than textual information alone. For instance, Bertero and Fung (2016a) built a humor dataset that employed canned laughter as an indicator to denote the humor scenes; Boccignone et al. (2017) proposed a multimodal dataset to detect humor from images. However, these datasets did not cover all modalities. As a result, they are coarse-grained in capturing the differentiation of multimodalities and perform suboptimal in conver-

[*] Corresponding author

LREC-COLING 2024, pages 11692–11698
20-25 May, 2024. © 2024 ELRA Language Resource Association: CC BY-NC 4.0

sational humor recognition.

To this end, this paper aims to construct a **mul**timodality corpus for **c**onversational **h**umor recognition, named **MUCH** [1], which was constructed based on a Chinese sitcom, *iPartment*. In order to better represent conversational humor, MUCH covers three modalities, including text, vision, and acoustics.

The main contributions of this paper are listed below:

- An annotation scheme for conversational humor recognition datasets is designed, involving three modalities (text, vision, and acoustics). Among them, the visual modality includes exaggerated facial expressions and movements, the acoustic modality involves abnormal intonation and homophones that can indicate humor. According to the scheme, a multimodal conversational humor corpus (MUCH) was constructed by manually annotating based on the Chinese sitcom *iPartment*.

- The MUCH corpus consists of 34,804 utterances in total, and 7,079 of them are humorous. Among the humorous utterances, 5,163 utterances evoke humor by unimodal expressions (i.e., textual (T), visual (V), and acoustic (A) modalities), and 1,916 required two or more modalities (i.e., T+V, T+A, V+A, T+V+A) to generate humor.

- To assess the MUCH corpus, we conducted several experiments to compare some classical methods, including both unimodal and multimodal. The experimental results showed that the multimodal method outperformed unimodal methods. As to unimodal methods, the textual method RoBERTa performed best, achieving the accuracy of 69.17% and the F1-score of 67.73%. As to the multimodal method, CLIP performed best, reaching the accuracy of 82.96% and the F1-score of 75.94%. This proved the suitability of the corpus for multimodal conversational humor recognition.

2. Related Work

Humor, an essential element in interpersonal communication, serves as a vital medium for expressing emotions in humans (Meyer, 2010), can be classified into two categories according to whether it is narrative, i.e., one-liners and conversational humor (Attardo et al., 2013).

Non-narrative one-liners are characterized by simple syntax and creative semantic structures. Mihalcea and Strapparava (2005) conducted humor analysis on one-liners, and Yang et al. (2015) utilized hand-crafted and non-neural models to recognize the underlying semantic structures of humor in one-liners.

Similar to one-liners recognition, research on conversational humor firstly focused on the textual modality for recognition. Zhang and Liu (2014) collected humor dataset from Twitter, and Chen and Lee (2017) collected and annotated TED speech transcripts.

In recent work, multimodal datasets were constructed to better understand semantics and employed to recognize conversational humor. Chandrasekaran et al. (2016) analyzed humorous expressions in abstract scenes, and Boccignone et al. (2017) proposed a multimodal dataset for detecting humor in images.

More recently, to better capture the humorous expressions in real conversations, some research on multimodal conversational humor datasets has been built based on sitcoms, including *Friends* (Poria et al., 2018), *The Big Bang Theory* (Patro et al., 2021), *Seinfeld* (Bertero and Fung, 2016b) etc.. However, most of the above datasets used acoustic modality, i.e., canned laughter, as an indicator for annotation, and other modalities were not fully exploited.

Therefore, in order to better recognize conversational humor, this paper constructed a multimodal humor corpus annotation scheme and manually annotated a fine-grained conversational humor corpus (MUCH) based on a Chinese sitcom.

3. Corpus Construction

According to the analysis of the multimodal conversational humor recognition task, the corpus for the task should satisfy the following requests (Hasan et al., 2019): 1) More than one speaker should be involved to form a conversation; 2) Different topics or scenes should be involved that can showcase various humor styles.

