# You Don't Say. . .
## Linguistic Features in Sarcasm Detection

Martina Ducret      Lauren Kruse      Carlos Martinez      Anna Feldman      Jing Peng

Montclair State University
Montclair, New Jersey, USA
fducretm1,krusel1,martinezcl1,feldmana,pengjg@montclair.edu

## Abstract

We explore linguistic features that contribute to sarcasm detection. The linguistic features that we investigate are a combination of text and word complexity, stylistic and psychological features. We experiment with sarcastic tweets with and without context. The results of our experiments indicate that contextual information is crucial for sarcasm prediction. One important observation is that sarcastic tweets are typically incongruent with their context in terms of sentiment or emotional load.

## 1 Introduction

Sarcasm, or verbal irony, is a figurative language device employed to convey the opposite meaning of what is actually being said. In verbal communication, a pause, intonation, or look can provide the cues necessary to determine whether there is sarcastic intent behind a comment. In writing, these social cues are inaccessible. Thus, we must rely on our understanding of the world, the speaker, and the context beyond the statement to discern between sarcasm and sincerity. This task has proven to be so subjective that social media users moderate their own comments using symbols and hashtags such as /s and #sarcasm to denote the sentiment on Reddit and Twitter, respectively. In fact, the dataset used in this paper was collected using such hashtags (Ghosh et al., 2020).

For machines, the lack of real-word knowledge is detrimental to their understanding of sarcasm as it hinders many natural language processing applications. Beyond social-media conversations, assessing product reviews as positive or negative requires an understanding of both rhetorical and literary devices. Back in 2012, BIC rolled out a "For

Her line of pens which led their intended female audience to poke fun at the misogynist message of the product. One reviewer commented, "Well at last pens for us ladies to use. . . now all we need is "for her" paper and I can finally learn to write!". While this review seems positive and gave the product four stars, our understanding of the social climate today leads us to conclude that this review is sarcastic and should be classified as such.

In social media communication, new slang words are introduced every day and emojis are often used to negate the sentiment of the text. In addition, stylistic devices and stylometric features are also often employed to convey a meaning opposite from its literal interpretation. While deep learning models can be very effective in their detection of sarcasm, they provide a "black box approach that gives linguists little to no insight into what features are characteristic of sarcasm. The purpose of the current work is to learn linguistic patterns associated with sarcastic tweets and their contexts and determine which are the strongest indicators of sarcasm. The next step is to combine these observations with transformer-based architectures to achieve a better prediction accuracy.

## 2 Previous Work

The field of automatic sarcasm recognition has become quite active in recent years. The most current event is the shared task (Ghosh et al., 2020) organized as a part of the 2nd FigLang workshop at ACL 2020. The task is typically framed as a binary classification task (sarcastic vs. non-sarcastic) considering either an utterance in isolation or in combination with contextual information. Early approaches to automatic sarcasm detection rely on different types of features, including sarcasm markers, word embeddings, emoticons, patterns between positive and negative sentiment (e.g., Davi-

dov et al. 2010; Tsur et al. 2010; González-Ibáñez et al. 2011; Riloff et al. 2013; Maynard and Greenwood 2014; Wallace et al. 2015; Ghosh et al. 2015; Joshi et al. 2015; Veale and Hao 2010; Liebrecht et al. 2013). Buschmeier et al. (2014) explore a range of features, mainly focused on sentiment, for the detection of verbal irony in product reviews. While this paper provides a good baseline for irony classification, our data differs in that it includes a multi-speaker thread of context prior to the sarcastic remark. More recent approaches apply deep learning methods (e.g., Ghosh and Veale 2016; Tay et al. 2018; Wallace et al. 2015). There is a great amount of research exploring the role of contextual information for sarcasm detection (e.g., Joshi et al. 2015; Bamman and Smith 2015; Misra and Arora 2019; Bamman and Smith 2015; Khattri et al. 2015; Amir et al. 2016; Rajadesingan et al. 2015; Ghosh and Veale 2017; Schifanella et al. 2016; Cai et al. 2019; Castro et al. 2019). Ghosh et al. (2020) report that almost all systems submitted as part of the shared task have used the transformer architecture, such as BERT (Turc et al. 2019) or RoBERTa (Liu et al. 2020), and other variants. They performed better than RNN architectures, even without any task specific fine-tuning. Unfortunately, it is difficult to interpret what these models capture about sarcastic tweets and their context. Our approach uses classical supervised algorithms to better understand which elements characterize sarcasm in a social media setting. We categorize linguistic features, experiment with different combinations, and take context into account when performing our experiments.

