The Rhetoricon Database: An overview and an appreciation

Randy Allen Harris and Chrysanne Di Marco

University of Waterloo, Waterloo ON, N2L 3G1 Canada

Abstract

We have developed a research track to realize several important lines of thought growing out of the Computational Rhetoric movement that the CMNA was so important for initiating and fostering. These lines include figure detection and annotation, ontological modelling, collocation and colligation of rhetorical figures, gamification, and the incorporation of Construction Grammar. We provide an overview, highlighting the ways the computational modelling of natural argumentation community has been instrumental to its development.

Keywords

Rhetorical figures, neurocognition, argument mining, figure detection, database

1.Introduction

Inspired by the arguments of Floriana Grasso, Jim Crosswhite, Chris Reed, and others in Computational Rhetoric and Argumentation community [1–4], —especially in conjunction with Jeanne Fahnestock's programme of figural logic [5–9] and the linguistic theory of Construction Grammar [10–13], which are more closely aligned than most practitioners of either are aware [14, 15]—we (RAH and CDM) have been developing a multipronged approach to rhetoric, argumentation, computational modelling, and cognitive neuroscience at the University of Waterloo now for what is closing in on two decades. CMNA has been among the most encouraging, receptive, and critical associations for this development. It's effectively where we got our start with these lines of research. So we thought it appropriate to report on our progress and the continuing promise of our approach at the 25th annual edition of the Workshop.

Our anchor of our project, The Rhetoricon Database, is founded on the view that many rhetorical figures are fundamentally form/function alignments whose form activates neurocognitive pattern biases and whose functions serve extra-grammatical communicative goals, including the epitomizing of lines of argument [5], so that figures are therefore essential for Advanced Natural Language Processing (ANLP) and full-scope Argument Mining (AM). The Rhetoricon Project has argument mining, figure detection, ontological, linguistic, cognitive neuroscience, database, Machine Learning (ML), Artificial Intelligence (AI), and game components.

2.The Rhetoricon Project

2.1.Computational rhetoric

We see our research as a direct outgrowth of the Computational Rhetoric / Argument Mining programme that coalesced first at the June 2000 Symposium on Argument and Computation at Bonskeid House, in Perthshire, Scotland, organized by Reed and Timothy J. Norman (Grasso participated in the event and describes the experience as "almost 'mystical'" [16, p.iv]), the Computational Models of Natural Argumentation (CMNA) workshops that began the following year, spearheaded by Reed and Grasso, and the journal, *Argument & Computation*, introduced by Grasso and Reed along with Iyad Rahwan and Guillermo R. Simari [3] before the end of that decade. You will find the workshops and the journal prominently in the references to our work, including the special issue on rhetorical figures of *Argument & Computation* we co-edited in 2017 [17]. We have offered two workshops of our own at Waterloo as well—under the handle, Computing figures, figuring computers—the second of which generated the articles for that special issue.

2.2.Figure detection

Our project effectively began when a MMath student, Jakub Gawryjolek wanted to build a tool to detect and annotate metaphors. CDM sent him to RAH, who told him that metaphors were a computational rats nest, but that there were many figures much easier to detect and annotate, especially figures of lexical repetition, like epanaphora (repetition at the beginning of phrases or clauses) and antimetabole (reverse repetition). Jakub built a tool for exactly those sorts of figures [18], which we demonstrated at CMNA IX [19], to a very warm reception. Claus Strommer, using ML and focusing exclusively on epanaphora, followed suit under our direction [20], with Reed as the External Examiner. This research sparked interest among a community of computational linguists and figure detection has grown into an active subfield of Computational Rhetoric [21–32].

2.3.Figure ontology

"We will assume that the argumentation process can take place on a certain set of objects of discourse," Grasso observed at the very beginnings of Computational Rhetoric, "forming an ontology" [4, p.204]; see also [16]. Her ontology was a large grained one, at the level of rhetors, ethoi, and stance, but we realized two things. First, that a finer grained ontology, at the level of rhetorical figures, would supplement this project in necessary ways. Second, that such an ontology would need to map the relations among rhetorical figures as form/function alliances with linguistic domains, where the function was most situated, and the neurocognitive pattern biases, which activated salience, mnemonic, and aesthetic effects most connected to the form [33–42]. For instance, the figure of epanaphora activates heightened attentional responses to repetition and position for lexemes relative to syntax (1), while the figure of rhyme also activates those same responses as well, but now for syllables relative to lexemes (2).

