A Corpus Study with German Data Sets into the Similarity of Irony and Satire

Marisa Schmidt
Faculty of Computer Science
University Koblenz
Universitätsstr. 1, 56070 Koblenz, Germany
marisaschmidt@uni-koblenz.de

Karin Harbusch
Faculty of Computer Science
University Koblenz
Universitätsstr. 1, 56070 Koblenz, Germany
harbusch@uni-koblenz.de

Abstract— **In deception detection, i.e., the falsification of news, satire detection is an import research area. This work strives for high accuracy in satire detection. We want to answer the question whether irony detection can serve the purpose of satire detection as well, or even better than specialized satire classification. The hypothesis underlying this claim follows the definition that satire is a genre that uses irony. Thus, we argue that irony should be indicative in a satire dataset. We contrast the results of runs with irony and satire annotated corpora with *Elmo4Irony,* an existing classifier for irony, and *Adversarial Satire*, an existing system for satire detection. In our evaluation, we use three different German data sets labeled with irony and satire, respectively. Our study corroborates the claim. Irony can indeed be found in a satirical dataset—even with higher accuracy. In order to supplement the finding, both systems are evaluated with typical examples from satire papers for deeper exploration. Unexpectedly, for the examples from the scholarly literature, both systems can hardly distinguish between irony/satire and neutral formulations.**

Keywords-Satire; Irony, Fake News; Deception Detection.

I. Introduction

Deception detection, i.e., the falsification of news in journalistic articles or social media, has become an increasingly important topic [1]. Another widely used term for false news is *fake news*. According to Zhou and Zafarani [2], it is now viewed as one of the greatest threats to democracy, journalism, and freedom of expression.

One strand of deception-detection research deals with *satire* detection. In the scholarly literature, definitions of satire can vary considerably (see, e.g., [3] and [4]). According to *The Oxford English Dictionary* [5], *satire* is "the use of humor, irony, exaggeration or ridicule to expose and criticize people's stupidity or vices." In other definitions, satire is tried to be demarcated from *irony.* In the comparison of definitions from the literature, Singh states: "Satire and irony are often closely related, but there are important distinctions between the two. As form of criticism, satire uses humor to accomplish its goals. One technique that satire uses is irony. Irony focuses on the discrepancies between what is said or seen and what is actually meant. Simply, satire and irony hardly differ because one, satire, often uses the other, irony." [6].

The Oxford English Dictionary defines *irony* as "the expression of meaning through the use of language signifying the opposite, typically for humorous effect." In the context of opinion analysis [7], Karoui and colleagues characterize irony as follows: "Irony denotes a discrepancy between discourse and reality, between two realities or, more generally, between two perspectives to incongruous effect."

Another term in this context often used as label in data sets is sarcasm. Karoui and colleagues demarcate it from irony as follows: "According to the Oxford English Dictionary, sarcasm is "the use of irony to mock or convey contempt". The utterance is bitter in nature and is intended to hurt the target [29]. Sarcasm is thus characterized by aggression, although not to the exclusion of mockery or teasing. Sarcasm is considered as a combination of the processes involved in both humor and irony, but is hurtful and overtly mocks the target. [...] Sarcasm is thus associated with aggression, insult and nastiness, traits that are not present in irony."

Given the subtle differences between the individual figurative language phenomena (cf. [7]), we want to explore whether comparative runs with the same labeled data sets but specific satire and irony detectors help to identify essential features that can lead to improved satire-classification results. We deploy *Adversarial Satire* [8] and *Elmo4Irony* [9] as prototypical detection components for satire and irony, respectively (see Sections III and IV for details). We run our study with the four German data sets outlined in Section II.

In our corpus study, we want to quantify how much irony can be detected in a satire annotated data set. As outlined above, satire is a genre that uses irony and therefore irony should be found in a satire dataset, i.e., irony detection is highly indicative to satire as well. Our study corroborates the claim. Irony can indeed be found in a satirical dataset—even with higher accuracy. In our comparative runs (cf. Section V), we illustrate that the irony detector Elmo4Irony performs better than the specialized satire classifier Adversarial Satire. As supporting evidence, we collected a small number of typical examples underpinning the different irony definitions. In order to make up a corpus, we add neutral facts (28.57% irony). However, for the examples from the scholarly literature, both systems can hardly distinguish between irony and neutral formulations. The implications from this unexpected finding require deeper inspection that is subject to further research (cf. Section VI).