## 3   Our Approach

Our approach utilizes a combination of complex, stylometric, and psychological linguistic features to automatically detect the presence or absence of sarcasm in a given text. We intentionally experiment with classical machine learning classification algorithms to get a better understanding of the linguistic features contributing to the sarcasm detection task. Our linguistic intuition is that there will be a discordance between the linguistic features corresponding with the responses and contexts labeled as sarcastic. Sarcastic tweets are likely to be semantically or emotionally incongruent with their preceding tweets, while non-sarcastic tweets show a greater harmony with their context. To measure the emotional load of a response and its context,

we extract a number of sentiment- and emotion-related features. We also look at the distribution of these features across the two classes. Furthermore, we test the performance of our classifier and importance of our features by considering just the response tweet versus the response with its accompanying context.

## 4   Data Set

We use the Twitter Corpus from the CodaLab shared task on sarcasm detection (Ghosh et al., 2020). The training data consists of 2,500 tweets labeled 'SARCASM and 2,500 tweets labeled 'NON SARCASM, the balanced test data consists of an additional 1,800 labeled tweets. Ghosh et al. (2020), this is a self-labeled data set where the tweets are annotated as sarcastic based on the hashtags used by the users. The non-sarcastic tweets are the ones that do not contain the sarcasm hashtags, but may be labeled with either positive or negative sentiment hashtags, such as '#happy'. Retweets, duplicates, quotes, etc., are excluded (see Ghosh et al. 2020 for more details). Each sarcastic and non-sarcastic tweet is accompanied with an hierarchical conversation thread, e.g., context/1 is the immediate context, context/0 is the context that preceded context/1, and so on. The training and test data include up to 19 preceding tweets labeled as context/0, context/1, . . . , context/19 (if available).

## 5   Feature Extraction

Our research focuses on the role linguistic features play in sarcasm detection. We classify our features into three categories: complexity, stylistic, and psychological. Abonizio et al. (2020) defines complexity features as linguistic features that capture the overall objective of the context at the word and sentence level. Stylistic features use natural language techniques to gain grammatical information to better understand the syntax and style of the document. Psychological features are closest related to emotions and the cognitive aspect of NLP. We expand on these psychological features by utilizing VAD (Valence, Arousal, Dominance) (Warriner et al., 2013), emotional embeddings, and LIWC (Tausczik and Pennebaker, 2010) . Lastly, we use word-level count vectors, word-level tf-idf, n-gram word-level tf-idf, n-gram character-level tf-idf. We stack these features and refer to them as count vectors for the remainder of this paper.

## 5.1 LIWC

LIWC (Tausczik and Pennebaker, 2010) is a text analysis program with a built-in dictionary that counts words in psychologically meaningful categories. After all the words have been reviewed, the module calculates the total percentages of words that are similar and match that of the user dictionary categories. We used LIWC to extract features to detect and categorize the meaning, emotional sentiment, and social relationship of the words in the data set.

## 5.2 Valence, Arousal, Dominance (VAD)

VAD (Valence Arousal Dominance) (Warriner et al., 2013) includes almost 14,000 lemmas rated on a 1-9 scale according to the emotions evoked by the terms. Valence refers to the pleasantness of the word, arousal determines how dull or exciting the emotion is, and dominance ranges from submission to feeling in control. The VAD dimensions allow us to further explore the affective meanings of tweets and determine their viability as a predictor of sarcasm. We compute VAD scores for each "response" and use the three scores obtained as a feature in our classifiers. Furthermore, we explore using the scores as a measure of congruity between our response and contexts. We calculate the VAD scores for each individual response and context and then subtract the response scores by their respective context scores. In other words, if a response receives a valence score of 8 and its context/0 receives a valence score of 2, the valence congruity score would be a 6. We hypothesize that sarcastic tweets might show very little affective congruity compared to their non-sarcastic counterparts.