For this purpose, we choose the classic Chinese sitcom *iPartment* as the basic dataset, which has 7 main characters, and each episode revolving around different plot topics being densely packed with laughter.

Based on the above sitcom dataset, we firstly divide each episode of *iPartment* into several dialogues, while each dialogue contains a series of sequential utterances. For each individual utterance, it will be annotated as humorous or non-humorous. More specifically, each humorous utterance is further annotated to indicate the presence of peculiar intonation, comical actions, or facial expressions. Additionally, each utterance is associated with other attributes, such as *Speaker* and *Sentence*. The detailed description and analysis of the MUCH will be

[1] The full MUCH corpus is publicly available for use at MUCH_Corpus.

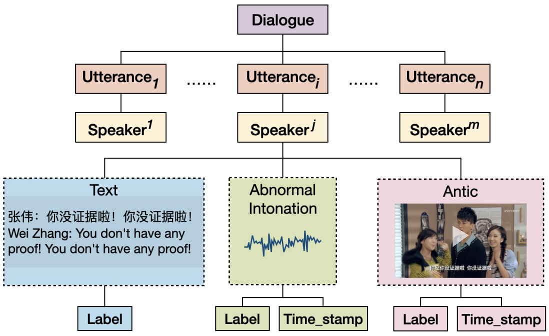


Figure 2: The annotation scheme.

shown in the following subsections.

3.1. Annotation Scheme

Without the loss of generality, we divide each episode of the sitcom into several dialogues based on difference scenes and plots. For each dialogue, there are *n* utterances, utterance$_i$ $(i = 1, 2, ..., n)$.

Figure 2 demonstrates our proposed annotation scheme for the conversational humor dataset, which has the following attributes:

Speaker: *Speaker* plays a particular role in the generation of humor, and different speakers have different styles of humor.

Text: *Text* represents the textual content of the utterance that indicates the specific manifestation of humor in the textual modality. This attribute is labeled as *1* when textual humor occurs and labeled as *0* otherwise.

Abnormal Intonation: *Abnormal intonation* can enhance the speaker's emotional expression. Moreover, due to language differences between Chinese and English, Chinese individuals often employ techniques such as rhyming and regional accents when expressing humor. In this case, humor cannot be recognized through the textual and visual modalities but can be through the acoustic modality. This attribute is labeled as *1* when an abnormal intonation occurs and labeled as *0* otherwise.

Antic: *Antic* refers to the characters exhibit comical expressions or gestures during the conversation. When comical expressions or gestures occur, *Antic* is labeled as *1*; otherwise, it is labeled as *0*. It allows for humor recognition in the visual modality through the annotation of comical actions and expressions.

In our annotation scheme, three modalities are all taken into consideration. Each utterance is annotated to indicate each modality. For example, if the textual modality is judged humorous, its corresponding label is annotated as *1*, otherwise, the label is annotated as *0*. Similarly, when humor is recognized in acoustic modality and visual modality, their label is annotated as *1*. Moreover, if the label of the acoustic or visual modality is *1*, we also recorded the time_stamp and the video clip of the utterance.

