

Harnessing rhetorical figures for argument mining

A Pilot Study in Relating Figures of Speech to Argument Structure

John Lawrence^a, Jacky Visser^{a,*} and Chris Reed^{a,b}

^a Centre for Argument Technology, University of Dundee, United Kingdom

E-mails: j.lawrence@dundee.ac.uk, j.visser@dundee.ac.uk, c.a.reed@dundee.ac.uk

^b Centre for Argument Technology, Polish Academy of Sciences, Poland

Abstract. The generalised, automated reconstruction of the reasoning structures underlying persuasive communication is an enormously challenging task. While this work in *argument mining* is increasingly informed by the rich tradition of argumentation studies outside the computational field, the rhetorical perspective on argumentation is thus far largely ignored. To explore the application of rhetorical insights in argument mining, we conduct a pilot study on the connection between *rhetorical figures* and argumentation structure. Rhetorical figures are linguistic devices that perform a variety of functions in argumentative discourse. The textual form of some of these figures is easy to identify automatically, such that an established connection between the figure and a preponderance of argumentative content would improve the performance of argument mining techniques. Furthermore, the automated mining of rhetorical figures could be used as an empirical, corpus-based testing ground for the claims made about these figures in the rhetorical literature. In the pilot study, we explore the connection between eight rhetorical figures the forms of which we expect to be relatively easy to identify computationally, and argumentation structure (concretely, we consider the six schemes ‘anadiplosis’, ‘epanaphora’, ‘epistrophe’, ‘epizeuxis’, ‘eutrepismus’, and ‘polyptoton’, and the two tropes ‘antithesis’ and ‘dirimens copulatio’, and relate their occurrences to relations of inference and conflict). The data of the study is collected in the MM2012c corpus of 39,694 words of argumentatively annotated transcripts from the BBC Radio 4’s Moral Maze discussion program. We show that some of the figures indeed correspond to passages of high argumentative density, relative to the text as a whole.

Keywords: Argument mining, argumentation structure, BBC’s Moral Maze, rhetorical figures

1. Introduction

1.1. Argument mining

The continuing growth in the volume of data that humanity produces has driven efforts to unlock the wealth of information this data contains. With respect to textual data, automatic techniques such as opinion mining and sentiment analysis [27,31] allow us to determine the views expressed in communication – for example, whether a product review is positive or negative. Enriching these polarity attributions with the reasons and arguments provided for the viewpoint requires the identification of more complex structural relations between concepts.

*Corresponding author: Jacky Visser, Queen Mother Building, University of Dundee, Dundee DD1 4HN, United Kingdom. Tel.: +44 138286584; E-mail: j.visser@dundee.ac.uk.

Argument(ation) mining [32] addresses this challenge by pursuing the automatic identification of the argumentative structure contained within pieces of natural language text. By automatically identifying this structure and its associated premises and conclusions, we are able to tell not just *what* views are expressed, but also *why* they are held. The field of argument mining has recently enjoyed rapid growth, propelled by three drivers: first, the academic and commercial success of the opinion mining and sentiment analysis techniques upon which argument mining builds; second, a strong commercial appetite for such technologies from companies such as IBM in the context of their focus on cognitive computing and the Watson suite of AI tools; and third, the progress in the (computation-oriented) theoretical understanding of argumentative discourse [4,47] and the development of supporting software tools and infrastructures [34].

The majority of the work carried out to date in the field of argument mining has used either a supervised machine learning approach [14,29,41], or a linguistic rule-based approach [30,51], to determine argumentative function. In both cases, these efforts are constrained by the sparsity of high quality annotated argument data. Whilst resources such as the Internet Argument Corpus (IAC) [52] and AIFdb [25] offer rapidly growing volumes of consistently annotated argumentative analyses, they do not yet provide the large volumes of data computationally required to train a robust classifier, or provide sufficient examples to determine a comprehensive set of rules, particularly when considered in relation to specific topics, institutional contexts or communicative domains.