## 5.3 VADER

VADER (Valence Aware Dictionary and sEntiment Reasoner) (Hutto and Gilbert, 2015) is a lexicon and rule-based tool built especially for sentiment analysis of social media texts. VADER maps lexical features to emotions and provides insight into the intensity of such emotions through a series of polarity indices. VADER considers capitalization, punctuation, degree modifiers, emojis, and negations to compute its negative, positive and neutral scores. Furthermore, VADER's compound score provides a normalized, weighted composite score for a given tweet.

| | | | Context/0 | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | Anger | Fear | Joy | Love | Sadness | Surprise | Total |
| Response | Not- Sarcasm | Anger | 221 | 35 | 176 | 11 | 69 | 6 | 518 |
| | | Fear | 48 | 27 | 51 | 1 | 25 | 4 | 156 |
| | | Joy | 279 | 72 | 898 | 34 | 131 | 20 | 1434 |
| | | Love | 4 | 2 | 48 | 7 | 5 | 2 | 68 |
| | | Sadness | 86 | 24 | 101 | 5 | 60 | 9 | 285 |
| | | Surprise | 16 | 3 | 13 | 0 | 4 | 3 | 39 |
| | | Total | 654 | 163 | 1287 | 58 | 294 | 44 | 2500 |
| | Sarcasm | Anger | 498 | 120 | 345 | 17 | 116 | 26 | 1122 |
| | | Fear | 83 | 33 | 61 | 3 | 12 | 6 | 198 |
| | | Joy | 333 | 72 | 289 | 13 | 96 | 22 | 825 |
| | | Love | 9 | 3 | 6 | 1 | 3 | 1 | 23 |
| | | Sadness | 98 | 24 | 88 | 2 | 41 | 3 | 256 |
| | | Surprise | 30 | 10 | 28 | 2 | 5 | 1 | 76 |
| | | Total | 1051 | 262 | 817 | 38 | 273 | 59 | 2500 |

Figure 1: Distribution of Emotions for Response vs. Context/0 in the training data.

## 5.4 Emotional Embeddings

The emotions conveyed in our data set are portrayed through emotional embeddings. Calculating the emotions of the text goes a level deeper than just looking at the word embeddings. Using a pre-trained model from Hugging Face (Saravia et al., 2018), we categorize the tweets into six emotions. The emotions include, joy, anger, fear, surprise, sadness and love. Figure 1 above represents an example of the distribution of emotions between response and context/0 in the balanced training data set. The results support our intuition that sarcasm is typically associated with negative emotions. When the context is labeled as "anger", non-sarcastic tweets tend to respond with joy, while sarcastic tweets usually respond with anger. By contrast, when the context is labeled as "joy", non-sarcastic tweets overwhelmingly respond with joy, while sarcastic tweets still largely respond with anger. There are 1,216 instances of the same emotion expressed in both response and context for the non- sarcasm class and 863 instances of this in the sarcasm class. Sarcastic tweets are generally incongruent with emotions throughout the response and context, unless associated with a negative emotion, e.g., anger.

## 5.5 Tweet-Context Similarity Scores

We use the standard document similarity estimation technique using word embeddings (GloVe, Pennington et al. 2014) and emotional embeddings (Saravia et al. 2018), which consists of measuring the similarity between the vector representations of the two documents. Let $x_1; \ldots x_m$ and $y_1; \ldots; y_n$ be the emotion (or word embedding) vectors of two documents. The cosine similarity value between the two documents (e.g., a tweet and its context)

centroids $C_x = \frac{1}{m}\sum_{i=1}^{m} x_i$ and $C_y = \frac{1}{n}\sum_{i=1}^{n} y_i$ is calculated as follows:

$$\cos(C_x; C_y) = \frac{\langle C_x; C_y \rangle}{\|C_x\| \|C_y\|};  \qquad (1)$$

where $\langle x; y \rangle$ denotes the inner product of two vectors $x$ and $y$.