The paper is organized as follows. In the next section, we present the four corpora used in our study. Sections III and IV elaborate on satire and irony detection, respectively. The results of our two experiments are presented in Section V. In the final section, we draw some conclusions.

II. Data sets

From public data collections, we use the three German data sets labelled with satire, irony, and sarcasm, respectively:

- C1: the satire data set by [8] with 329,859 articles from 15 different newspapers (2.82% satirical ones),
- C2: two subsets of a big Reddit corpus labeled for irony [10]: (C2a) *SARC 2.0* with 321,748 entries and (C2b) *SARC 2.0 pol* (17,074 entries), and
- C3: a Twitter data set from SemEval-2018 [11] that is labeled with #irony, #sarcasm and #not. The corpus provides 4,792 tweets, where both, irony and sarcasm, have a percentage of 50%.

In our study, we use only a subset of the satire corpus C1 (dubbed C1SUB) with 125 newspaper articles, 45 of which are satire (36%) to run it on a less powerful system compared to the settings in [8] (according to personal communication, their system has 256 gigabyte (GB) memory). With 60 GB, the classification accuracy with the reduction of the amount of data leads to comparable results with the numbers published in [8]. The other two corpora, i.e., C2a, C2b, and C3, are fully used in the study.

Moreover, we test all models with a newly set up corpus, called C4 here, that aims at a broad collection of prototypical examples from the irony literature used there to illustrate the definition (cf. example (1) in [12]).

(1) *Ich würde dieses Buch Freunden empfehlen, die an Schlaflosigkeit leiden oder die ich absolut verachte.*
'I would recommend this book to friends, who either suffer from insomnia or whom I despise.'

Although, we call C4 a 'corpus', we have to emphasize that it is still in its infancy. Currently, C4 comprises 10 ironic examples from different articles. Moreover, we thought up 5 ironic ones ourselves as a kind of control instance in contrast to the outstanding quality of the literature examples (cf. example (2)) and 6 neutral definitions of facts labelled not-ironic (cf. example (3)). The preliminary size does not create a problem here, for we use it as a kind of litmus test for the models only.

(2) *Oh ja! Du bist definitiv der klügste Mensch, den ich kenne!*
'Oh yes! You are definitively the most clever man I met.'

(3) *Gänseblümchen haben weiße Blüten.*
'Daisies have white blossoms.'

The evaluation with all four data sets is outlined in Section V. In the next two sections, we first sketch the satire and irony detection component, individually, before we employ both system with the four data sets in our study.

III. Satire detection

The challenging task of satire detection has been tackled from various points of view: lexically, syntactically, and semantically. Thu and Aung give an historical overview for systems from the different viewpoints [13].

Additionally, we cite more recent approaches here. McHardy and colleagues extend a satire detector with an adversarial component to control for the confounding variable of publication source [8]. The system, called Adversarial Satire, is based on Tensorflow [14] and uses Word2Vec embeddings [15], [16]. For the evaluation, the German satire corpus (dubbed C1 in Section II), was set up. Li and colleagues [17] propose a multimodal method for satire detection using textual and visual cues. Razali et al. [18] suggest a context-driven satire-detection component deploying Deep Learning.

In our study, we decided to use Adversarial Satire so that the original evaluation results for C1 can directly be compared with our implementation (the code can be found here: https://gitlab.uni-koblenz.de/marisaschmidt/irony-detection) running C1SUB (see Table I). We use the Linux [19] distribution Ubuntu [20]—deploying the CUDA 11— with 50 GB kernel memory plus 500 GB extra; the system runs on 4 CPUs and 1 GPU with 35 GB; this set up requires some smaller adaptions we skip here for reasons of space).

Table I illustrates the results for the smaller corpus C1SUB compared to the original results—for reasons of space, we only outline the results for one setting (confounding variable=0.0). For all settings in the overall evaluation, the quality favorably compares. Thus, we can use the component with the reduced corpus C1SUB in our study.

IV. Irony and Sarcasm detection

For a good overview on satire-detection systems, subdivided into surface and semantic approaches, as well as pragmatic ones, see [7]. Here, we cursorily sum up other approaches.

Ilić and colleagues propose a model that uses character-level vector representations of words, based on Embeddings from Language Model (ELMo [21]). The system is called ElMo4Irony [9]. Kumar and Harish propose to extract five sets of linguistic features fused with features selected using two stages of a feature selection method [22]. Lin and colleagues compare different machine-learning methods for irony detection [23]. Jiang and colleagues present an approach mainly based on fine-tuned BERT models using a Grid-Search and Data Augmentation with MLM (Masked Language Model) substitution [24] based on BERTimbau for smoothing the use of a small data set. Tomás and colleagues propose a transformer-based model for multimodal irony detection [25].