Previous efforts to overcome this issue have looked at discourse indicators [53] – explicit linguistic expressions signalling the discursive role played by an utterance [49] – as easily identified patterns that require no prior annotation [26,55]. Whilst discourse indicators have been shown to be a reliable way of determining argumentative structure, the majority of arguments is not explicitly signalled by means of such an indicator, limiting their effectiveness in argument mining.

1.2. *Figures of speech in the rhetorical tradition*

The rhetorical study of the effective process of persuasion by means of argumentative discourse stretches back to antiquity, frequently interacting with the logical and dialectical perspectives on argumentation [54]. The rhetorical perspective is often applied to particular communicative contexts or events in particular domains, such as the legal or the political (thereby intersecting with other applied fields such as law and political science) [45,56]. Related to the domain of scientific communication, Fahnestock's [12] work marks a major advance in the rhetorical study of argumentation. Fahnestock makes a compelling case for the conception of rhetorical figures of speech (henceforth 'rhetorical figures') as couplings of linguistic form and function. Drawing on a tradition that links figures to *topoi*, running back to Aristotle [2], Fahnestock argues that figures "map function onto form or perfectly epitomize certain patterns of thought or argument" (p. 26). She demonstrates this claim for a specific group of figures related to organisation. To the extent that the claim is true – that there is, in Fahnestock's terms, a "figural logic" at work in language – the potential for argument mining and other computational explorations of language is promising.

The argumentation-functional approach to rhetorical figures is also exemplified in the integration of the rhetorical perspective in the pragma-dialectical theory of argumentation [46]. In the pragma-dialectical theory, figures are principally considered part of the presentational devices used by arguers to perform argumentative moves. Together with the choice from the topical potential and adaptation to the audience, the presentational device constitutes the three aspects of strategic manoeuvring. An arguer's selection with respect to the three aspects of strategic manoeuvring is motivated by their goal of

defending (or attacking) a standpoint in a way that is both dialectically reasonable and rhetorically effective within the particular communicative context (or activity type) at hand. The strategic function as a means of presenting argumentative moves in discourse has been examined of some specific examples of rhetorical figures, e.g.: *conciliatio* by van Eemeren and Houtlosser [48], *metonymy*, *rhetorical questions*, *praeteritio*, and *hyperbole* by Snoeck Henkemans [37–40], and *litotes* by van Poppel [50].

The most complete compilation of rhetorical figures of which we are aware is the online resource *Silva Rhetoricae* (rhetoric.byu.edu) maintained by Burton [6]. The website assembles a large and diverse set of figures with descriptions, examples and detailed cross-referencing. An alternative, in some sense richer and more ambitious, but also as yet less comprehensive, online resource is the RhetFig website of Harris and Di Marco's group in Waterloo (artsresearch.uwaterloo.ca/rhetfig) [3,24]. Because of their digital format and comprehensiveness, both of these Web-based resources serve as invaluable starting points for the exploration of rhetorical figures in the context of what might be called 'computational rhetoric'.

The *Silva Rhetoricae* and RhetFig websites both suffer from the maladies that plague any such endeavour to compile comprehensive and consistent categorisations of rhetorical figures: the tradition of studying rhetorical figures goes back at least two and a half millennia, crosses many scholarly periods, many languages, and many fields and approaches (for instance, grammatical, psychological, structuralist, and Biblical, to name just the most prominent), and is plagued by terminological confusion, inertia and undisciplined innovation. A taxonomy or ontology of rhetorical figures facilitates the systematic categorisation of the historically compiled, *ad hoc*, and unprincipled collection of rhetorical figures and similar presentational devices. We use the four-way division of Chien and Harris [8,19].¹ Its principal virtues for our purposes are its explicit cognitive foundation and the precise definitions it offers in specific linguistic and semiotic terms. Explicating all of its categories is not relevant to the current paper, but two of the classifications are central for computational purposes: the ancient categories of *scheme* and *trope*. Schemes are defined by Harris [19] as "formal alterations, shifts away from conventional, baseline expectations in the usage of signifiers [vehicles, signantid]" (p. 571) – not to be confused with the notion of 'argument(ation) schemes' as commonly used in argument mining and argumentation theory. Tropes are defined as "conceptual alterations, shifts away from conventional, baseline expectations in the usage of signifieds [interpretants, signata]" ([19], p. 571). For computational purposes, patterns of form are much easier to detect than conceptual ones. For example, a scheme like *polyptoton* (repetition of stems with different affixes: 'hate the sin but not the sinner') is easier to algorithmically detect than a trope like *metaphor* (a cross-domain mapping: 'Juliet is the sun').²

