The Elephant in the Room: Ten Challenges of Computational Detection of Rhetorical Figures

Ramona Kühn
Innstraße 43
94032 Passau
University of Passau
ramona.kuehn@uni-passau.de

Jelena Mitrović
Innstraße 43
94032 Passau
University of Passau/
Institute for AI R&D of Serbia
jelena.mitrovic@uni-passau.de

Abstract

Computational detection of rhetorical figures focuses mostly on figures such as metaphor, irony, or sarcasm. However, there exist many more figures that are neither less important nor less prevalent. We want to pinpoint the reasons why researchers often avoid other figures and shed light on the challenges they struggle with when investigating those figures. In this comprehensive survey, we analyzed over 40 papers dealing with the computational detection of rhetorical figures other than metaphor, simile, analogy, sarcasm, and irony. We encountered recurrent challenges from which we compiled a ten point list. Furthermore, we suggest solutions for each challenge to encourage researchers to investigate a greater variety of rhetorical figures.

1 Introduction

Rhetorical figures such as metaphor, alliteration, or irony are present in our daily lives. They make language vivid, more emotional, or more persuasive. Each figure has a special function, e.g., figures with repetition create more emphasis (Fahnestock, 2002), while sarcasm and irony are often used in the context of hate speech (Frenda et al., 2023). To understand the often non-literal meaning and subtle nuances of a text containing rhetorical figures, it is important to reliably detect those figures computationally. Furthermore, the performance of classical NLP tasks improves when taking features of rhetorical figures into account. This was demonstrated for sentiment analysis (Nguyen et al., 2015), argumentation mining (Mitrović et al., 2017), text summarization (Alliheedi and Di Marco, 2014), and hate speech and abusive language detection (Lemmens et al., 2021).

Most detection approaches only consider the rhetorical figures metaphor (Shutova et al., 2013; Ghosh et al., 2015; Bizzoni et al., 2017; Bizzoni and Ghanimifard, 2018; Chakrabarty et al., 2022; Rai and Chakraverty, 2020; Tong et al., 2021; Ge et al., 2023), or irony and sarcasm (Ghosh et al., 2015; Wallace, 2015; Joshi et al., 2017; Yaghoobian et al., 2021). However, the Silva Rhetoricae[1] (Burton, 2007), an online resource for rhetorical figures and their descriptions, lists 435 different rhetorical figures. Most of those figures are neither less present nor less important than metaphor. For example, the figure antithesis is important in environmental (Green, 2021) or populist communication (Kühn et al., 2024), litotes is important in sentiment analysis (Karp et al., 2021), and polyptoton can highlight similarities while showing a distinction (Fahnestock, 2002).

We believe that it is essential to pay attention to the other figures, too. In this survey, we investigate the main challenges and problems researchers struggle with when computationally dealing with those figures. We examined over 40 papers describing computational detection approaches for rhetorical figures other than metaphor, simile, analogy, sarcasm, and irony. The figures range from A like alliteration to Z like zeugma. Table 3 in Appendix A illustrates the distribution of figures across the papers we examined, showing the frequency of appearance for each figure. The investigated papers were published between 2006 and 2024.

We focus on papers that consider the detection of rhetorical figures in written text, as speech or multimodal approaches further increase both the complexity and challenges. We were looking on Google Scholar[2] for relevant work by searching for figure names along with "detection", and including relevant related work. We explicitly did not look only into libraries such as the ACL anthology[3], as the field of rhetorical figure detection is not that represented at big conferences. From these results, we compiled a comprehensive list of ten key chal-

[1] http://rhetoric.byu.edu/
[2] https://scholar.google.com/
[3] https://aclanthology.org/

lenges and problems that show a recurrent pattern. We also provide suggestions for overcoming those challenges in order to further strengthen the field of computational detection of rhetorical figures in the future.

2 Rhetorical Figure Detection: Ten Challenges

We present ten challenges that most researchers face when trying to computationally identify rhetorical figures. We also suggest solutions for each of the challenges.