1. Easy come, easy go
2. By hook or by crook.

Our ontology research led quite directly to insights about figural collocation, which is extraordinarily common in all languages, varieties, registers, and genres, but which is also extraordinarily neglected by researchers. For instance, Example (1) includes antithesis as well as epanaphora (in the antonyms, *come* and *go*); Example (2) includes epanaphora as well as rhyme. Both Examples include parison (syntactic structure repetition) and isocolon (prosodic intonation repetition) as well. Moreover, both the form and the function of figured instances is a product of collocation, rather than of individual figures as conventionally assumed [5, 9]. These collocational effects are frequently mediated by iconicity [5, 7, 43].

Our ontological approach to rhetorical figures also influenced computational linguists, most notably Cliff O'Reilly [44] and Jelena Mitrović [45], with whom we have collaborated, along with Mitrović's colleagues and students at Universität Passau [33, 36, 41, 46]; Yetian Wang, our student at Waterloo, and Ramona Kühn, Mitrović's at Passau, have done especially impressive work [42, 47, 48]. Ramona successfully defended her thesis earlier this year; Yetian will be defending this week, just after we submit this paper, with Grasso as External Examiner.

2.4.Figure annotation

The ontology required proof and testing with instances, which begat a database, which has been anchoring our project since 2015, growing to thousands of instances annotated for hundreds of figures, very frequently in collocations. We developed an annotation scheme using Extensible Markup Language (XML), which we sketched out and presented at CMNA XVI [49], which generated valuable feedback. That first iteration used XML attributes but feedback encouraged us to keep exploring options and we subsequently refined it into a standoff markup system utilizing JavaScript Object Notation (JSON) [55].

2.5.Gamification

Our work has generated a number of hypotheses about form/function correlations, many of them adapted from Fahnestock's work, especially in connection with figural collocation and grammatical colligation, which we plan to probe in corpora. We also plan to build our own Pretrained Language Model tuned to rhetorical figuration (and therefore, neurocognitive affinities). To both of those ends, we will need to populate our database both to greater volume and to greater breadth (genres, registers, modes) and have developed a game to crowdsource its population. [50–52], GoFigure; see also Robert Clapperton's RAH-directed thesis [53].

3.Figures and constructions

Our most important insights, we feel, are in the areas of figural collocation and grammatical colligation. "Typically," Jeanne Fahnestock points out, "the antimetabole epitomizes arguments concerning reciprocal causality, a causal influence that goes in opposite directions, or a reversible process" [5, p.141]. This is certainly true and Fahnestock's work continues to inspire us, but it is incomplete. Other figures necessarily collocate with antimetabole to convey reciprocality. Take the famous "all for one and one for all." It conveys reciprocal obligation. The *all* is obliged to defend and uphold the interests of the *one*, and the *one* is obliged to defend and uphold the interests of the *all*.

But now compare "all for one, one with all;" still an antimetabole (reverse lexical repetition of *all* and *one*), but without the medial repetition of *for* (the figure is mesodiplosis): no reciprocality. Now compare "all for one, one for defending and upholding the interests of all;" still antimetabole, still mesodiplosis, but the syntax is no longer parallel (i.e., the figure of parison is no longer manifest; and, admittedly the new syntactic structure introduces lexical complications). What is necessary for the reciprocality is in fact a conspiracy of these three figures: antimetabole, mesodiplosis, and parison.[1] If you repeat two NPs on the opposite side of the same two place predicate and you maintain the syntactic structure (in English), you also reverse the syntactic and semantic roles of the NPs. In this example, first *one* is the head noun and *all* is the prepositional object, then *all* is the head noun and *one* is the prepositional object. First *one* is the TRAJECTOR and *all* is the LANDMARK, then *all* is the LANDMARK and *one* is the TRAJECTOR.

Now compare "all and one, one and all." We have antimetabole, mesodiplosis, and parison but again no reciprocality. Grammatical colligation is just as important as figural collocation. They work together. In this case, we no longer have a two-place predicate occupying the mesodiplosis. We have a conjunction. The communicative function now is irrelevance of order. These insights have led us to increasingly integrate our programme with Construction Grammar [15, 43, 54–56].