To our knowledge, the first work on computational rhetoric was at the Symposium on Argument & Computation held in 2000 in the Scottish Highlands [35]. In the volume that was the output of the symposium, Crosswhite et al. [9] explore how basic tenets of rhetoric might be accounted for in computational systems. They start with the classical distinction between *ethos*, *pathos* and *logos* as three means of persuasion, and then take aspects of Perelman and Olbrechts-Tyteca's *New Rhetoric* [33] in particular, and formalise them using a context logic. Grasso [17] develops these starting points further, with a particular view to building natural language generation systems that make explicit use of insights from rhetoric.

While both of these works include passing mentions of rhetorical figures, and Grasso emphasises their importance for "creating a computational model of rhetorical argumentation" [17], neither gives

¹Chien and Harris [8] make the three-way distinction between *scheme* (based on forms), *trope* (based on concepts) and *chroma* (based on intentions). Harris [19] adds the additional category of *move*, for discourse-level configurations.

²The introduction to the issue of *Argument & Computation* that our paper appears in, includes a glossary with definitions of the rhetorical figures used in the papers.

any indication of how such a project might proceed. The first work to impose some computational order on figures was initiated by Harris and Di Marco [18]. This ongoing thread of research has more recently started to deliver results that can lay a foundation for detailed text processing (see, e.g., Mladenović and Mitrović's application to Serbian [28], and Gladkova's broader 'conspiracy of features' approach to argument diagnostics [16]), and the mining of rhetorical figures [15,21].

1.3. Motivation

The computational detection of rhetorical figures is tremendously promising, as Crosswhite et al. [9], Grasso [17], Harris and Di Marco [18], and Ruan et al. [36], among others, seem to agree. It remains nevertheless an equally undeveloped area. Any foray, therefore, into the computational detection of figures could be a valuable contribution to argument mining, and argumentation studies and rhetoric in general (whether computational or not). In the current paper we report on a pilot study looking at the relation between arguers' usage of rhetorical figures and the structure of their argumentation.³ If we regard rhetorical figures as linguistic devices used to present moves in argumentative discourse, some such presentational devices can be expected to systematically relate to specific argumentation structures (indeed following the form-function pairing of figures described in rhetorical scholarship).

One of the main challenges in argument mining is the automated recognition of the structure of argumentation. In contrast, it is relatively easy to identify the formal dimension of some rhetorical figures with automated techniques. The identification of figural forms can then be harnessed for argument mining as a feature for machine learning techniques, or equally serve in combination with other linguistic cues, domain knowledge, and common patterns of reasoning to mirror the complex process followed by a human annotator in reconstructing the structure of argumentation. Underlying our assumed connection between the discursive structure characteristic of rhetorical figures, and the structure of argumentative is the claim in Fahnestock's figural logic that specific figural *forms* are systematically associated with corresponding figural *functions* [12] (see also Tindale [44] and Harris [19]). In other words, a figure can be thought of as exhibiting a formal dimension and a functional dimension, with the systematic association of a particular form and a particular function constituting a particular type of rhetorical figure. This assumption forms the basis of our approach to improving the automated search for argumentation structures (related to figural *functions*) by means of searching for dialogical structures (instances of figural *forms*).