2.1 Inconsistent Definitions and Binary Classification

Although rhetorical figures have been studied for hundreds of years from a linguistic perspective, their spellings and definitions are often inconsistent (Harris et al., 2018; Gavidia et al., 2022; Kühn and Mitrović, 2024). This leads to different interpretations of what a rhetorical figure consists of. Consider, for example, the figure antithesis ("working all day, sleeping all night"). Most definitions agree on the antonymous relation (working vs. sleeping), but not every definition requires syntactic parallelism. Another example comes from the work of Dubremetz and Nivre (2017), in which the figure chiasmus is described, but the authors actually refer to a more specific form of chiasmus called antimetabole (Schneider et al., 2021). A further problem is that some figures are language dependent, i.e., a rhetorical figure in English does not have a matching counterpart in another language (Kühn et al., 2022; Zhu et al., 2022). For example, metaphor and simile are considered one figure in Chinese (Zhu et al., 2022), or figures with the same name have deviating definitions in different languages (Wang et al., 2022). We think these inconsistent definitions cause problems when figures are binary annotated, e.g., as *present* or *not present*, because figures deviate in their salience.

Suggested solution: Consulting different sources before approaching the figure detection task is a good way to start. More importantly, we think that the detection of rhetorical figures should not be considered a binary classification task. We suggest a ranking scheme (e.g., continuous values) tailored to every figure based on its salience and conspicuousness or how many properties from the textual definitions are fulfilled. Rankings for rhetorical figures have already proven to be useful (Dubremetz and Nivre, 2015; Troiano et al., 2018; Zhang and Wan, 2021). For example, in the case of antithesis, sentences that contain both parallelism and antonyms can be ranked higher than sentences with antonyms and no parallelism. Nevertheless, it is necessary to remember that annotations with continuous values are often more unreliable than binary annotations (Bagdon et al., 2024). To avoid this problem, we suggest a comparison-based annotation, e.g., best-worst scaling. This method already performed well in emotion intensity annotation with language models (Bagdon et al., 2024), which we consider related to rhetorical figure annotation.

2.2 Defining Boundaries and Intentional Usage

Another problem that most researchers encountered is the definition of the boundary in which to look for figures. As figures can span over multiple sentences, paragraphs, or the whole text, it is important to define where to start and where to end. If a repetition of two words is too far apart, it is not recognized as salient anymore by humans, while automatic parsers detect the repetition (Strommer, 2011). Properly defining boundaries determines the success of rhetorical figure detection (Strommer, 2011). An additional challenging aspect is to decide whether the figure is accidentally or intentionally present. Especially repetitions can occur without a rhetorical purpose (Strommer, 2011; Dubremetz and Nivre, 2015). This leads to the problem that annotators often cannot agree if it is actually a figure and which figure it is, decreasing the agreement between annotators and the reliability of the annotation itself. Strommer (2011) describes that in the case of his 156 instances, the annotators agreed only on two of them to be an intentional anaphora. Troiano et al. (2018) also mention that they had diverse annotations in their hyperbole dataset.

Suggested solution: It is important for future dataset construction to not only include one or two sentences containing the figure itself but also to consider larger text chunks. A ranking scheme mentioned in Section 2.1 can also help with expressing the salience of figures and deciphering between a rather accidental or intentional use.

2.3 Lack of Data/Datasets

When considering popular figures such as metaphor, irony, or sarcasm, researchers can profit from users that tag their posts in social media, e.g., `#sarcasm` (Ranganath et al., 2018). This makes it easier to compile larger annotated datasets. Other figures that also play an important role in persuasive communication, but are not that present in the minds of the average social media users are often neglected. It can be more difficult to find instances of those figures (Dubremetz and Nivre, 2015). Another problem when creating datasets could be an inherent bias, as only sentences with salient rhetorical figures are chosen. This means that edge cases, where it is arguable whether it is a rhetorical figure or not (see Section 2.2,) are not included in the dataset.

Suggested solution: Generative large language models (LLMs) can help create sentences containing rhetorical figures. A downside is, however, that the LLM was probably pre-trained on data in which rhetorical figures other than metaphor are not explicitly annotated, making the generation more difficult. Furthermore, one must be aware of the vicious cycle that LLMs can only generate sentences with rhetorical figures they already know. If the LLM does not know the construction rules of a rhetorical figure, it cannot reliably generate sentences containing the figure. It is still necessary that human annotators oversee the process, as in Chakrabarty et al. (2022) where three annotators verified the texts generated by GPT-3. Another solution to collect more annotated data is to develop platforms where users can submit instances of rhetorical figures in a game-like scenario (Kühn and Mitrović, 2023).