3.2. Annotation Process

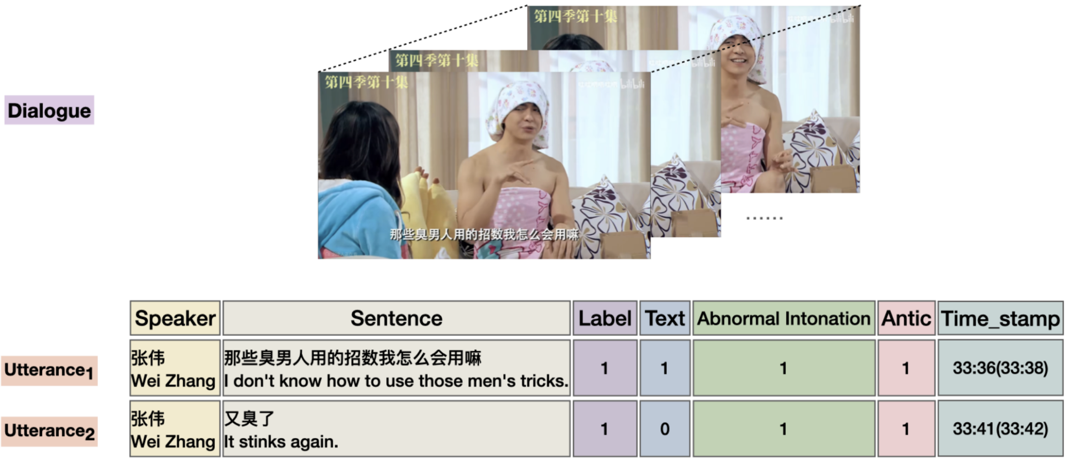
Unlike other NLP annotation tasks, the recognition of conversational humor could vary from one person to another. Therefore, we found 12 annotators to annotate the MUCH dataset, all of whom have experience in data annotation. The annotators were divided into four groups for annotation.

The annotation processing has the following steps: (1) Divide each episode into several dialogues based on different plots and scenes; (2) Record all the utterances in each dialogue and record the speaker and the content of each utterance; (3) Judge whether each utterance embodies humor in textual, visual, and acoustic modalities following the proposed scheme. At the same time, when humor is expressed visually or acoustically, the time_stamp of the corresponding utterance is also provided.

Toward a specific instance, the processing of annotation can be classified into three cases: (1) If 3 annotators achieved the agreement, e.g., they all regarded the instance as humorous or non-humorous, the instance was labeled as *1* or *0*, and the annotation was completed; (2) If 2 annotators labeled the instance as humorous, the instance was labeled as *1* according to the majority rule; (3) If only 1 annotator considered the instance to be humorous, considering the contingency of humor, it required the other 9 annotators to annotate the instance and determine the final label by applying the majority rule.

During our annotation process, we also discovered that sitcoms often use canned laughter to induce laughter in the audience. For the utterance with canned laughter, which our annotators did not think was humorous in any modality, we also marked it and retained the time_stamp and other information for further fine-grained research of the corpus in the future.

Figure 3 provides an example of our annotation. Based on *Wei Zhang*'s performance, the annotators firstly determine whether the overall utterance is humorous (labeled as *1*) or non-humorous (labeled as *0*). Then, the annotators judge whether it is humorous in each modality. In the textual modality, taking *utterance*$_1$ as an example, *"I don't know how to use those men's tricks,"* conflicts with speaker's masculine gender and plays a taunting role, so it is annotated as humorous in the textual modality (labeled as *1*). At the same time, *Wei Zhang* uses a feminine tone and dress to conduct dialogue, so it is annotated as humorous in both acoustic and visual modalities.



	Speaker	Sentence	Label	Text	Abnormal Intonation	Antic	Time_stamp
Utterance$_1$	张伟 Wei Zhang	那些臭男人用的招数我怎么会用嘛 I don't know how to use those men's tricks.	1	1	1	1	33:36(33:38)
Utterance$_2$	张伟 Wei Zhang	又臭了 It stinks again.	1	0	1	1	33:41(33:42)

Figure 3: An annotation example.

Filed		Value
# Dialogue		1,626
# Utterance		34,804
# Speaker		423
# Humorous utterance		7,079
Total duration in hour		62
Avg. duration of dialogue (minutes)		2.76
Avg. duration of utterance (seconds)		2.89
# Humor in unimodal	T	3,661
	V	647
	A	855
# Humor in multimodal	T+V	615
	T+A	514
	V+A	347
	T+V+A	440

Table 1: Statistics of MUCH corpus. Here, '#' denotes number, 'Avg.' denotes average, 'T' denotes text, 'V' denotes vision, and 'A' denotes acoustics.

