A Dataset for Detecting Humor in Arabic Text

Hend Al-Khalifa*, Fetoun AlZahrani, Hala Qawara, Reema AlRowais, Sawsan Alowa and Luluh AlDhubayi
Information Technology Department, College of Computer and Information Sciences
King Saud University
Riyadh, Saudi Arabia
*Hendk@ksu.edu.sa

Abstract

Humor detection is a complex and ambiguous task in natural language processing. This has made automatic humor detection challenging, particularly for languages with limited resources such as Arabic. In this paper, we attempt to solve this task by collecting and annotating Arabic humorous tweets in dialects and Modern Standard Arabic (MSA) text then performing automatic humor detection on the collected data. We experimented on the collected dataset by fine-tuning seven Arabic Pre-Trained language models (PLMs) which are: AraBERTv02, Arabertv02-twitter, QARIB, MarBERT, MARBERTv2, CAMeLBERT-DA, and CAMeLBERT-MIX to establish a baseline classification system. We concluded that CAMeLBERT-DA was the best-performing model and it achieved an F1-score and accuracy of 72.11%.

1 Introduction

Humor is defined as the attempt of practices to provoke laughter and provide a sort of amusement. Humor permeates human life in both dimensions' reality and virtually (Zhang and Liu, 2014). Humor has become a frequent subject of research due to its coercive power in communication and other fields (Meyer, 2000). For instance, computational and linguistic research in humor have been going for more than twenty years, and according to (Amin and Burghardt, 2020), they mainly focus on Humor Recognition, Humor Adoption Systems, Computational Humor Evaluation, Computational Humor Applications and Computational Humor Datasets and Corpora. In fact, the power of social media and the popularity of humor in it encouraged some Natural Language Processing (NLP) researchers to conduct their studies about humor on Twitter and other social media platforms such as (Zhang and Liu, 2014), (Khandelwal et al. 2018) and (Meaney et al. 2021). Drawing on the humor context, Arabic linguistic research on humor is sparse, some of it focused on certain Arabic dialects and humor aspects e.g. (Nayef and El-Nashar, 2014), (Omar, 2017), (Banikalef, 2014) and (Bzour, 2021) conducted an analysis of Arabic humor.

According to (Altin 2019), the main benefit of detecting humor is to improve human machine communication, as the machines are not fully capable to understand humor as humans do. Therefore, detecting humor is very essential to make progression in human alike and intelligent systems.

The lack of high quality annotated data for conducting various experiments is one of the main obstacles to developing any detection models in low-resource languages. In order to address this issue, this study is the first to our knowledge to detect humor in text written in Arabic language, whether it is MSA or Arabic Dialect. The study includes details about the process of collecting, preprocessing, analyzing the data, and the encountered challenges during the process. The dataset we collected is publicly available[1].

The rest of the paper is organized as follows: Section 2 presents previous work in the area of humor detection. Section 3, provides detailed description of the dataset. Section 4, briefly describes the used experiments and analysis of results. Finally, Section 5 concludes the paper with future work.

2 Related work

The task of detecting humor in text has drawn the attention of many researchers in different languages. Although there is no single definition of

[1] https://github.com/iwan-rg/Arabic-Humor

humor, most systems focus on identifying jokes, irony, and other forms of verbal play. Early humor detection systems were based on rule-based approaches, which relied on hand-crafted rules to identify humor. However, these systems were limited in their ability to generalize to new types of humor (Rajakumar et al., 2010).

More recent systems have employed machine learning techniques to automatically learn rules for identifying humor. These approaches have shown promise, but still face challenges in accurately detecting humor. For instance, Hossain et al.(2019) analysed regular English news headlines to predict whether or not an edited headline is funny. Kao et al. (2016) did a humor detection in puns in English text at a fine-grained level. Similarly, a novel annotation scheme in Chinese text contained annotated humorous text and keywords that triggered humor, with 9,123 Chinese jokes and 39,977 sentences in total (Zhang et al. 2019). HEMOS (Humor-EMOji-Slang-Based) is a deep learning system for sentiment classification of Chinese language in two collected lexicons (slang expression and converted Weibo[2] emojis). It has a binary annotation (optimistic humorous and pessimistic humorous) which added to the standard positive and negative sentiments. The carried experiment was implemented on both lexicons to an attention-based bi-directional long short-term memory recurrent neural network (AttBiLSTM). It resulted in a substantial improvement in predicting sentiment polarity on Weibo (Li et al. 2020). Furthermore, a Spanish based task by (Castro et al. 2016) to detect humor on a crowdsourced corpus of labelled data classified by implementing number of features such as adult slang, dialogue, hashtag, keywords, etc. Likewise, a neural network for humor recognition using (biLSTM) models used to classify tweets in Spanish as humorous or not. Additionally, it scored the level of funniness in the context based on Human annotation from one to five where one (not funny) and five (excellent) (Altin et al. 2019).