2.4 Imbalanced Datasets and Deceptive Performance Metrics

If datasets for rhetorical figures are constructed, researchers like Bhattasali et al. (2015); Dubremetz and Nivre (2017); Ranganath et al. (2018); Adewumi et al. (2021); Kühn et al. (2023) face highly imbalanced datasets, i.e., the majority of data points are not a rhetorical figure. Using then accuracy as a performance metric can be highly deceptive. In a dataset where 99 % of instances are not a rhetorical figure, a model that consistently predicts a particular class will achieve a classification accuracy of 99 %. Also, other metrics such as precision and recall have to be considered carefully as their problems became obvious in the work of Gawryjolek (2009) and Java (2015). Furthermore, with only a few datasets with positive examples of rhetorical figures, it is more difficult to train machine models on (Dubremetz and Nivre, 2017; Zhang and Wan, 2021) or fine-tune language models to achieve better performance.

Suggested solution: Augmentation techniques or over- or undersampling can help decrease the imbalance. LLMs can also help create more sentences containing rhetorical figures. Evaluation metrics have to be chosen wisely.

2.5 Not Including Ontologies

Formal domain ontologies of rhetorical figures have the goal of overcoming the problem of inconsistent definitions and spellings (see Section 2.1). There exist ontologies such as the English RhetFig ontology (Harris et al., 2017), the Ploke (Wang et al., 2021), the Serbian Retfig (Mladenović and Mitrović, 2013), the German GRhOOT (Kühn et al., 2022), and a multilingual ontology (Wang et al., 2021). They all represent rhetorical figures in the form of classes and relations, describing how they are constructed, where they appear, and which cognitive effects they have. However, we realized that none of the investigated approaches use those ontologies.

Suggested solution: We suggest including those ontologies in the process of detecting rhetorical figures. We are confident that those ontologies can help improve detection rules or help annotators achieve higher agreement. Further applications are also possible when the ontologies are combined with LLMs, especially in a retrieval augmented generation (RAG) system (Lewis et al., 2020), where the context of an LLM is enhanced with rhetorical knowledge from the ontologies. In addition, it is possible that the data generation and annotation capabilities of LLMs are improved, too.

2.6 Missing Context

Rhetorical figures are often implicit, subtle, and can only be understood with context knowledge (Lawrence et al., 2017; Ranganath et al., 2018; Troiano et al., 2018). Some figures can even be used both in a figurative and literal meaning, e.g., rhetorical questions, which are syntactically not different from regular questions (Ranganath et al., 2018), or hyperboles that can also have both a literal and a figurative meaning, depending on context: Troiano et al. (2018) give the example of *"It took ages to build the castle"* vs. *"It took ages to build the castle. After a few minutes, my little* *brother had already destroyed it!"* For an efficient detection of rhetorical figures, it is important to understand the semantics, syntax, and pragmatics (Medková, 2020).

Suggested solution: For the detection of most figures, it is necessary to include sentences/paragraphs pre- and succeeding the sentence of interest for context knowledge. In addition, LLMs can help to resolve contextual ambiguities and syntactic knowledge about figure formation can be extracted from ontologies.

2.7 Focus on Rule-based Methods

While deep-learning methods are already implemented successfully for the detection of metaphors (Bizzoni et al., 2017; Bizzoni and Ghanimifard, 2018), we observe a focus on rule-based approaches for lesser-known figures. We are certain that approaches based on LLMs will massively increase in the future and may overcome the performance of current state-of-the-art rule-based approaches. Zhu et al. (2022) experience lower performance with rule-based approaches for various rhetorical figures. They note that a complex task such as the detection of rhetorical figures cannot be solved by identifying "shallow and obvious patterns." Similar to the field of mail spam detection, there is no use in creating lists with known rhetorical figures, as humans are creative and come up with new metaphors or analogies. From the over 40 papers we investigated, the authors implemented 87 different detection techniques for various figures (see Table 1). 68.97 % are rule-based approaches, whereas only 27.59 % are model-based or deep learning approaches. Only one approach from Kühn et al. (2024) combines a rule-based with a model-based approach to detect the figure antithesis.

Suggested solution: We suggest using LLMs. However, as even powerful language models show a decreased performance in the understanding of rhetorical figures compared to humans (Liu et al., 2022), we believe that the combination of LLMs and rule-based approaches can be fruitful. For example, the presence of figures with perfect lexical repetition can be better verified by rules.