While we apply the computational techniques on a limited dataset in our pilot study, the approach can be extended for a more detailed investigation, or for the empirical testing, of the specific correspondences between figural forms and functions as claimed in the rhetorical literature. Finding an instance matching the form of a particular rhetorical figure does not necessarily mean that the figure itself can be said to be activated: when the form occurs without the function (and similarly vice versa, obviously), we might only call this a quasi-instance of the rhetorical figure. The claims about specific form-function pairings of the rhetorical figures should however be grounded in empirical observation. Quantitative evidence could therefore suggest that some of the qualitatively demonstrated pairings should be revised, and the approach we explore in the pilot study suggests a method for the production of such data.

³The focus on argumentation structure means we are concerned with *logos* (persuasion by means of an appeal to reasoning). The function of rhetorical figures can equally be associated with *ethos* (an appeal to the appreciation of the speaker's character) or *pathos* (an appeal to emotions) [43]. Argument mining is, however, presently mostly concerned with *logos* in terms of argumentation – although work on the mining of ethotic attacks and defences [11] shows there are exceptions.

2. Data and methodology

2.1. BBC's moral maze

The data used in this paper is taken from the BBC Radio 4 program *Moral Maze*.⁴ Specifically, we consider transcripts of five episodes from the 2012 summer season. The episodes consecutively deal with the banking system,⁵ the welfare state,⁶ the British empire,⁷ problem families,⁸ and the morality of money.⁹ Each episode follows the same format. The BBC Radio presenter serves as chair and starts the conversation by introducing the topic before getting the opinions of four regular panellists. Three guest participants are then introduced and interviewed in turn by the panellists. This format encourages a regular back-and-forth between all participants, taking turns to speak. The episodes, lasting one hour each, contain on average approximately 600 dialogue contributions, leading to a combined word-count for the five transcripts of 39,694.

2.2. The annotation of argument structure

The argumentative structure of the five *Moral Maze* episodes has been annotated by expert annotators on the basis of Inference Anchoring Theory [5], which provides a framework for the computation-oriented analysis of both the inferential and the dialogical aspects of argumentative discourse. The annotation was done using the OVA (Online Visualisation of Argumentation) annotation software (ova.arg.tech) [22], with the output compiled in the MM2012c corpus (freely available at corpora.aifdb.org/mm2012c) in the graph-based Argument Interchange Format (AIF) standard [7].

As part of the annotation, the transcripts are segmented into the constitutive dialogue units and associated propositional units – in the AIF ontology, these constitute L-nodes (for locutions as dialogue units) and I-nodes (for information in propositional units). Subsequently, the dialogue relations of transition and illocution (respectively TA- and YA-nodes in AIF terms), and propositional relations of inference, conflict and rephrase (RA-, CA-, and MA-nodes) are annotated.

For the purposes of connecting the form of rhetorical figures to their function in terms of argumentation structure, most relevant are the I-nodes as the building blocks (elementary units of propositional content), and the propositional relations of inference between an I-node and another that it supports, conflict between an I-node and another that it contradicts, and rephrase between an I-node and another that it paraphrases. These together constitute the structure of the argumentation. More specifically, in our pilot study, we look at the relative frequencies of the inference and conflict relations both in and around I-nodes associated with occurrences in the dialogue of several types of rhetorical figures' form as an indication of their argumentative function.

2.3. Rhetorical figures

Rhetorical figures constitute a somewhat open class. Despite the aforementioned efforts of Silva Rhetoricae [6] and RhetFig [3], no currently available list of figures seems to be exhaustive. A great

⁴Available online at bbc.co.uk/programmes/b006qk11.

⁵Available online at bbc.co.uk/programmes/b01kbj37.

⁶Available online at bbc.co.uk/programmes/b01k29ph.

⁷Available online at bbc.co.uk/programmes/b01ks9zl.

⁸Available online at bbc.co.uk/programmes/b0110kcc.