To tackle the problem of humor detection the attention of research has shifted to deep learning approaches as seen in the two previous studies and as in (Annamoradnejad and Zoghi 2020), who used sentence embeddings and utilized the linguistic structure of humor in designing their proposed model. Their model outperforms the baseline models and achieved high accuracy in detecting humor in their ColBERT experiment on labelled dataset, which is technically injecting BERT sentence embedding into a neural network model that processes sentences separately in parallel hidden layers. Also, Fan et al.(2020) proposed an internal and external attention neural network (IEANN) for the humor detection task by merging two types of attention mechanisms to capture the incongruity and ambiguity in humorous text. An experiment conducted on two humor datasets to test the proposed model and the results showed that their model has better interpretability as they claimed. Researchers have shown an increased interest in automatic humor detection. A data collected from Twitter and Kaggle[3] was evaluated using several models to the purpose of detecting and rating humor and offensive language, interestingly, they also originally predict humor controversy which result from the variance of annotators rating of certain jokes (Meaney et al. 2021).

Other than text-based detection of humor, multimodal language dataset were used in (Hasan et al. 2019) and (Bertero and Fung, 2016). The first one, determined the effects of using (text, vision and audio) all in one dataset called UR-FUNNY for humor detection task and present the performance results of the task. The latter proposed a deep neural network framework of integration of audio and language features for their dataset that collected from three TV shows that have canned laughter which used as indication of the humor occurrence.

Finally, related works for humor detection in Arabic language is scarce. They are usually considered as part of sarcasm detection task, therefore, we aim in this study to build the first Arabic humor dataset and evaluate it using different Arabic PLMs, to establish a baseline classification system for this downstream task.

3 Arabic Humor Dataset

3.1 Data Collection

To collect humorous text in Arabic language different tools were used. Twint [4] and Sketch Engine[5] are used to create a corpus of more than 10,000 entries. Firstly, tweets were scrapped from Twitter using Twint by following two approaches.

[2] is the largest Chinese social network.
[3] https://www.kaggle.com/
[4] https://github.com/twintproject/twint
[5] https://www.sketchengine.eu/

The first approach, scrapping using username where multiple accounts known for their humorous content were selected. The second approach, scrapping using trending hashtags such as "نفسيه"and "نكت" which are related to humor aspect. The list of used usernames and hashtags are given in Table 1.

Secondly, Sketch Engine was used for scrapping humorous text by using the web search tool that was provided by the Sketch Engine. By entering a set of keywords that are likely used for triggering humor to collect data that are relatively contain humorous text. The list of used keywords is given in Table 1.

Twitter		**Web Search**
Usernames	**Hashtags**	**Keywords**
7amadQalam, _3urf, Fake_SteveJobs1 iiim7mdz, mnm2900, Jokes_HD, Jokes_LoooL, nektah_time	#وليد_مع_اكشن نفسيه#	مرة محشش، في محشش، محشش، مرة واحد، في عجوز، غبي، مصري، حضرمي، حضرميه، فيه واحد، واحد تزوج، فيه سوداني، فيه مجنون، قصيمي، كان فيه قصيمي، جحا، فيه صعيدي، حوطي، حواطى، فيه جنوبي، جيزاني، غمد وزهارين، مسطول، منسم

Table 1: Lists of Usernames, Hashtags, and Keywords

3.2 Data Cleaning and Preprocessing

The collected text for the dataset are unstructured with slang terms, abbreviations, and orthographic errors that should be cleaned and pre-processed before performing the corpus analysis properly. Thus, the unstructured text must be converted to a structured format using pre-processing techniques, which are an important step in natural language processing (Aljabri et al. 2021).