3.3. Statistics

The MUCH corpus consists of 4 seasons of *iPartment*, 80 episodes in total, with each episode approximately 45 minutes long. It consists of 34,804 utterances in total, and 7,079 of them are humorous. Of the humor utterances, 5,163 utterances generate humor by unimodal expressions (i.e., textual (T), visual (V), and acoustic (A) modalities), and 1,916 required two or more modalities (i.e., T+V, T+A, V+A, T+V+A) to generate humor. Details are shown in Table 1.

We also compare the MUCH with other current multimodal humor datasets. It can be seen in Table 2 that the corpus annotation scheme we proposed can label humor for different modalities separately, not just the overall humor.

4. Experiment

4.1. Approaches for Comparison

In order to assess the effectiveness and suitability of MUCH corpus for the conversational humor recognition task, we conducted several experiments using both unimodal and multimodal methods.

Unimodal Method: We firstly adopted several unimodal methods for each individual modality and test the performance.

- Textual modality methods: The specific textual content of the dialogue is the main part of the production of conversational humor. In this paper, we employed the **BERT** (Kenton and Toutanova, 2019) and **RoBERTa** (Liu et al., 2019) for testing.

- Visual modality methods: Conversational humor is evoked by comical postures or facial expressions. Therefore, we employed **ViT** (Dosovitskiy et al., 2020) to recognize facial expressions and **OMNIVORE** (Girdhar et al., 2022) to recognize actions in the visual modality.

- Acoustic modality method: Acoustic rhythmic features such as pitch and energy have been demonstrated to be associated with humorous expression in sitcoms (Purandare and Litman, 2006). We employed **openSMILE** (Eyben et al., 2010) to extract audio features.

Multimodal Method: To assess our MUCH corpus, we also tested a classic multimodal method, **CLIP** (Radford et al., 2021), which could capture the feature from at least two modalities.

Dataset	**Source**	**Annotation Process**	**Modalities**	**Language**
UR-Funny (Hasan et al., 2019)	TED speech	Provide videos and their transcripts from the TED portal.	T, V, A	English
MuStARD (Castro et al., 2019)	Sitcom	Provide the utterance and the corresponding original fragment, while also proving contextual information.	T, V, A	English
TBBT (Kayatani et al., 2021)	Sitcom	Provide the utterance and the corresponding original fragement; Only the overall label.	T, V	English
MHD (Patro et al., 2021)	Sitcom	Annotation based on canned laughter.	T, V, A	English
M2H2 (Chauhan et al., 2021)	Sitcom	Provide the utterance and the corresponding original fragement; Only the overall label.	T, V, A	Hindi
MUCH (Ours)	Sitcom	Provide an overall label and labels for each of the three modalities for each utterance.	T, V, A	Chinese

Table 2: Comparison between MUCH and other multimodal humor datasets. It can be seen that compared with these datasets, MUCH has three respective annotations on the three modalities, not just the overall label.

Modality		**Method**	**Acc.**(%)	**P**(%)	**R**(%)	**F1**(%)
Unimodal	Text	BERT	65.18	60.72	61.44	61.08
		RoBERTa	**69.17**	**70.37**	**65.28**	**67.73**
	Vision	ViT	65.72	65.17	60.25	62.61
		OMNIVORE	60.13	60.37	59.12	59.47
	Acoustic	openSMILLE	56.41	55.93	61.01	58.36
Multimodal		CLIP	**82.96**	**74.86**	**77.05**	**75.94**

Table 3: Performance of conversational humor recognition via unimodal and multimodal approaches on the MUCH corpus.

4.2. Evaluation Metrics

We divided the MUCH corpus into training, development, and testing sets with a ratio of 65%, 15%, and 20%. Following Liang et al. (2022), we use accuracy (Acc.), precision (P), recall (R), and F1-score (F1) as the evaluation metrics.

4.3. Experimental Result

The performance of the MUCH corpus for recognizing conversational humor by both unimodal and multimodal methods is shown in Table 3.