We compute two similarity scores: 1) semantic cosine similarity using word embeddings; 2) cosine similarity using emotional embeddings. Our linguistic intuition is that a sarcastic response is going to be semantically or emotionally incongruent with its context and this is what creates the sarcasm effect.

| | Message |
|---|---|
| c/0 | Its no secret that this president has routinely targeted religious and ethnic minorities. He has fanned the flames of hate against refugees, Muslims, Africans, immigrants, women and all racial and religious minorities. |
| c/1 | He is routinely and openly hostile to any legitimate Congressional oversight. He has made clear his wanton corruption by soliciting a bribe from a foreign government for his personal political gain. |
| R | Yassss queen, youre so brave and bold. |

Table 1: Sarcastic Tweet.
Response=R; Context0=C/0; Context1=C/1

| | Message |
|---|---|
| c/0 | A2 I revert back to Canvas. I am sure you can post assignments for parents in this, (haven't done this yet). Canvas = #thebomb #KidsDeserveIt |
| c/1 | Can you telk me more about Canvas? I haven't heard of it. |
| R | It's Edmodo with #MorePower You can create assignments in it, post all work, the assignments can be auto graded and imported into your Skyward grade book. |

Table 2: Non Sarcastic Tweet.
Response=R; Context0=C/0; Context1=C/1

Table 1 is an example of a sarcastic tweet whose context/0, context/1 and response received an emotion of anger, anger, and joy, respectively. Table 2 represents a non-sarcastic thread of tweets where each message was classified as joy. This indicates that non-sarcastic tweets tend to be more emotionally similar to the preceding context while sarcastic tweets tend to shift in emotion. As a result, when compared to its contexts, the sarcastic tweet received lower emotional similarity scores than the non-sarcastic tweet.

## 5.6 Feature Analysis

After running all of the features on the training data, we implemented SHAP (SHapley Additive exPlanations) (Lundberg and Lee, 2017) to determine which features are the most important for classification. SHAP is a theoretic output technique that explains predictions of our model, by producing a SHAPLEY score that plots the most important features in our model. The features produced by SHAP were used in our experiments and are referred to as our "select linguistic features". The top 20 features SHAP selects contain a combination of character features such as character count, as well as a number of sentiment features, including VADER scores, emotion scores for both a response and its context as well as VAD features.

## 6 Experimental Evaluation

### 6.1 Data Preprocessing

Our preprocessing procedure consists of steps to remove noisy and unnecessary data. First, we tokenize and lemmatize the tweets using NLTK (Loper and Bird, 2002). We also remove any instance of "@USER" due to the repetition of this token in the beginning of most tweets. Prior research demon-strated that classifiers did not tend to benefit from large quantities of additional context and we no-ticed that a majority of the tweets only contained context/0 and context/1. While we plan to experi-ment further with additional context layers, in this work we only report on experiments that involve context/0 and context/1. We did not remove any stop words due to the small amount of text in each tweet. We also maintained punctuation and emojis as they proved to be useful information during the extraction of certain features, such as VADER.

## 7 Results

We use a Random Forest classifier and run 21 different experiments of which the most relevant ones are outlined in Table 3. The baseline scores represent an attention based LSTM model described in Ghosh et al. (2018) and used in the CodaLab Shared Task. We look at how each feature performed on just the response versus the response and context. We notice that for response, a combination of all count features and all linguistic features achieves the best F1 score of 67%. This score is further increased to 70% when the context is considered.

| Experiments | P | R | A | F1 |
|---|---|---|---|---|
| Baseline 1 Shared Task | 70% | 66.9% | N/A | 68% |
| R (All Ct. Ft.) | 72% | 61% | 63% | 66% |
| R(All Ling Ft.) | 56% | 60% | 59% | 52% |
| R(Sel. Ling Ft.) | 54% | 60% | 59% | 57% |
| R(All Ling Ft. +All Ct.) | 70% | 64% | 65% | 67% |
| R(Sel. Ling Ft. +All Ct.) | 65% | 51% | 51% | 57% |
| R+C/0+C/1 (All Ct. Ft) | 67% | 61% | 62% | 64% |
| R+C/0+C/1(All Ling Ft.) | 64% | 60% | 60% | 53% |
| R+C/0+C/1(Sel. Ling Ft.) | 71% | 60% | 62% | 64% |
| R+C/0+C/1(All Ling Ft. +All Ct.) | 80% | 62% | 66% | 70% |
| R+C/0+C/1(Sel. Ling Ft. +All Ct.) | 70% | 65% | 66% | 67% |

Table 3: Random Forest; Various feature combinations. Response=R; Context0=C/0; Context1=C/1; Count=Ct; Linguistics= Ling; Features= ft.