As stated in Section I, sarcasm and irony are closely related, i.e., are often judged to stand in a sub-super relationship. So, we round out our state of the art with a survey article on sarcasm detection: Joshi and colleagues describe various datasets, approaches, trends and issues [26].

TABLE I. EVALUATION OF ADVERSARIAL SATIRE WITH C1SUB

Data	*C1SUB*			*C1*		
	P	R	F1	P	R	F1
dev	0,999	0,667	0,799	0,989	0,526	0,687
test	0,818	0,643	0,719	0,990	0,501	0,665

Concentrating on irony here, we deploy Elmo4Irony in our study. The approach considers a wide variety of features (e.g., capitalizations or emoticons; cf. [7] for a study on the impact of syntactic, semantic and pragmatic features). ElMo4Irony uses PyTorch [27] and GloVe embeddings [28].

For Elmo4Irony, we also skip here the details of our implementation under the above-mentioned system settings. In Table II, we exemplarily sketch the results for dropout = 0.1 to demonstrate that the results favorably compare to the numbers in [9].

V. COMPARATIVE RUNS WITH THE DATA SETS

Two experiments are conducted using the systems deploying the data sets presented in the previous sections. *Experiment 1* is devoted to the research question of whether irony can also be found in a satire dataset. As follow-up question from the positive findings in Experiment 1, *Experiment 2* probes examples of irony given in the literature with the two systems, i.e., tests the models with C4.

As outlined in Section I, satire is defined as a genre that uses irony. This definition leads to the hypothesis that an irony detection system—in our case Elmo4Irony (cf. Section IV) — should find irony in the satire-data set. To test this hypothesis, Elmo4Irony and the satire classifier Adversarial Satire (cf. Section III) both employ the data set C1SUB.

Both methods are trained over 10 epochs with a batch size of 16. Elmo4Irony is trained with dropouts of 0.0, 0.1 and 0.5. For Adversarial Satire, different values for the adversarial weight are used: the confounding variable = 0.0, 0.2, 0.3 and 0.7. For these variable settings, Elmo4Irony performs always better than Adversarial Satire (for two exemplary variable setting, the overall results are outlined in Table III). In fact, the irony classifier provides better results on the satire dataset than the specialized satire classifier. This observation confirms the hypothesis of Experiment 1. Irony is an indicative feature to satire detection.

TABLE II. EVALUATION OF ELMO4IRONY WITH C2 AND C3

Data	*Our implementation*			*Original numbers*		
	P	R	F1	P	R	F1
C2a	0,707	0,704	0,703	0,760	0,760	0,760
C2b	0,687	0,686	0,685	0,720	0,720	0,720
C3	0,685	0,688	0,686	0,696	0,697	0,696

TABLE III. COMPARISON OF IRONY AND SATIRE DETECTION

Data	*Adversarial Satire*			Elmo4Irony		
	Confounding variable = 0.0			Confounding variable = 0.0		
	P	R	F1	P	R	F1
C1SUB	0,622	0,617	0,618	0,895	0,800	0,816
	Confounding variable = 0.7			Confounding variable = 0.1		
	P	R	F1	P	R	F1
C1SUB	0,708	0,617	0,603	0,857	0,867	0,839

In Experiment 2, both systems are evaluated on the new dataset C4 (the training of the models still happens on their regular datasets). C4 is labelled for irony. Based on Experiment 1, we argue that irony can serve as satire feature. However, it is less obvious that a satire classifier will find irony on irony data. It is therefore to be expected that Adversarial Satire will find less satire on this data set with ironic examples. Again, we tested the different values for the dropout in Elmo4Irony and the adversarial weight in Adversarial Satire. Table IV provides the numbers of samples that were correctly classified as irony (TP), wrongly classified as irony (FP), correctly classified as regular (TN) and wrongly classified as regular (FN). The numbers in brackets show the results probing additionally provided neutral text to obtain article length in C4 aiming at improving the quality of Adversarial Satire.

Interestingly, the results show that most models classify all examples as ironic. In the initial scenario, the Elmo4Irony model, which is trained with a dropout of 0.0, finds the least irony. However, it still classifies almost all the non-ironic examples as ironic, while the ironic examples are classified as non-ironic. A second Elmo4Irony model that correctly classifies at least one example as non-ironic is the model trained with a drop rate of 0.5.