Based on the experimental results, we have the following observations:

- For the unimodality, the BERT and RoBERTa were employed for humor recognition in the textual modality, and achieved the accuracy of 65.18% and 69.17%, and the F1-score of 61.08% and 67.73%, respectively. The visual modality employed the ViT and OMNIVORE, achieved the accuracy of 65.72% and 60.13% and the F1-score of 62.61% and 59.47%, respectively. In the acoustic modality, the openSMILE was employed and attained the accuracy of 56.41% and the F1-score of 58.36%. For the multimodality, the CLIP performed well, achieved the accuracy of 82.96% and the F1-score of 75.94%.

- The experimental results of the MUCH corpus in both unimodal (textual, visual, and acoustic) and multimodal methods demonstrated the applicability of the corpus annotation scheme we have constructed.

- Method using multimodality outperformed methods that used unimodalities. Between text and nonverbal behaviours (vision and acoustic), text proved to be the most important modality. In most cases, multimodal methods are performs better than text alone for humor recognition.

5. Conclusion

In this paper, we proposed a new multimodal conversational humor annotation scheme and manually annotated the MUCH corpus. The MUCH corpus was constructed based on a Chinese sitcom and includes three modalities: text, vision, and acoustics. It consists of 34,804 utterances in total, and 7,079 of them are humorous. We conducted several experiments based on some classical methods, including both unimodal and multimodal. We discovered that the multimodal approach surpassed the performance of most unimodal methods.

During the annotation process, we also annotated the speaker of each utterance, which plays an important role in the generation of humor. We will conduct future research on the effect of the speaker on the conversational humor recognition task.

6. Acknowledgements

This paper was partially supported by National Natural Science Foundation of China (Grant number: 61976066), Beijing Natural Science Foundation (Grant number: 4212031), and Research Funds for NSD Construction, University of International Relations (Grant numbers: 2021GA07).

We would also like to thank Ningyuan Gao, Kexin Guo, Zihan Wu, Chenguang Wang, Yu Zeng, Yining Zhang, and others for their hard work on the annotations.

7. Ethical Consideration

We constructed the MUCH corpus based on the Chinese sitcom, *iPartment*, which has been licensed for academic research, and the annotation for the MUCH corpus was done by human experts, who are regular employees of our research group. The MUCH corpus is freely available and will be used only for the purpose of academic research. There are no other issues to declare.

8. References

Salvatore Attardo, Lucy Pickering, Fofo Lomotey, and Shigehito Menjo. 2013. Multimodality in conversational humor. *Review of Cognitive Linguistics. Published under the auspices of the Spanish Cognitive Linguistics Association*, 11(2):402–416.

Dario Bertero and Pascale Fung. 2016a. Deep learning of audio and language features for humor prediction. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 496–501.

Dario Bertero and Pascale Fung. 2016b. Predicting humor response in dialogues from tv sitcoms. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5780–5784. IEEE.

Giuseppe Boccignone, Donatello Conte, Vittorio Cuculo, and Raffaella Lanzarotti. 2017. Amhuse: a multimodal dataset for humour sensing. In *Proceedings of the 19th ACM international conference on multimodal interaction*, pages 438–445.

Santiago Castro, Devamanyu Hazarika, Verónica Pérez-Rosas, Roger Zimmermann, and Soujanya Poria. 2019. Towards multimodal sarcasm detection (an $_obviously_perfect paper)$.

Arjun Chandrasekaran, Ashwin K Vijayakumar, Stanislaw Antol, Mohit Bansal, Dhruv Batra, C Lawrence Zitnick, and Devi Parikh. 2016. We are humor beings: Understanding and predicting visual humor. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4603–4612.

Dushyant Singh Chauhan, Gopendra Vikram Singh, Navonil Majumder, Amir Zadeh, Asif Ekbal, Pushpak Bhattacharyya, Louis-philippe Morency, and Soujanya Poria. 2021. M2h2: A multimodal multiparty hindi dataset for humor recognition in conversations. In *Proceedings of the 2021 International Conference on Multimodal Interaction*, pages 773–777.