## 8 Conclusion

In this paper we explored the role various linguistic features play in computational sarcasm detection. We investigated a combination of text and word complexity features, stylistic and psychological features. The result of our experiments indicate that contextual information is crucial for sarcasm detection. We also observed that sarcastic tweets are often incongruent with their context in terms of sentiment or emotional load. Using a Random Forest classifier and the features we extracted we obtain promising results. Our current work is concerned with combining these observations with transformer-based architectures to achieve a better prediction accuracy.

## Acknowledgments

## References

Hugo Queiroz Abonizio, Janaina Ignacio de Morais, Gabriel Marques Tavares, and Sylvio Barbon Junior. 2020. Language-independent fake news detection: English, portuguese, and spanish mutual features. Future Internet, 12(5):87.

Silvio Amir, Byron C Wallace, Hao Lyu, and Paula Carvalho Mário J Silva. 2016. Modelling context with user embeddings for sarcasm detection in social media. arXiv preprint arXiv:1607.00976.

David Bamman and Noah Smith. 2015. Contextualized sarcasm detection on twitter.

Konstantin Buschmeier, Philipp Cimiano, and Roman Klinger. 2014. An impact analysis of features in a classification approach to irony detection in product reviews. In Proceedings of the 5th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, pages 42–49, Baltimore, Maryland. Association for Computational Linguistics.

Yitao Cai, Huiyu Cai, and Xiaojun Wan. 2019. Multimodal sarcasm detection in twitter with hierarchical fusion model. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 2506–2515.

Santiago Castro, Devamanyu Hazarika, Verónica Pérez-Rosas, Roger Zimmermann, Rada Mihalcea, and Soujanya Poria. 2019. Towards multimodal sarcasm detection (an _obviously_perfect paper). arXiv preprint arXiv:1906.01815.

Dmitry Davidov, Oren Tsur, and Ari Rappoport. 2010. Semi-supervised recognition of sarcasm in Twitter and Amazon. In Proceedings of the Fourteenth Conference on Computational Natural Language Learning, pages 107–116, Uppsala, Sweden. Association for Computational Linguistics.

Aniruddha Ghosh and Tony Veale. 2016. Fracking sarcasm using neural network. In Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, pages 161–169, San Diego, California. Association for Computational Linguistics.

Aniruddha Ghosh and Tony Veale. 2017. Magnets for sarcasm: Making sarcasm detection timely, contextual and very personal. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pages 482–491.

Debanjan Ghosh, Alexander R Fabbri, and Smaranda Muresan. 2018. Sarcasm analysis using conversation context. Computational Linguistics, 44(4):755–792.

Debanjan Ghosh, Weiwei Guo, and Smaranda Muresan. 2015. Sarcastic or not: Word embeddings to predict the literal or sarcastic meaning of words. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pages 1003–1012, Lisbon, Portugal. Association for Computa-tional Linguistics.

Debanjan Ghosh, Avijit Vajpayee, and Smaranda Muresan. 2020. A report on the 2020 sarcasm detection shared task. In Proceedings of the Second Workshop on Figurative Language Processing, pages 1–11, Online. Association for Computational Linguistics.

Roberto Gonzalez-Ibanez, Smaranda Muresan, and Nina Wacholder. 2011. Identifying sarcasm in Twitter: A closer look. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, pages 581–586, Portland, Oregon, USA. Association for Computational Linguistics.

C.J. Hutto and Eric Gilbert. 2015. Vader: A parsimonious rule-based model for sentiment analysis of social media text.

Aditya Joshi, Vinita Sharma, and Pushpak Bhattacharyya. 2015. Harnessing context incongruity for sarcasm detection. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers), pages 757–762, Beijing, China. Association for Computational Linguistics.