4.Conclusion and appreciation

Our programme, we believe, is one of the most promising to develop within the Computational Rhetoric paradigm that the CMNA workshops helped to initiate and certainly have fostered. It follows from the ontological arguments of Grasso, the figural logic of Fahnestock, the obvious importance of cognitive neuroscience, and the natural relevance of AI. Along the way we have discovered, or at least dramatically expanded, the role of collocation and colligation in figural logic, and the deep compatibility of Construction Grammar. We won't say that none of this would have been possible without the CMNA, but it would have taken a much different character and felt more like we were swimming against the current rather than joining a vibrant, growing, and encouraging community.

Acknowledgements

This research has been funded by several agencies over the years, most notably by the Social Sciences and Humanities Research Council of Canada and the Natural Science and Engineering Research Council of Canada.

References

1. Reed, C., Norman, T.J. eds: Argumentation machines: new frontiers in argument and computation. Kluwer Acad. Publ, Dordrecht (2004).

---

[1] There are other figures present in "all for one, one for all" as well (epanalepsis, anadiplosis), but they are less criterial for reciprocality.

2. Crosswhite, J.: Rhetoric and computation. Presented at the Symposium on Argument and Computation , Bonskeid House, Perthshire, Scotland. June 27 (2000).
3. Grasso, F., Rahwan, I., Reed, C., Simari, G.R.: Introducing Argument & Computation. Argument & Computation. 1, 1–5 (2010). https://doi.org/10.1080/19462161003734804.
4. Grasso, F.: Towards Computational Rhetoric. IL. 22, (2002). https://doi.org/10.22329/il.v22i3.2589.
5. Fahnestock, J.: Rhetorical figures in science. Oxford University Press, New York (1999).
6. Fahnestock, J.: Figures of Argument. Informal Logic. 24, 115–135 (2004). https://doi.org/10.22329/il.v24i2.2139.
7. Fahnestock, J.: Rhetorical Style: The Uses of Language in Persuasion. Oxford University Press, New York (2011).
8. Harris, R.A.: Figural Logic in Gregor Mendel's "Experiments on Plant Hybrids." Philosophy & rhetoric. 46, 570–602 (2013).
9. Tindale, C.W. (Christopher W.: Rhetorical argumentation : principles of theory and practice. In: Rhetorical argumentation : principles of theory and practice. Sage Publications, Thousand Oaks, Calif (2004).
10. Fillmore, C.J.: The mechanisms of 'Construction Grammar'. Berkeley Linguistic Society. 14, 35–55 (1988).
11. Fillmore, C.J., Kay, P., O'Connor, M.C.: Regularity and Idiomaticity in Grammatical Constructions: The Case of Let Alone. Language. 64, 501 (1988). https://doi.org/10.2307/414531.
12. Goldberg, A.: Constructions: A new theoretical approach to language. Trends in Cognitive Science. 7, 219–224 (2003).
13. Goldberg, A.E.: Explain Me This: Creativity, Competition, and the Partial Productivity of Constructions. Princeton University Press, Princeton (2019).
14. Turner, M.: Figure. In: Cacciari, C., Gibbs, R., Jr., Katz, A., and Turner, M. (eds.) Figurative Language and Thought. pp. 44–87. Oxford University Press, New York (1998).
15. Harris, R.A.: Grammatical constructions and rhetorical figures: The case of chiasmus. LACUS Forum. 38, (2022).
16. Grasso, F.: Characterizing rhetorical argumentation, https://cgi.csc.liv.ac.uk/~floriana/TesiFlo.pdf, (2003).
17. Harris, R.A., Di Marco, C. eds: [Special issue] Rhetorical figures. Argument & computation. 8, 211–231 (2017).
18. Gawryjolek, J.: Automated annotation and visualization of rhetorical figures, https://uwspace.uwaterloo.ca/handle/10012/4426, (2009).
19. Gawryjolek, J., Di Marco, C., Harris, R.A.: Automated annotation and visualization of rhetorical figures. In: 9th International Workshop on Computational Models of Natural Argument, International Joint Conference on Artificial Intelligence. , Pasadena, CA (2009).
20. Strommer, C.: Using rhetorical figures and shallow attributes as a metric of intent in text, (2011).
21. Dubremetz, M.: Detecting Rhetorical Figures based on repetition of words: Chiasmus, epanaphora, epiphora, https://www.kalendarium.uu.se/Evenemang?eventId=31460, (2017).
22. Dubremetz, M.: Vers une identification automatique du chiasme de mots. In: Actes de la 15e Rencontres des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues. pp. 150–163. , Les Sables d'Olonne, France (2013).
23. Dubremetz, M., Nivre, J.: Rhetorical Figure Detection: Chiasmus, Epanaphora, Epiphora. Frontiers in Digital Humanities. 5, 10 (2018). https://doi.org/10.3389/fdigh.2018.00010.