Additionally, we tested Adversarial Satire (which was trained on whole articles instead of single sentences) with adding a neutral text to the example sentences. With this extended input, Elmo4Irony classifies everything as non-ironic with most variable settings. The only Adversarial Satire model that classifies one example as non-ironic in the scenario with additional text is the model trained with an adversarial weight of 0.2. This model correctly classifies one of our self-created neutral examples as non-ironic. Under the condition of no additional text, the same model also classifies one example as non-ironic, however, this is actually an ironic one. In essence, additional neutral text does not have a positive impact on the classification of adversarial satire.

In order to sum up the findings of Experiment 2, unexpectedly, the features calculated by both systems are not suitable for this new data set, as almost everything is classified as ironic. The small size of C4 cannot be the reason for failure given that the corpus is only used as test set. Deeper analysis of the features is required here (cf. [7] and [13]).

TABLE IV. EVALUATION OF C4

drop-out	TP	FP	TN	FN	***adv. weight***	TP	FP	TN	FN
0.0	3 (0)	5 (0)	1 (6)	12 (15)	0.0	15 (15)	6 (6)	0 (0)	0 (0)
0.1	15 (0)	6 (0)	0 (6)	0 (15)	0.2	14 (15)	6 (5)	0 (1)	1 (0)
0.5	15 (0)	5 (0)	1 (6)	0 (15)	0.3	15 (15)	6 (6)	0 (0)	0 (0)
					0.7	15 (15)	6 (6)	0 (0)	0 (0)

VI. CONCLUSIONS

We have presented the results of a corpus study into the relationship between satire and irony. Based on the definition that satire uses irony, we could verify that irony detection can serve as satire classification very well. Experiment 2 was designed to better understand the irony features. However, the results were unexpectedly poor. We plan to extend C4 to a full development/test corpus with a larger collection of examples from very divergent sources. The goal is to obtain a richer set of features to classify irony.

REFERENCES

[1] V. L. Rubin, N. J. Conroy, Y. Chen, and S. Cornwell, "Fake news or truth? Using satirical cues to detect potentially misleading news," Proc. NAACL-HLT, pp. 7–17, Jun. 2016.

[2] X. Zhou and R. Zafarani, "A survey of fake news: fundamental theories, detection methods, and opportunities," ACM Computing Surveys, vol. 53, no. 5, art. 109, Sept. 2020. https://dl.acm.org/doi/pdf/10.145/3395046 [retrieved: 23.03.23]

[3] L. Colletta, "Political satire and postmodern irony in the age of Stephen Colbert and Jon Stewart," The Journal of Popular Culture, vol. 42, no. 5, pp. 856-874, 2009. https://doi.org/10.1111/j.1540-5931.2009.00711.x [retrieved: 23.03.23]

[4] C. Condren, "Satire and definition," Humor, vol. 25, no. 4 , 2012, https://doi.org/10.1515/humor-2012-0019 [retrieved: 23.03.23]

[5] OED, Oxford Univ. Press, Oxford. https://www.oed.com/public/freeoed/loginpage [retrieved: 23.3.23]

[6] R. K. Singh, "Humour, irony and satire in literature," IJEL, vol. 3, no. 4, pp. 65-72, 2012. https://www.academia.edu/4541187/Humour_Irony_and_Satire_in_Literature [retrieved: 23.03.23]

[7] J. Karoui, F. Benamara, and V. Moriceau, "Automatic detection of irony: opinion mining in microblogs and social media," London: ISTE, 2019. https://doi.org/10.1002/9781119671183 [retrieved: 23.03.23]

[8] R. McHardy, H. Adel, and R. Klinger, "Adversarial training for satire detection: controlling for confounding variables," Proc. NAACL-HLT, pp. 660–665, Jun. 2019. https://doi.org/10.48550/arXiv.1902.11145 [retrieved: 23.03.23]

[9] S. Ilić, E. Marrese-Taylor, J. A. Balazs, and Y. Matsuo, "Deep contextualized word representations for detecting sarcasm and irony," Proc. 9th WASSA, pp. 2–7, Oct. 2018. https://aclanthology.org/W18-6202 [retrieved: 23.03.23]