2.8 Focus on English

Existing datasets of rhetorical figures mainly contain sentences in English. This makes it even more challenging to investigate rhetorical figures in other languages. A direct translation from English into another language is often not possible without losing the original form of the rhetorical figure, especially if it contains syntactical aspects (Kühn et al., 2023). Another problem is that English is uncased and has neither a grammatical gender nor inflection. Some figures based on a change in inflection (such as polyptoton) appear less frequently than in languages with strong inflection, e.g., German (Fahnestock, 2002). Furthermore, English does not have separable verbs. These are verbs where the prefix is split from the main verb. This can create repetitions without a rhetorical purpose: "Wir fingen an, an danach zu denken" (*"We began to think about what comes after."*), where "an" is repeated while referring to different concepts. This highlights once again why rule-based approaches can fail (see Section 2.7). Table 2 shows that 66.81 % of the investigated approaches focus on rhetorical figures in English. When authors consider figures in multiple languages (e.g., Hromada (2011) investigates English, Latin, French and German, or Lagutina et al. (2019) in Russian and English), we counted them individually for every language.

Approach category	#Approaches	In Percent
Rule-based	**60**	**68.97 %**
Model-based	24	27.59 %
Rule-& Model-based	1	1.15 %
Unknown	2	2.30 %

Table 1: Distribution of the approach categories over the 86 approaches.

Language	#Approaches	In Percent
English	**74**	**69.81 %**
German	10	9.43 %
Russian	8	7.55 %
French	4	3.77 %
Latin	4	3.77 %
Chinese	3	2.83 %
Czech	2	1.89 %
Japanese	1	0.94 %

Table 2: Distribution of languages.

The focus on English leads to another problem. Most NLP tools are developed for English. According to the #BenderRule (Bender, 2019), it is "undesirable" that language technologies are only developed for one or two popular languages. This leads to a vicious cycle: The more tools are tailored to the English language, the more researchers only focus on the detection of rhetorical figures in English. Because appropriate tools are lacking for other languages, identifying rhetorical figures is more challenging and might be neglected. As we mentioned previously, translating the data into English to be able to use existing tools is not an option.

The focus on English already created inequalities regarding model creation, leading to a lower acceptance rate at NLP conferences for papers not dealing with English (Søgaard, 2022).

Suggested solution: This is not an easy challenge to overcome as it affects the entire discipline of NLP. Nevertheless, we would like to encourage researchers to perform their work in languages other than English. Also, we think that it is necessary to reward research that focuses on other languages. Another solution can be the creation of adequate tools in multiple languages.

2.9 Neglecting Cognitive Effects

Another point of critique is that research about rhetorical figures focuses on detection but often forgets about the cognitive effects of the figures (Mitrović et al., 2020). This seems to be especially the case when approaching rhetorical figures from a computational perspective, as it is already challenging to implement detection algorithms. Often, the interpretation of the figure in the given context is then neglected. However, as every form of a figure has a certain function (Givón, 1995), it is important to not only identify figures but also interpret their usage.

Suggested solution: It is important to have a holistic look at the task of rhetorical figure detection. We suggest including explanations of what the usage of a certain figure in a given context actually means and analyzing which emotions are created for readers and listeners.

2.10 Lack of Interdisciplinary Efforts

Dealing with rhetorical figures is a highly interdisciplinary task that includes all obstacles from other disciplines. From an NLP perspective, rhetorical figures are not only syntactic constructions. They also include semantic features, have a transferred meaning, or depend on sound. For certain figures, it is necessary to identify negation, which is still a hard task in NLP. As rhetorical figures appear in all areas of our daily lives, we encounter them in the domain of advertising, politics, sentiment analysis, hate speech, machine translation, and many more. Rhetorical figures are also interesting for neuroscience in terms of their effect on the human brain. Green (2021) showed how rhetorical figures are applied in environmental arguments. Fahnestock (2002) highlights the importance of rhetorical figures in disciplines such as biology or chemistry, among others. If those fields understand rhetorical figures better, they can communicate more effectively with convincing arguments. In the field of law, there is a growing body of work devoted to argumentation and deciphering the effects of figures on persuasiveness (Al Zubaer et al., 2023).

Suggested solution: Researchers coming from different disciplines should join forces to build a holistic view of rhetorical figures, their purpose, function, and effect. Computer scientists and linguists can benefit from one another especially. Other disciplines can also profit from collaboration and open up new areas of research.