⁹Available online at bbc.co.uk/programmes/b01jrjgg.

variety of figures have been named and catalogued based on their form and their rhetorical effects – with the treatment of figures in rhetorical textbooks often limited to considerations of ‘aesthetics’ or ‘emotion’. The result is that many of the catalogues are lengthy and diverse, but not complete. Silva Rhetoricae lists nearly 500 figures, for instance, ranging from those that are familiar and in some cases well-studied in computational linguistics (such as *metaphor* and *metonymy*) to those that are obscure, complex or peculiarly specific (such as *anemographia*, the creation of an illusion of reality through description of the wind; and *antiprosopopoeia*, the representation of persons as inanimate objects). Because many of the figures are known by a variety of names in the rhetorical literature, we strive to provide some principal alternative nomenclature with the figures we focus on. Our main goal in the pilot study is to employ insights about rhetorical figures as a way of improving the automated reconstruction of the structure of argumentation in persuasive discourse (part of the functional dimension). To that end, those rhetorical figures of which the formal dimension is particularly easy to recognise by automatic means are our priority. We will therefore mainly be working with the category of schemes.

The forms of schemes of structured repetition and (to a lesser extent) structured order should often be straightforward to identify automatically, which is why these figures were chosen in previous computational work as well [10,43]. Silva Rhetoricae lists 49 figures of repetition and 25 figures of order. Of the figures of repetition, we exclude the tropes (i.e. those concerned with higher level semantic repetition, such as the repetition of concepts or ideas), because they are more demanding computationally. We also exclude the phonological schemes of repetition (such as the repetition of prosodic inflection), again because of computational considerations. After removing the figures involving synonyms (and near-synonyms), hypernyms and hyponyms from the list, we are left with five schemes of repetition: *anadiplosis*, *epanaphora*, *epistrophe*, *epizeuxis*, and *polyptoton*.

Amongst the figures of order, many require complex knowledge representation (such as *parecbasis*, a departure from the logical order of presentation) or semantic representation (such as *catacosmesis*, the ordering of words from greatest to least in dignity), and are therefore unsuitable for our current purpose. Some of these are tropes, or fall into the categories of chroma or move (see Footnote 1), creating additional computational challenges, but even some of the schemes require complex syntactic processing (such as *hysterologia*, the interposition of a phrase between a preposition and its object). This leaves *eutrepismus* (an explicit numbering of parts), as a figure of order the formal dimension of which should be relatively easy to identify automatically.

While tropes are generally more challenging computationally, some are easier to process than others as a result of conventional explicit lexical signalling of form. *Dirimens copulatio* (the mentioning of counter-considerations), for instance, is a figure of semantic balance which is stereotypically marked by one of only a small number of discourse indicators. Similarly, *antithesis* (a figure of more general semantic opposition) often leverages antonyms, which provide a clear search path.

In our pilot study, we explore a set of eight rhetorical figures, consisting of six schemes (anadiplosis, epanaphora, epistrophe, epizeuxis, eutrepismus and polyptoton) and two tropes (antithesis and dirimens copulatio). In addition to being selected for their clear linguistic signalling and therefore ease of automated processing, these figures represent a diverse range of rhetorical attributes. Our approach is to build recognisers for the formal dimension of each figure (see Section 3) and subsequently explore the relations between occurrences of the forms of the figures and their functional correlation to particular structural constellations of arguments (see Section 4).

3. Defining and identifying rhetorical figures

In this section, we define each of the eight selected rhetorical figures for the purposes of our pilot study on the basis of the definitions provided in the *Silva Rhetoricae* and *RhetFig* repositories [3,6]. We illustrate the defined figures with examples from the MM2012c corpus. Finally, we discuss how we are able to automatically label text spans from the Moral Maze transcripts to be part of the form of the rhetorical figure, and how we can subsequently transfer this labelling of text spans to I-nodes in the MM2012c corpus – recall that in accordance with the AIF standard I-nodes represent information in the graph-based ontology [7]. Because each of the rhetorical figures is considered in isolation, any text span and corresponding I-node(s) can be labelled as part of several figures. This also means that the examples we use can contain instances of various other figures besides the one it illustrates. Within the scope of this pilot study we will not look into the interactions between, and co-occurrences of, the various rhetorical figures.