Lei Chen and Chong Min Lee. 2017. Convolutional neural network for humor recognition. *arXiv preprint arXiv:1702.02584*.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

Florian Eyben, Martin Wöllmer, and Björn Schuller. 2010. Opensmile: the munich versatile and fast open-source audio feature extractor. In *Proceedings of the 18th ACM international conference on Multimedia*, pages 1459–1462.

Bobak Farzin, Piotr Czapla, and Jeremy Howard. 2019. Applying a pre-trained language model to spanish twitter humor prediction. *arXiv preprint arXiv:1907.03187*.

Rohit Girdhar, Mannat Singh, Nikhila Ravi, Laurens van der Maaten, Armand Joulin, and Ishan Misra. 2022. Omnivore: A single model for many visual modalities. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16102–16112.

Md Kamrul Hasan, Wasifur Rahman, Amir Zadeh, Jianyuan Zhong, Md Iftekhar Tanveer, Louis Philippe Morency, Mohammed, and Hoque. 2019. Ur-funny: A multimodal language dataset for understanding humor.

W. John Hutchins. 1995. Machine translation: A brief history. In E.F.K. KOERNER and R.E. ASHER, editors, *Concise History of the Language Sciences*, pages 431–445. Pergamon, Amsterdam.

Yuta Kayatani, Zekun Yang, Mayu Otani, Noa Garcia, Chenhui Chu, Yuta Nakashima, and Haruo Takemura. 2021. The laughing machine: Predicting humor in video. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 2073–2082.

Jacob Devlin Ming-Wei Chang Kenton and Lee Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of naacL-HLT*, volume 1, page 2.

Bin Liang, Chenwei Lou, Xiang Li, Min Yang, Lin Gui, Yulan He, Wenjie Pei, and Ruifeng Xu. 2022. Multimodal sarcasm detection via cross-modal graph convolutional network. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 1767–1777. Association for Computational Linguistics.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Rod A Martin. 2007. The psychology of humor: an integrative approach. *Academic Pr Inc*.

Steve Mckee. 2012. Reading comprehension, what we know: A review of research 1995 to 2011. *Language Testing in Asia*, 2(1):45.

John C Meyer. 2010. Humor as a double-edged sword: Four functions of humor in communication. *Communication Theory*, 10(3):310–331.

Rada Mihalcea and Carlo Strapparava. 2005. Making computers laugh: Investigations in automatic humor recognition. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 531–538.

Badri N Patro, Mayank Lunayach, Deepankar Srivastava, Hunar Singh, Vinay P Namboodiri, et al. 2021. Multimodal humor dataset: Predicting laughter tracks for sitcoms. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 576–585.

Soujanya Poria, Devamanyu Hazarika, Navonil Majumder, Gautam Naik, Erik Cambria, and Rada Mihalcea. 2018. Meld: A multimodal multi-party dataset for emotion recognition in conversations. *arXiv preprint arXiv:1810.02508*.

Amruta Purandare and Diane Litman. 2006. Humor: Prosody analysis and automatic recognition for f*r*i*e*n*d*s*. In *Empirical Methods in Natural Language Processing*.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR.

Maite Taboada. 2016. Sentiment analysis: An overview from linguistics. *Annual Review of Linguistics*, 2(1):325–347.

Julia M Taylor. 2004. Computationally recognizing wordplay in jokes. *Proceedings of Cogsci*, 53(1):1315–1320.

Caleb Warren, Adam Barsky, and A Peter McGraw. 2018. Humor, comedy, and consumer behavior. *Journal of Consumer Research*, 45(3):529–552.

Diyi Yang, Alon Lavie, Chris Dyer, and Eduard Hovy. 2015. Humor recognition and humor anchor extraction. In *Proceedings of the 2015 conference on empirical methods in natural language processing*, pages 2367–2376.

Zekun Yang, Yuta Nakashima, and Haruo Takemura. 2023. Multi-modal humor segment prediction in video. *Multimedia Systems*, pages 1–10.

Renxian Zhang and Naishi Liu. 2014. Recognizing humor on twitter. In *Proceedings of the 23rd ACM international conference on conference on information and knowledge management*, pages 889–898.