3 Conclusion

Our comprehensive review of over 40 papers highlights the prevalent challenges in computationally detecting rhetorical figures. As each rhetorical figure plays a crucial role in our daily communication, we urge researchers to tackle the presented challenges. When we can understand the non-literal and subtle meaning of rhetorical figures, we can improve existing systems and better understand language. In the future, we would like to see some of the suggestions implemented. Furthermore, we aim to inspire researchers to also focus on the detection of lesser-known figures.

Acknowledgements

The project on which this report is based was funded by the German Federal Ministry of Education and Research (BMBF) under the funding code 01IS20049. The authors are responsible for the content of this publication. Furthermore, we thank the anonymous reviewers for their valuable feedback, insightful comments, and new perspectives that enhanced the quality of our work.

SPONSORED BY THE

Federal Ministry
of Education
and Research

References

Tosin P Adewumi, Roshanak Vadoodi, Aparajita Tripathy, Konstantina Nikolaidou, Foteini Liwicki, and Marcus Liwicki. 2021. Potential idiomatic expression (pie)-english: Corpus for classes of idioms. *arXiv preprint arXiv:2105.03280*.

Abdullah Al Zubaer, Michael Granitzer, and Jelena Mitrović. 2023. Performance analysis of large language models in the domain of legal argument mining. *Frontiers in Artificial Intelligence*, 6.

Mohammed Alliheedi and Chrysanne Di Marco. 2014. Rhetorical figuration as a metric in text summarization. In *Advances in Artificial Intelligence: 27th Canadian Conference on Artificial Intelligence, Canadian AI 2014, Montréal, QC, Canada, May 6-9, 2014. Proceedings 27*, pages 13–22. Springer.

Christopher Bagdon, Prathamesh Karmalker, Harsha Gurulingappa, and Roman Klinger. 2024. " you are an expert annotator": Automatic best-worst-scaling annotations for emotion intensity modeling. *arXiv preprint arXiv:2403.17612*.

Emily Bender. 2019. The #benderrule: On naming the languages we study and why it matters. *The Gradient*, 14.

Shohini Bhattasali, Jeremy Cytryn, Elana Feldman, and Joonsuk Park. 2015. Automatic identification of rhetorical questions. 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 743–749.

Yuri Bizzoni, Stergios Chatzikyriakidis, and Mehdi Ghanimifard. 2017. “deep” learning : Detecting metaphoricity in adjective-noun pairs. In *Proceedings of the Workshop on Stylistic Variation*, pages 43–52, Copenhagen, Denmark. Association for Computational Linguistics.

Yuri Bizzoni and Mehdi Ghanimifard. 2018. Bigrams and bilstms two neural networks for sequential metaphor detection. In *Proceedings of the Workshop on Figurative Language Processing*, pages 91–101.

Gideon O Burton. 2007. The forest of rhetoric. *Silva Rhetoricae*, 6.

Tuhin Chakrabarty, Arkadiy Saakyan, Debanjan Ghosh, and Smaranda Muresan. 2022. Flute: Figurative language understanding through textual explanations. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 7139–7159.

Marie Dubremetz and Joakim Nivre. 2015. Rhetorical figure detection: The case of chiasmus. In *Proceedings of the Fourth Workshop on Computational Linguistics for Literature*, pages 23–31.

Marie Dubremetz and Joakim Nivre. 2017. Machine learning for rhetorical figure detection: More chiasmus with less annotation. In *Proceedings of the 21st Nordic Conference on Computational Linguistics*, pages 37–45.

Jeanne Fahnestock. 2002. *Rhetorical figures in science*. Oxford University Press on Demand.

Simona Frenda, Viviana Patti, and Paolo Rosso. 2023. When sarcasm hurts: Irony-aware models for abusive language detection. In *International Conference of the Cross-Language Evaluation Forum for European Languages*, pages 34–47. Springer.

Martha Gavidia, Patrick Lee, Anna Feldman, and Jing Peng. 2022. Cats are fuzzy pets: A corpus and analysis of potentially euphemistic terms. *arXiv preprint arXiv:2205.02728*.

Jakub Jan Gawryjolek. 2009. Automated annotation and visualization of rhetorical figures. Master's thesis, University of Waterloo.

Mengshi Ge, Rui Mao, and Erik Cambria. 2023. A survey on computational metaphor processing techniques: From identification, interpretation, generation to application. *Artificial Intelligence Review*, pages 1–67.