- Arabic normalization:
  The unification of some characters that have many forms, was employed on the dataset by replacing (ئ،ى) with (ي), letter (ؤ) with (و), letter (ة) to (أ،إ،آ) to (ا) and (گ) to (ك).
- Deleted unrelated and unwanted text:
  The extracted data from the web was not limited to certain tags; it extracted all the content from the web pages that match the search keywords. As a result, the data contained unrelated information such as information about the topic writer, number of contributions and contact information; hence most of the collected web data were from forums.

Moreover, the data cleaning steps that were applied on web data were as follow:

- Remove unrelated information:
  A dictionary of unrelated words was created. The dictionary contains the following keywords: " عضو"، "المحتويات"، "كاتب الموضوع"، "عدد المساهمات"، "عدد الرسائل"، "معلومات الاتصال"،"بيانات العضو"،"مساهمة". After that a special query is created to search through the data frame to drop any row that contains one or more of the keywords in the dictionary.
- Remove diacritics or Arabic Tashkeel:
  Diacritics is a vowel mark used to indicate short vowels which are: fatha, damma, and kasra, in addition to sukun and shadda which indicates the absence of a vowel. Diacritics are removed to reduce data sparsity.
- Remove numbers, English letters, special characters, symbols, emojis, and newlines.
- Remove all words that contain underscore and hashtag sign from the data.
- Remove repeated characters such as "هههه".
- Remove duplicate rows from the dataset.

Finally, after applying the preprocessing and cleaning of the data, we performed tokenization where we split the text into small pieces called tokens. A simple tokenizer was implemented which is space-based. This tokenizer splits the text into chunks by its spaces using regular expression (regex). Table 2 demonstrated an example from the dataset and how it changed during the data preprocessing and cleaning process.

Process	**Change**
Raw Tweet	محشش حط على السيارة برقع ليييييييش علشانها نكت# 🤭👏🤦 باسم امه
Normalization	محشش حط علي السياره برقع ليييييييش علشانها نكت# 🤭👏🤦 باسم امه
Remove Hashtags	محشش حط علي السياره برقع ليييييييش علشانها 🤭👏🤦 باسم امه
Remove Emojis	محشش حط علي السياره برقع ليييييش علشانها باسم امه
Remove Repeating Char	محشش حط علي السياره برقع ليش علشانها باسم امه
Simple Tokenization using regex (space-based)	['محشش'، 'حط'، 'علي'، 'السياره'، 'برقع'، 'ليش'، 'علشانها'، 'باسم'، 'امه"]

Table 2: Pre-processing and Data Cleaning in Tweet.

3.3 Dataset Description

The purpose of the collected dataset is to detect humor in Arabic text which contains different Arabic dialects like Gulf, Egyptian, Levantine and Maghrebi besides MSA (Modern Standard Arabic). The size of the dataset was 10039 entries collected from Twitter and the web. The text length varied from 3 words minimum up to 446 words maximum. Additionally, the dataset contains 201,095 tokens and 36814 word types. The word type is defined here by counting the number of vocabularies (distinct words) in the corpus.

3.4 Dataset annotation

The annotation guidelines were explained to the annotators through online and face to face meetings. The given guidelines were as follows:

- The labels are defined to the annotators as: Humor - The author has written a comic or amusing text. Non-Humor - The author has not written comic or amusing text.
- Perspective: The text should be considered humorous or not from the annotator's perspective.
- Ambiguity: If the text is not understandable or you cannot get the joke, check non-humor, also if it does not have any keyword related to humorous aspects.
- Incomplete text is caused by retrieving tweets without their corresponding memes, so if the text conveys any funny, amusement or humorous meaning check humor.
- The context of the text could be understood from certain keywords in the text that are related to certain groups or subjects that are known to be subjected to humor examples such as (واحد محشش، قروي، صعيدي، فيه واحد بخيل). Table 3 shows examples of each label.

We tried as much as possible to help annotators in their decision of labelling the data by providing certain guidelines. However, it is difficult to identify something that has no standard definition and ambiguously stated (Castro et al. 2016). In general text comprised of any tales, imaginative, jokes, and abusive are explained as humorous unless it doesn't convey any delightful, entertaining or amusement as in (Khandelwal et al. 2018).

As mentioned above people have different sense of humor so deciding whether the text is humorous or not will be best if it is from the annotator perspective as in (Castro et al. 2016). Since the text collected in MSA and DA (Dialectal Arabic) are different and some of them cannot be understandable by the annotators especially if they were in Magrabi dialect, so we think that the text can be cracked by some known keywords that mostly used in humorous context.