The definitions we provide lack some of the depth and detail that might be preferable in many other studies relating to rhetorical figures. For our current purposes, however, these definitions should provide enough detail to distinguish them and to serve as a starting point for their operationalisation in terms of criteria for the automated identification of the figural forms in the text. Less strict definitions or more encompassing operationalisations in terms of recognisers would presumably lead to the identification of more potential instances of the figural forms in the text at the expense of precision. For our purpose of improving argument mining by exploring the introduction of rhetorical figures as features, finding only instances of a particular figure (high precision) is more important than finding all instances of the figure (high recall). As an additional feature to improve argument mining, confidence that what we find actually has certain properties (minimisation of false positives) outweighs completeness in locating all instances with those properties (minimisation of false negatives).

3.1. Identifying anadiplosis

Anadiplosis (also known as *reduplicatio*) takes the form of a repetition of the last word (group) from the previous line, clause, or sentence at the beginning of the next. Example (1) is an instance of the figural form of anadiplosis in the corpus, where the word ‘poverty’ recurs both as the last word of the first sentence and as the first word of the last sentence.¹⁰ (The first occurrence of the word ‘poverty’ in Example (1) is not part of the anadiplosis. It is nevertheless part of another figure: *ploce*, the non-place-specific proximal repetition of words.)

- (1) Certain factors make it much more likely that you will not exit poverty, or you won’t prevent **poverty**. **Poverty** being more than income, by the way.

The figural form of anadiplosis is easily identified by splitting the original text into clauses based on punctuation, and then comparing words on either side of these boundaries. To map instances of anadiplosis in the MM2012c corpus to I-nodes, we straightforwardly locate the node or nodes in which the repeated word or words occur.

¹⁰An annotated version of Example (1) is available in OVA at arg.tech/ova-6305.

3.2. Identifying epanaphora

Epanaphora (also known under other names, such as *repetitio* or *anaphora* in the rhetorical literature – the latter being a term we will avoid due to the obvious potential for confusion with the linguistic term) is a figure of repetition in which the same word (or group of words) occurs at the beginning of successive phrases or clauses. In Example (2) from our corpus, ‘It can be’ is used as the opening of three consecutive sentences.¹¹

- (2) [Saving] is not necessarily a moral good. **It can be** a way of demonstrating independence. **It can be** risk averse. **It can be** conservative.

We search for instances of epanaphora forms in the corpus by first locating adjacent clauses that begin with the same word(s), and then filtering the results to remove those where the clauses merely begin with a common word (e.g., ‘the’, ‘a’, and ‘it’) without any further extension with consecutive repeated words. Similarly to the other figures of repetition, instances of epanaphora are mapped to any I-nodes which contain the repeated text. So, returning to Example (2), any I-nodes containing the string ‘it can be’ corresponding to consecutive utterances by a speaker are considered part of this instance of the figure.

3.3. Identifying epistrophe

Epistrophe (also known as *conversio*) is virtually a mirror image of epanaphora. Rather than repeated words (or word groups) at the beginning of consecutive phrases or clauses, with an epistrophe, the repetition occurs at the end. (3) is an example of epistrophe identified in the MM2012c corpus.¹²

- (3) Otherwise you’ve got **slavery**. I mean, if you get into debt, and you can never get out of it, that’s **slavery**.

The mapping from epistrophe to I-nodes is carried out in exactly the same way as for epanaphora. In the case of Example (3), all I-nodes within one turn (or one speaker’s consecutive turns) containing the word ‘slavery’ are considered part of the identified instance of epistrophe.

3.4. Identifying epizeuxis

Epizeuxis (also known as *geminatio*) is the repetition of a word or small group of words, with no other words in between, as illustrated by Example (4) from our corpus.¹³ The figure is closely related to anadiplosis (see Section 3.1), but does not require the repeated words (or word groups) to be located at the end and subsequent start of two consecutive sentences.

- (4) Well, as we just heard, because the system seems to be **very, very** heavily weighted against them. A lot of savers have spent much of their life putting money by, in the hope that they can provide for themselves in their old age.