Aniruddha Ghosh, Guofu Li, Tony Veale, Paolo Rosso, Ekaterina Shutova, John Barnden, and Antonio Reyes. 2015. Semeval-2015 task 11: Sentiment analysis of figurative language in twitter. In *Proceedings of the 9th international workshop on semantic evaluation (SemEval 2015)*, pages 470–478.

Talmy Givón. 1995. Isomorphism in the grammatical code. *Iconicity in language*, pages 47–76.

Nancy L Green. 2021. Some argumentative uses of the rhetorical figure of antithesis in environmental science policy articles. In *Proceedings of the Workshop on Computational Models of Natural Argument*, pages 85–90.

Randy Allen Harris, Chrysanne Di Marco, Ashley Rose Mehlenbacher, Robert Clapperton, Insun Choi, Isabel Li, Sebastian Ruan, and Cliff O'Reilly. 2017. A cognitive ontology of rhetorical figures. *Cognition and Ontologies*, pages 18–21.

Randy Allen Harris, Chrysanne Di Marco, Sebastian Ruan, and Cliff O'Reilly. 2018. An annotation scheme for rhetorical figures. *Argument & Computation*, 9(2):155–175.

Daniel Hromada. 2011. Initial experiments with multilingual extraction of rhetoric figures by means of perl-compatible regular expressions. In *Proceedings of the Second Student Research Workshop associated with RANLP 2011*, pages 85–90.

James Java. 2015. *Characterization of prose by rhetorical structure for machine learning classification*. Ph.D. thesis, Nova Southeastern University.

Aditya Joshi, Pushpak Bhattacharyya, and Mark J Carman. 2017. Automatic sarcasm detection: A survey. *ACM Computing Surveys (CSUR)*, 50(5):1–22.

M Karp, N Kunanets, and Y Kucher. 2021. Meiosis and litotes in the catcher in the rye by jerome david salinger: text mining. In *CEUR Workshop Proceedings*, volume 2870, pages 166–178.

Ramona Kühn, Jelena Mitrović, and Michael Granitzer. 2022. Grhoot: Ontology of rhetorical figures in german. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 4001–4010.

Ramona Kühn and Jelena Mitrović. 2023. Multilingual domain ontologies of rhetorical figures and their applications. In *UniDive 1st General Meeting*.

Ramona Kühn and Jelena Mitrović. 2024. Status Quo der Entwicklungen von Ontologien Rhetorischer Figuren in Englisch, Deutsch und Serbisch. In *Book of Abstracts - DHd2024*. Zenodo.

Ramona Kühn, Jelena Mitrović, and Michael Granitzer. 2023. Hidden in plain sight: Can german wiktionary and wordnets facilitate the detection of antithesis? In *Proceedings of the 12th Global Wordnet Conference*, pages 106–116.

Ramona Kühn, Khouloud Saadi, Jelena Mitrović, and Michael Granitzer. 2024. Using pre-trained language models in an end-to-end pipeline for antithesis detection. In *Proceedings of the 14th Language Resources and Evaluation Conference*. European Language Resources Association.

Nadezhda Stanislavovna Lagutina, Kseniya Vladimirovna Lagutina, Elena Igorevna Boychuk, Inna Alekseevna Vorontsova, and Il'ya Vyacheslavovich Paramonov. 2019. Automated search of rhythm figures in a literary text for comparative analysis of originals and translations based on the material of the english and russian languages. *Modelirovanie i Analiz Informatsionnykh Sistem*, 26(3):420–440.

John Lawrence, Jacky Visser, and Chris Reed. 2017. Harnessing rhetorical figures for argument mining. *Argument & Computation*, 8(3):289–310.

Jens Lemmens, Ilia Markov, and Walter Daelemans. 2021. Improving hate speech type and target detection with hateful metaphor features. In *Proceedings of the Fourth Workshop on NLP for Internet Freedom: Censorship, Disinformation, and Propaganda*, pages 7–16.

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al. 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems*, 33:9459–9474.

Emmy Liu, Chen Cui, Kenneth Zheng, and Graham Neubig. 2022. Testing the ability of language models to interpret figurative language. *arXiv preprint arXiv:2204.12632*.

Helena Medková. 2020. Automatic detection of zeugma. In *RASLAN*, pages 79–86.

Jelena Mitrović, Cliff O'Reilly, Miljana Mladenović, and Siegfried Handschuh. 2017. Ontological representations of rhetorical figures for argument mining. *Argument & Computation*, 8(3):267–287.