Labels	Examples in Arabic	English Translation
Humor: Something written to be comical or funny	-فيه قروي شاف نافورة جاب لها سباك	- A villager saw a fountain and brought it a plumber.
	-صعيدي راح ل ماكدونالدز وقال لهم عندكم حلقات بصل قالوا ايه قال عطوني الحلقة الأخيرة	- An upper egyptian went to McDonald's and asked them do you have onion rings? ", They said - yes- he said: give me the last episode.
	-في واحد محشش جالس امام المكيف يقوله لا تنفخ لا اكسر راسك	- A stoned sitting in front of the air-conditioner saying: Do not blow or , I will smash your head.
	-فيه واحد بخيل قال الي اولاده اذا نجحو با وريكم سيارة الأيسكريم	-There is a miser told his children that: If you succeed, I will show you the ice cream car.
Non-Humor: Something written not to be humorous or funny	اللي جابو لنا امراض نفسيه عاملين فيها دكاترة نفسيين	Those who brought us Psychiatric illness act like psychiatrists

Table 3: Examples of Humor and non-Humor text in the dataset.

Many tweets contain an image (memes) with a corresponding label that results in losing the full context, here the decision can be made depending on the available text (label or caption) if it conveys any humorous aspect or not, or if it can be understood without the image so it is considered humor, if it is not, it will be non-humor.

Inter-Annotator Agreement (IAA) is a measurement used for the reliability and credibility of the annotators to ensure the quality of the annotation process. It compares the annotators' labels of the text and shows how much annotators agreed upon a particular text or disagreed.

Moreover, it gives insights into how challenging it was to identify each text label and the reasons behind this labeling decision. Also, it indicates whether the annotators fully understand the guidelines given to them or not (Khandelwal et al. 2018).

Since we have three annotators for each text, we choose Fleiss's Kappa (Fleiss, 1971) to measure IAA in our annotated corpus. We calculated Fleiss's Kappa for the 10039 texts and the result of Kappa was 0.73 which is considered substantial according to Kappa statistics (Landis and Koch, 1977) .

Table 4 shows the statistics of the annotated dataset which indicates that the dataset is almost balanced because there is no significant difference between the two classes.

Class	Number of entities	Number of Tokens	Percentage
Humor (1)	4455	86989	44.38%
Non-Humor (0)	5584	114171	55.62%
All	10039	201095	100%

Table 4: Distribution of Humor and non-Humor text in the dataset.

4 Experiments and Results

4.1 Model selection

In this study we selected seven Arabic Pre-Trained Language Models based on their performance on other text classification tasks for Arabic language and also being pre-trained on dialectal Arabic (DA). The models are:

1. **AraBERTv02**[6] is a transformer-based model for Arabic language based on the BERT-Base model.
   AraBERTv02 was trained on the Arabic version of the Books Corpus and the Arabic Wikipedia. The model achieves a state-of-the-art performance on a number of Arabic natural language understanding tasks, including question answering and natural language inference.
2. **Arabertv02-twitter**[7] same as AraBERTv02 and trained on ~60M Arabic tweets (filtered from a collection on 100M).
3. **QARIB**[8]: QCRI Arabic and Dialectal BERT (QARiB) model, was trained on a collection of ~ 420 Million tweets and ~ 180 Million sentences of text.
4. **MarBERT**[9] **:** is a large-scale pre-trained masked language model based on the BERT-Base model and focused on both Dialectal Arabic (DA) and MSA.
5. **MARBERTv2**[10] : same as MarBERT with further pre-training the stronger model, on the same MSA data as ARBERT in addition to AraNews dataset.
6. **CAMeLBERT-DA**[11]: is a BERT model pre-trained on 54GB of dialectal Arabic (DA).
7. **CAMeLBERT-MIX**[12]: is a BERT model pre-trained on 167GB of mix text including Modern Standard Arabic (MSA), dialectal Arabic (DA), and classical Arabic (CA).

The selected models were fine-tuned to perform a binary classification task to detect a given text if it is Humor or Non-Humor.

The dataset was split 85:15 for training and testing respectively using scikit-learn library[13].