To locate examples of epizeuxis in our corpus, we consider, in turn, unigrams, bigrams and trigrams and search for locations where at least two successive occurrences are found. These examples are mapped to the single I-node in which the repeated text appears.

¹¹An annotated version of Example (2) is available in OVA at arg.tech/ova-6133.

¹²An annotated version of Example (3) is available in OVA at arg.tech/ova-6814.

¹³An annotated version of Example (4) is available in OVA at arg.tech/ova-6122.

3.5. Identifying eutrepismus

Eutrepismus (also known as *ordinatio*) is the numbering and ordering of parts under consideration. In Example (5) the words ‘firstly’ and ‘secondly’ explicitly indicate an enumeration of two suggested causes for the ‘welfarism’ under discussion.¹⁴

- (5) The growth of what you call ‘welfarism’ is to do with, **firstly** the disappearance of so many skilled middle income jobs; and **secondly** due to the failure to build social housing.

Instances of eutrepismus are located by searching each speaker’s turn for occurrences of both of the words (roots) ‘first’ and ‘second’. This includes a derived form, such as the ‘firstly’ and ‘secondly’ in Example (5). Erring on the safe side, we presently do not consider common alternative explicit ordering indicators, such as ‘a’/‘b’, or ‘one’/‘two’. The reason for this exclusion is to arrive at more reliable results, the prevalence of the alternative indicators in other discursive uses, not indicative of eutrepismus, has the potential to greatly increase the number of false positives (i.e. I-nodes classified as eutrepismus, that actually are not).

Of the rhetorical figures we consider in this pilot study, eutrepismus is one of the easiest to search for automatically, due to the uncomplicated lexical cues. The mapping of instances of eutrepismus onto its corresponding I-nodes, however, is not as straightforward as for some of the other figures. The numberings can often be analysed as structural indicators outside the bounds of the argumentation structure. This means that the numbering words will not be included in the I-nodes of the annotation, rather being incorporated in the annotation as indicative of discourse relations (S-nodes for schemes of dialogue protocol, illocutionary connection, and inferential, conflict, and rephrase relations). For this reason, we look at each number in turn and take either the node in which it occurs, or, if a number occurs ‘between’ I-nodes, then we take the consecutive I-node (of which the corresponding text span, the number is most closely related to).

3.6. Identifying polyptoton

Polyptoton (also known as *adnominatio*) is the repetition of words derived from the same root but with different morphology (possibly non-overt). The occurrence of ‘individualistic’, ‘individual’, and ‘individualised’ all based on the common ‘individu-’ in (6) constitute an example of polyptoton.¹⁵

- (6) People find it very difficult in a very strongly **individualistic** culture to understand what’s being meant by the phrase ‘the common good’. So they find it hard to put the two at odds, and to understand that they ought to be leaning towards the common good and not towards personal and **individual** interest. Because that’s a hard concept to get hold of in a very **individualised** society, and of course the free market system itself is based up on a very **individualistic** anthropology.

In order to locate occurrences of polyptoton in our test data, we consider each turn in the transcript. For each individual turn, we use the Python Natural Language Toolkit (nltk.org) Porter Stemmer to determine the root form of each word, and, where the same root occurred at least twice but with different endings, we consider that turn to contain an instance of polyptoton. Once we identify the presence of polyptoton within a specific turn in the dialogue, we consider all I-nodes within that turn which contain a word derived from the identified root as corresponding to the same instance of the figure.

¹⁴An annotated version of Example (5) is available in OVA at arg.tech/ova-6278.

¹⁵An annotated version of Example (6) is available in OVA at arg.tech/ova-5641.

3.7. Identifying antithesis

Antithesis (also known as *contentio*) is the juxtaposition of contrasting words or ideas, often, though not always, within parallel grammatical structures. In (7), we see an example of such parallel use of the antonyms ‘top’ and ‘bottom’ from the corpus.¹⁶

- (7) There is lots of research been done to how cultures are being changed; it’s very difficult, but one things is very clear: if there is something rotten at the **top**, there is no hope for reform at the **bottom**.