Jelena Mitrović, Cliff O'Reilly, Randy Allen Harris, and Michael Granitzer. 2020. Cognitive modeling in computational rhetoric: Litotes, containment and the unexcluded middle. In *ICAART (2)*, pages 806–813.

Miljana Mladenović and Jelena Mitrović. 2013. Ontology of rhetorical figures for serbian. In *Text, Speech, and Dialogue*, pages 386–393, Berlin, Heidelberg. Springer.

Hoang Long Nguyen, Trung Duc Nguyen, Dosam Hwang, and Jason J Jung. 2015. Kelabteam: A statistical approach on figurative language sentiment analysis in twitter. In *Proceedings of the 9th international workshop on semantic evaluation (SemEval 2015)*, pages 679–683.

Sunny Rai and Shampa Chakraverty. 2020. A survey on computational metaphor processing. *ACM Computing Surveys (CSUR)*, 53(2):1–37.

Suhas Ranganath, Xia Hu, Jiliang Tang, Suhang Wang, and Huan Liu. 2018. Understanding and identifying rhetorical questions in social media. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 9(2):1–22.

Felix Schneider, Björn Barz, Phillip Brandes, Sophie Marshall, and Joachim Denzler. 2021. Data-driven detection of general chiasmi using lexical and semantic features. In *Proceedings of the 5th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, pages 96–100.

Ekaterina Shutova, Simone Teufel, and Anna Korhonen. 2013. Statistical metaphor processing. *Computational Linguistics*, 39(2):301–353.

Anders Søgaard. 2022. Should we ban english nlp for a year? In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 5254–5260.

Claus Walter Strommer. 2011. Using rhetorical figures and shallow attributes as a metric of intent in text.

Xiaoyu Tong, Ekaterina Shutova, and Martha Lewis. 2021. Recent advances in neural metaphor processing: A linguistic, cognitive and social perspective. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4673–4686.

Enrica Troiano, Carlo Strapparava, Gözde Özbal, and Serra Sinem Tekiroğlu. 2018. A computational exploration of exaggeration. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3296–3304.

Byron C Wallace. 2015. Computational irony: A survey and new perspectives. *Artificial intelligence review*, 43:467–483.

Yetian Wang, Randy Allen Harris, and Daniel M Berry. 2021. An ontology for ploke: Rhetorical figures of lexical repetitions. In *JOWO*.

Yetian Wang, Ramona Kühn, Randy Allen Harris, Jelena Mitrović, and Michael Granitzer. 2022. Towards a unified multilingual ontology for rhetorical figures. In *Proceedings of the 14th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management. Valletta, Malta: SCITEPRESS-Science and Technology Publications*, pages 117–127.

Hamed Yaghoobian, Hamid R Arabnia, and Khaled Rasheed. 2021. Sarcasm detection: A comparative study. *arXiv preprint arXiv:2107.02276*.

Yunxiang Zhang and Xiaojun Wan. 2021. Mover: Mask, over-generate and rank for hyperbole generation. *arXiv preprint arXiv:2109.07726*.

Dawei Zhu, Qiusi Zhan, Zhejian Zhou, Yifan Song, Jiebin Zhang, and Sujian Li. 2022. Configure: Exploring discourse-level chinese figures of speech. *arXiv preprint arXiv:2209.07678*.

A Appendix

Following Table 3 shows which figures were considered in the papers and how often they are investigated. If multiple figures are investigated in one paper, we counted them multiple times.

Figure	# Approaches
Alliteration	2
Anadiplosis	5
Anaphora/Epanaphora	7
Antimetabole	**10**
Antithesis	5
Assonance	1
Chiasmus	8
Conduplicatio	1
Diacope	1
Dirimens copulatio	1
Duality	1
Dysphemism	1
Epanalepsis	3
Epanaphora	1
Epiphora/Epistrophe	6
Epizeuxis	4
Euphemism	3
Eutrepismus	1
Hyperbole	4
Isocolon	2
Litotes	3
Meiosis	1
Metonymy	6
Oxymoron	3
Parallelism	3
Personification	1
Ploke/Ploce	2
Polyptoton	4
Polysyndeton	3
Quote	1
Repetition	1
Rhetorical question	3
Symploke	2
Synaesthesia	1
Zeugma	1

Table 3: Frequency of appearance for each figure in the investigated papers.