24. Dubremetz, M., Nivre, J.: Rhetorical figure detection: The case of chiasmus. In: NAACL-HLT Fourth Workshop on Computational Linguistics for Literature. pp. 23–31. Curran Associations, New York (2015).
25. Dubremetz, M., Nivre, J.: Syntax matters for rhetorical structure: The case of chiasmus. In: Proceedings of the Fifth Workshop on Computational Linguistics for Literature, NAACL-HLT. pp. 47–53. , San Diego, CA (2016).
26. Dubremetz, M., Nivre, J.: Machine learning for rhetorical figure detection: More chiasmus with less annotation. In: Proceedings of the 21st Nordic Conference of Computational Linguistics (2017).
27. Java, J.: Characterization of prose by rhetorical structure for machine learning classification5, (2015).
28. Kühn, R., Mitrovic, J., Granitzer, M.: Hidden in plain sight: Can german wiktionary and wordnets facilitate the detection of antithesis? In: 12th international global wordnet conference (2023).
29. Kühn, R., Mitrović, J.: The elephant in the room: Ten challenges of computational detection of rhetorical figures. In: Ghosh, D., Muresan, S., Feldman, A., Chakrabarty, T., and Liu, E. (eds.) Proceedings of the 4th workshop on figurative language processing (FigLang 2024). pp. 45–52. Association for Computational Linguistics, Mexico City, Mexico (Hybrid) (2024). https://doi.org/10.18653/v1/2024.figlang-1.6.
30. Green, N.L.: Some Argumentative Uses of the Rhetorical Figure of Antithesis in Environmental Science Policy Articles. In: CMNA'21: Workshop on Computational Models of Natural Argument. CEUR-WS.org (2021).
31. Green, N.L.: The use of antithesis and other contrastive relations in argumentation. Argument & Computation. 14, 1–16 (2022).
32. Lawrence, J., Visser, J., Reed, C.: Harnessing rhetorical figures for argument mining. Argument & Computation. 8, 289–310 (2017). https://doi.org/10.3233/AAC-170026.
33. Harris, R.A., Di Marco, C., Mehlenbacher, A.R., Clapperton, R., Choi, I., Ruan, S., O'Reilly, C.: A cognitive ontology of rhetorical figures. Presented at the CAOS - Cognition and Ontologies , University of Bath April 20 (2017).
34. Kelly, A.R., Abbott, N.A., Harris, R.A., DiMarco, C.: Toward an ontology of rhetorical figures. In: Anacleto, J.C., de Mattos Fortes, R.P., and Costa, C.J. (eds.) Proceedings of the 28th annual international conference on design of communication, SIGDOC 2010, são carlos, são paulo state, brazil, september 26-29, 2010. pp. 123–130. ACM (2010). https://doi.org/10.1145/1878450.1878471.
35. Harris, R.A., Di Marco, C.: Constructing a rhetorical figuration ontology. In: Symposium on persuasive technology and digital behavior intervention, convention of the society for the study of artificial intelligence and simulation of behaviour (AISB). pp. 47–52. , Edinburgh, Scotland (2009).
36. O'Reilly, C., Wang, Y., Bott, S., Pacheco, P., Black, L.A. (formerly T., Harris, R.A.: A climactic ontology. Presented at the International Association for Cognitive Semiotics , Toronto, Ontario (Ryerson U) July 14 (2018).
37. Black, L.A., Tu, K., O'Reilly, C., Wang, Y., Pacheco, P., Harris, R.A.: An Ontological Approach to Meaning Making through PATH and Gestalt Foregrounding in Climax. American journal of semiotics. 35, 217–249 (2019).
38. Kara-Yakoubian, M., Walker, A.C., Assadourian, G.A., Bisnar Griffin, D., Fugelsang, J., Harris, R.A.: Chiasmus and the Keats effect. Presented at the RhetCanada , Edmonton, Alberta (U of Alberta) June 5 (2021).
39. Harris, R.: Review of Fahnestock, Rhetorical figures in science. Rhetoric Society Quarterly. 31, 89–104 (2001). https://doi.org/10.1080/02773940109391216.
40. Fahnestock, J.: Rhetoric in the Age of Cognitive Science. In: Graff, R. (ed.) The viability of rhetoric. pp. 59–179. State University of New York Press, New York (2005).