4.2 Models' Results

PLM Model	Accuracy	Precision	Recall	F1-score
AraBERTv02	68.72	69.06	69.19	68.71
Arabertv02-twitter	70.91	72.13	71.86	70.89
QARIB	67.72	68.03	68.16	67.71
MarBERT	70.91	72.13	71.86	70.89
MARBERTv2	71.71	72.94	72.67	71.69
CAMeLBERT-DA	**72.11**	**72.75**	**72.79**	**72.11**
CAMeLBERT-MIX	70.31	72.55	71.64	70.20

Table 5: Results of various models trained and tested on our dataset

The selected models were evaluated using different metrics which are accuracy, precision, recall and F1 scores as shown in Table 5. The results of the four metrics show the performance of each model on the testing set. CAMeLBERT-DA

[6] https://huggingface.co/aubmindlab/bert-base-arabertv02

[7] https://huggingface.co/aubmindlab/bert-base-arabertv02-twitter

[8] https://huggingface.co/qarib/bert-base-qarib

[9] https://huggingface.co/UBC-NLP/MARBERT

[10] https://huggingface.co/UBC-NLP/MARBERTv2

[11] https://huggingface.co/CAMeL-Lab/bert-base-arabic-camelbert-da

[12] https://huggingface.co/CAMeL-Lab/bert-base-arabic-camelbert-mix

[13] https://scikit-learn.org

obtained the highest scores in all specified metrics followed by MARBERTv2, this is mainly due to the increase of the vocabulary size and the amount and type of the training data i.e. dialectal Arabic (DA). The performance of Arabertv02-twitter and MarBERT were the same in all metrics. The CAMeLBERT-MIX model came next with relatively small difference, lastly QARIB came with the lowest scores in all metrics.

Comparing our task to other similar works are not feasible as they use different datasets and different languages. Since our task is the only one using this dataset which was collected for this task with the goal of classifying humor in Arabic. Therefore, we developed a baseline model using different measures.

5 Challenges and Issues

As mentioned early the task of humor detection is challenging in general and it is even harder in Arabic language due to the nature of the language. In general, detecting humor needs more knowledge to fully understand it (Khandelwal et al. 2018). The nature of humor also is a challenge by itself as people perceived humors differently, what might be funny to one might not be for another. Lacking a benchmark for humor detection make it difficult to evaluate models' performance. Moreover, one of the main causes of the lag in progress on humor research is the scarcity of public datasets (Hossain et al. 2019).

For Arabic language, (Biniz et al. 2018) stated that analyzing and automating tasks in Arabic is difficult than other languages due to: its morphological richness, its complex syntax, and its difficult semantics. The main challenge of our work was having multiple dialects that are different in syntax that make the detection harder.

6 Conclusion

This study discussed the methods we used to collect and construct humor dataset of different Arabic dialects and MSA. The dataset contained 10039 tweets and was annotated with two labels humor and non-humor. Three annotators have manually annotated the corpus, and Fleiss's Kappa was calculated to ensure the quality of the annotation process. Also, we carried out a set of experiments using a number of PLMs on the dataset to test it. The best model was CAMeLBERT-DA with an accuracy and an F1 score of 72.11%, which indicates potential for improvement and reflects the problematic nature of the humor dataset and the challenge of humor detection in general.

As a future work, we plan to understand and identify which features are essential for humor detection that contribute to enhancing the models prediction and troubleshooting unexpected model outputs.

References

Altin, Lutfiye Seda Mut. 2019. "LaSTUS/TALN at HAHA: Humor Analysis Based on Human Annotation." 6.

Amin, Miriam, and Manuel Burghardt. 2020. "A Survey on Approaches to Computational Humor Generation." Pp. 29–41 in *Proceedings of the The 4th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*. Online: International Committee on Computational Linguistics.

Annamoradnejad, Issa, and Gohar Zoghi. 2020. *ColBERT: Using BERT Sentence Embedding for Humor Detection*.

Banikalef, Eddin Abdullah Ahmed. 2014. "Linguistic Analysis of Humor in Jordanian Arabic among Young Jordanians Facebookers." Retrieved December 8, 2021 (https://www.semanticscholar.org/paper/Linguistic-Analysis-of-Humor-in-Jordanian-Arabic-Banikalef/e6aaed4aeb4e627678b62203791d959c792dd625).

Bertero, Dario, and Pascale Fung. 2016. "Multimodal Deep Neural Nets for Detecting Humor in TV Sitcoms." Pp. 383–90 in *2016 IEEE Spoken Language Technology Workshop (SLT)*.