Anupam Khattri, Aditya Joshi, Pushpak Bhattacharyya, and Mark Carman. 2015. Your sentiment precedes you: Using an authors historical tweets to predict sarcasm. In Proceedings of the 6th workshop on computational approaches to subjectivity, sentiment and social media analysis, pages 25–30.

Christine Liebrecht, Florian Kunneman, and Antal van den Bosch. 2013. The perfect solution for detecting sarcasm in tweets #not. In Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, pages 29–37, Atlanta, Georgia. 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. 2020. RoBERTa: A Robustly Optimized BERT Pretraining Approach.

Edward Loper and Steven Bird. 2002. NLTK: The Natural Language Toolkit. In In Proceedings of the ACL Workshop on Effective Tools and Methodolo-gies for Teaching Natural Language Processing and Computational Linguistics. Philadelphia: Associa-tion for Computational Linguistics.

Scott M Lundberg and Su-In Lee. 2017. A unified approach to interpreting model predictions. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, Advances in Neural Information Processing Systems 30, pages 4765–4774. Curran Associates, Inc.

Diana Maynard and Mark Greenwood. 2014. Who cares about sarcastic tweets? investigating the impact of sarcasm on sentiment analysis. In Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), pages 4238–4243, Reykjavik, Iceland. European Language Resources Association (ELRA).

Rishabh Misra and Prahal Arora. 2019. Sarcasm detection using hybrid neural network. ArXiv, abs/1908.07414.

Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. Glove: Global vectors for word representation. In Empirical Methods in Natural Language Processing (EMNLP), pages 1532–1543.

Ashwin Rajadesingan, Reza Zafarani, and Huan Liu. 2015. Sarcasm detection on twitter: A behavioral modeling approach. In Proceedings of the eighth ACM international conference on web search and data mining, pages 97–106.

Ellen Riloff, Ashequl Qadir, Prafulla Surve, Lalindra De Silva, Nathan Gilbert, and Ruihong Huang. 2013. Sarcasm as contrast between a positive sentiment and negative situation. In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, pages 704–714, Seattle, Washington, USA. Association for Computational Linguistics.

Elvis Saravia, Hsien-Chi Toby Liu, Yen-Hao Huang, Junlin Wu, and Yi-Shin Chen. 2018. CARER: Contextualized affect representations for emotion recognition. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pages 3687–3697, Brussels, Belgium. Association for Computational Linguistics.

Rossano Schifanella, Paloma de Juan, Joel Tetreault, and Liangliang Cao. 2016. Detecting sarcasm in multimodal social platforms. In Proceedings of the 24th ACM international conference on Multimedia, pages 1136–1145.

Yla R. Tausczik and James W. Pennebaker. 2010. The psychological meaning of words: Liwc and computerized text analysis methods. Journal of Language and Social Psychology, 29(1):24–54.

Yi Tay, Anh Tuan Luu, Siu Cheung Hui, and Jian Su. 2018. Reasoning with sarcasm by reading in-between. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 1010–1020, Mel-bourne, Australia. Association for Computational Linguistics.

Oren Tsur, Dmitry Davidov, and Ari Rappoport. 2010. Icwsm - a great catchy name: Semi-supervised recognition of sarcastic sentences in online product reviews. In ICWSM. The AAAI Press.

Iulia Turc, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Well-read students learn better: On the importance of pre-training compact models. arXiv preprint arXiv:1908.08962v2.

Tony Veale and Yanfen Hao. 2010. Detecting ironic intent in creative comparisons. In ECAI 2010 - 19th European Conference on Artificial Intelligence, Lisbon, Portugal, August 16-20, 2010, Proceedings, volume 215 of Frontiers in Artificial Intelligence and Applications, pages 765–770. IOS Press.

Byron C. Wallace, Do Kook Choe, and Eugene Charniak. 2015. Sparse, contextually informed models for irony detection: Exploiting user communities, entities and sentiment. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint

Conference on Natural Language Processing (Volume 1: Long Papers), pages 1035–1044, Beijing, China. Association for Computational Linguistics.

Amy Warriner, Victor Kuperman, and Marc Brysbaert. 2013. Norms of valence, arousal, and dominance for 13,915 english lemmas. Behavior research methods, 45.