[10] M. Khodak, N. Saunshi, and K. Vodrahalli, "A large self-annotated corpus for sarcasm," Proc. 11th LREC, pp. 641-646, May 2018. https://doi.org/10.48550/arXiv.1704.05579 [retrieved: 23.03.23]

[11] C. van Hee, E. Lefever, and V. Hoste, "SemEval-2018 task 3: irony detection in English tweets," Proc. 12th SemEval, pp. 39-50, Jun. 2018. http://dx.doi.org/10.18653/v1/S18-1005 [retrieved: 23.02.23]

[12] J. Ling, and R. Klinger, "An empirical, quantitative analysis of the differences between sarcasm and irony," Proc. *European* Semantic Web Conference, pp. 203-216, Jun. 2016. https://doi.org/10.1007/978-3-319-47602-5_39 [retrieved: 23.02.23]

[13] P. P. Thu and T. N. Aung. "Implementation of emotional features on satire detection," International Journal of Networked and Distributed Computing, Vol. 6, No. 2, pp. 78-87, 2018.

[14] https://www.tensorflow.org [retrieved: 23.02.23]

[15] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," 2013a. https://doi.org/10.48550/arXiv.1301.3781 [retrieved: 23.02.23]

[16] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," Proc. NeurIPS: Advances in neural information processing systems, vol. 26, pp. 3111-3119, 2013b. https://doi.org/10.48550/arXiv.1310.4546 [retrieved: 23.02.23]

[17] L. Li, O. Levi, P. Hosseini, and D. A. Broniatowski, "A multimodal method for satire detection using textual and visual cues," Proc. 3rd NLP4IF, pp. 33–38, Dec. 2020. https://arxiv.org/abs/2010.06671 [retrieved: 23.02.23]

[18] M. S. Razali, A. Abdul Halin, Y. W. Chow, N. Mohd Norowi, and S. Doraisamy, "Context-driven satire detection with deep learning," IEEE Access, vol. 10, pp. 78780-78787, 2022. https://ieeexplore.ieee.org/document/9841563 [retrieved: 23.02.23]

[19] https://www.linuxfoundation.org/ [retrieved: 23.02.23]

[20] https://ubuntu.com [retrieved: 23.03.23]

[21] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, L. Zettlemoyer, "Deep contextualized word representations," Proc. NAACL-HLT, vol. 1, pp. 2227–2237, Jun. 2018. https://aclanthology.org/N18-1202.pdf [retrieved: 23.02.23]

[22] H. M. Kumar, and B. S. Harish, "Automatic irony detection using feature fusion and ensemble classifier," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 5, no. 7, pp. 70–79, 2019. https://www.ijimai.org/journal/sites/default/files/files/2019/07/ijimai20195_7_7_pdf_17438.pdf [retrieved: 23.02.23]

[23] C. L. Lin, M. Ptaszynski, and F. Masui, "Exploring machine learning techniques for irony detection," Proc. 33rd JSAI, Jun. 2019. https://www.jstage.jst.go.jp/article/pjsai/JSAI2019/0/JSAI2019_2A4E203/_pdf/-char/en [retrieved: 23.02.23]

[24] S. Jiang, C. Chen, N. Lin, Z. Chen, and J. Chen, "Irony detection in the Portuguese language using BERT," Proc. IberLEF 2021, pp 891-897, Sept. 2021. http://ceur-ws.org/Vol-2943/idpt_paper1.pdf [retrieved: 23.02.23]

[25] D. Tomás, R. Ortega-Bueno, G. Zhang, P. Rosso, and R. Schifanella, "Transformer-based models for multimodal irony detection," Journal of Ambient Intelligence and Humanized Computing, 2022. doi.org/10.1007/s12652-022-04447-y [retrieved: 23.02.23]

[26] A. Joshi, P. Bhattacharyya, and M. J. Carman, "Automatic sarcasm detection: a survey," ACM Computing Surveys, vol. 50. no. 5, art. 73, pp.1–22, 2017. https://doi.org/10.1145/3124420 [retrieved: 23.02.23].

[27] https://pytorch.org [retrieved: 23.02.23]

[28] J. Pennington, R. Socher, and C. D. Manning, "GloVe: global vectors for word representation," Proc. EMNLP, pp. 1532-1543, Oct. 2014. http://dx.doi.org/10.3115/v1/D14-1162 [retrieved: 23.02.23]

[29] V. Simédoh, "Humour and irony in sub-Saharan Francophone literature: from critical issues to a poetics of laughter," Berlin: Peter Lang, 2012.