Biniz, Mohamed, Samir Boukil, Fatiha Adnani, Loubna Cherrat, and Abd Moutaouakkil. 2018. "Arabic Text Classification Using Deep Learning Technics." *International Journal of Grid and Distributed Computing* 11:103–14. doi: 10.14257/ijgdc.2018.11.9.09.

Bzour, Assistant Prof Anwar Fayez Al. 2021. "Arabic Humorous Texts: An Attempt to Analyze." *Review of International Geographical Education Online* 11(4):1480-1492.

Castro, Santiago, Matías Cubero, Diego Garat, and Guillermo Moncecchi. 2016. “Is This a Joke? Detecting Humor in Spanish Tweets.” *ArXiv:1703.09527 [Cs]* 10022:139–50. doi: 10.1007/978-3-319-47955-2_12.

Fan, Xiaochao, Hongfei Lin, Liang Yang, Yufeng Diao, Chen Shen, Yonghe Chu, and Yanbo Zou. 2020. “Humor Detection via an Internal and External Neural Network.” *Neurocomputing* 394. doi: 10.1016/j.neucom.2020.02.030.

Fleiss, Joseph L. 1971. “Measuring Nominal Scale Agreement among Many Raters.” *Psychological Bulletin* 76(5):378–82. doi: 10.1037/h0031619.

Hasan, Md Kamrul, 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.” *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* 2046–56. doi: 10.18653/v1/D19-1211.

Hossain, Nabil, John Krumm, and Michael Gamon. 2019. “‘President Vows to Cut <Taxes> Hair’: Dataset and Analysis of Creative Text Editing for Humorous Headlines.” *ArXiv:1906.00274 [Cs]*.

Kao, Justine T., R. Levy, and Noah D. Goodman. 2016. “A Computational Model of Linguistic Humor in Puns.” *Cogn. Sci.* doi: 10.1111/cogs.12269.

Khandelwal, Ankush, Sahil Swami, Syed S. Akhtar, and Manish Shrivastava. 2018. “Humor Detection in English-Hindi Code-Mixed Social Media Content : Corpus and Baseline System.” *ArXiv:1806.05513 [Cs]*.

Landis, J. Richard, and Gary G. Koch. 1977. “The Measurement of Observer Agreement for Categorical Data.” *Biometrics* 33(1):159. doi: 10.2307/2529310.

Li, Da, Rafal Rzepka, Michal Ptaszynski, and Kenji Araki. 2020. “HEMOS: A Novel Deep Learning-Based Fine-Grained Humor Detecting Method for Sentiment Analysis of Social Media.” *Information Processing & Management* 57:102290. doi: 10.1016/j.ipm.2020.102290.

Meaney, J. A., Steven Wilson, Luis Chiruzzo, Adam Lopez, and Walid Magdy. 2021. “SemEval 2021 Task 7: HaHackathon, Detecting and Rating Humor and Offense.” Pp. 105–19 in *Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021)*. Online: Association for Computational Linguistics.

Meyer, John C. 2000. “Humor as a Double-Edged Sword: Four Functions of Humor in Communication.” *Communication Theory* 10(3):310–31. doi: 10.1111/j.1468-2885.2000.tb00194.x.

Nayef, Heba, and Mohamed El-Nashar. 2014. “‘Dissecting the Poisoned Honey’ Sexist Humor in Egypt: A Linguistic Analysis of Sexism in Colloquial Cairene Arabic Jokes.” *Anàlisi* 131. doi: 10.7238/a.v0i50.2324.

Omar, Hidaya Moulay. 2017. “A Linguistic Analysis of Humour in Algerian Arabic : The Case of ‘Sultan Ashur X’ Sitcom.” Thesis, Yarmouk University.

Zhang, Dongyu, Heting Zhang, Xikai Liu, Hongfei Lin, and Feng Xia. 2019. “Telling the Whole Story: A Manually Annotated Chinese Dataset for the Analysis of Humor in Jokes.” Pp. 6402–7 in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Hong Kong, China: Association for Computational Linguistics.

Zhang, Renxian, and Naishi Liu. 2014. “Recognizing Humor on Twitter.” Pp. 889–98 in *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, *CIKM ’14*. New York, NY, USA: Association for Computing Machinery.