41. Mitrović, J., O'Reilly, C., Granitzer, M., Harris, R.A.: Cognitive Modeling in Computational Rhetoric: Litotes, Containment and the Unexcluded Middle. In: Proceedings of the 12th International Conference on Agents and Artificial Intelligence. , Valleta, Malta.
42. Wang, Y., Harris, R.A., Berry, D.: An Ontology for Ploke: Rhetorical Figures of Lexical Repetitions. Presented at the CAOS@JOWO 2021 (Cognition and Ontologies @ Joint Ontologies Workshop) , The Research Centre for Knowledge and Data of the Free University of Bozen-Bolzano September 17 (2021).
43. Harris, R.A.: Chiastic iconicity. In: Fischer, O., Ljunberg, C., Tabakowska, E., and Lenninger, S. (eds.) Iconicity in Cognition and across Semiotic Systems. John Benjamins Publishing Company, Amsterdam (2022).
44. O'Reilly, C., Paurobally, S.: Lassoing rhetoric with OWL and SWRL, (2010).
45. Mladenović, M., Mitrović, J.: Ontology of Rhetorical Figures for Serbian. In: Habernal, I. and Matoušek, V. (eds.) Text, Speech, and Dialogue. pp. 386–393. Springer, Berlin, Heidelberg (2013). https://doi.org/10.1007/978-3-642-40585-3_49.
46. Mitrović, J., O'Reilly, C., Mladenović, M., Handschuh, S.: Ontological representations of rhetorical figures for argument mining. Argument & Computation. 8, 267–287 (2017). https://doi.org/10.3233/AAC-170027.
47. Wang, Y., Kühn, R., Harris, R.A., Mitrović, J., Granitzer, M.: Towards a unified multilingual ontology for rhetorical figures. In: 14th international joint conference on knowledge discovery, knowledge engineering and knowledge management - KEOD. pp. 117–127 (2022).
48. Kühn, R., Mitrovic, J.: Multilingual domain ontologies of rhetorical figures and their applications. In: UniDive 1st general meeting (2023).
49. Ruan, S., Marco, C.D., Harris, R.A.: Rhetorical figure annotation with XML. In: Bex, F., Grasso, F., and Green, N.L. (eds.) Proceedings of the 16th workshop on computational models of natural argument co-located with IJCAI 2016, new york, USA, july 9th, 2016. pp. 23–33. CEUR-WS.org (2016).
50. Atienza, A., Bisnar Griffin, D., Chan, F., Harris, R.A.: GoFigure: Citizen Science meets the gamification of rhetoric. Presented at the Play on! , Montreal, Quebec May 13 (2020).
51. Bisnar Griffin, D., Chan, F., Harris, R.A.: Gamifying the Study of Rhetorical Figures in the Undergraduate Rhetorical Studies Curriculum. Presented at the Northeast Modern Languages Association (NeMLA) , Boston MA USA March 7 (2020).
52. Harris, R.A., Luan, R., Ogunsanya, A.: The future of Machine Learning: Free-range gamesourcing a rhetorical figure database. GI Seed Symposium. , The Games Institute, University of WAterloo (2024).
53. Clapperton, R.: [Thesis directed] Ametros : A Technogenetic Simulation Game for Professional Communication Coursework, (2014).
54. Harris, R.A.: Rhetorical schemes and Construction Grammar: The argument from chiasmus. Constructions and Frames. (Forthcoming).
55. Harris, R.A.: Rules are rules: Rhetorical figures as algorithms. In: Loukanova, R., LeFanu Lumsdaine, P., and Muskens, R. (eds.) Logic and algorithms in computational linguistics. pp. 217–260. Springer, Berlin (2023).
56. Hashemi, R., Chen, H., Harris, R.A.: Define Constructions Before Constructions Define You: Rhetorical Figures in the AB before BA Construction. Rhetor. 9, (2024).
57. Reed, C., Grasso, F.: Recent advances in computational models of natural argument. Int. J. Intell. Syst. 22, 1–15 (2007). https://doi.org/10.1002/int.20187.

A. Online Resources

Our Rhetoricon Database and the related gamification for figure harvesting, GoFigure, are not ready for prime time as we submit this project description, but we invite CMNA habitués to visit and comment upon them so that we can continue to develop our project: https://rhetoricon.uwaterloo.ca/; https://gofigure.uwaterloo.ca/