Like many tropes, antithesis is a complex figure with many variants. As a result of the limited scope of our current pilot study, we confine our search to antonyms, temporarily disregarding relevant distinctions between, e.g., contradictory and contrary positions.¹⁷ Again, we look at individual turns in the dialogue to locate instances of the figure. For each turn, we first remove any common English stop-words (using the list from the Natural Language Toolkit at nltk.org), and then use WordNet (wordnet.princeton.edu) to collect antonyms for each of the remaining words. Where a word and one of its antonyms are both found to occur in the same turn, we consider the corresponding I-node(s) as containing an instance of antithesis. Because one dialogue turn can extend over several I-nodes, the text spans in the original transcript that correspond to these I-nodes could be non-adjacent.

3.8. Identifying *dirimens copulatio*

Dirimens copulatio (Latin for “a joining that interrupts”) is a figure by which one juxtaposes one statement with a qualifying statement in a superficially negative environment. An example of this rhetorical figure can be found in Example (8) from the corpus.¹⁸

- (8) Trust **not only** requires moral integrity, **but** also competence. And I think that competence is lacking in the banking industry.

Instances of *dirimens copulatio* are commonly conveyed by ‘not only...but also’ clauses. We locate instances of the figure in the MM2012c corpus by searching for sentences which contain one of the bigrams ‘not only’, ‘not just’, or ‘not merely’ with a subsequent occurrence of the word ‘but’. We map these onto I-nodes by first taking the text from the beginning of the clause in which the bigram appears until the end of the clause in which ‘but’ appears, and then considering any I-nodes which this text overlaps as corresponding to the figure. This may be either a single node if the figure is being used purely for emphasis of the original point, or two nodes if the parts are more noticeably in contrast.

4. Analysis of the results and discussion

Before moving on to the discussion of the individual results for each of the eight figures, we have some observations about the corpus as a whole. Table 1 summarises the frequencies of the identified forms of the eight types of rhetorical figure within the text. The first column shows the number of dialogue turns resembling the formal dimension of the figure. The second column gives the number of I-nodes onto which the figure maps (details of how occurrences of each figure are mapped onto specific I-nodes are

¹⁶An annotated version of Example (7) is available in OVA at arg.tech/ova-5782.

¹⁷Fahnestock [12] discusses such issues relating to antithesis, as well as ‘double antitheses’, in detail (pp. 45–85).

¹⁸An annotated version of Example (8) is available in OVA at arg.tech/ova-5627.

Table 1

The count of rhetorical figures in the corpus, in terms of number of dialogue turns, number of I-nodes, and number of words

	Turns	I-nodes	Words
MM2012c Corpus	926	2,037	39,694
Anadiplosis	3	7	96
Epanaphora	11	21	285
Epistrophe	10	20	281
Epizeuxis	17	15	221
Eutrepismus	3	5	70
Polyptoton	7	24	307
Antithesis	68	98	1,316
Dirimens Copulatio	7	8	128

Table 2

The connection between rhetorical figures and argument structure, in terms of average number of support and conflict relations, and counts of incoming and outgoing relations

	I-nodes	Support	Incoming	Outgoing	Conflict	Incoming	Outgoing
MM2012c Corpus	2,037	0.795	–	–	0.159	–	–
Anadiplosis	7	0.334	0	1	1.667	1	4
Epanaphora	21	0.894	4	10	0	0	0
Epistrophe	20	0.916	5	7	0.134	1	2
Epizeuxis	15	0.854	7	7	0.176	0	3
Eutrepismus	5	1.334	2	4	0.167	0	1
Polyptoton	24	1.119	4	7	0.202	1	3
Antithesis	98	0.853	23	28	0.147	3	8
Dirimens Copulatio	8	0.929	5	3	0.571	2	2

provided in the subsections below). Finally, the third column shows a word-count of the sentential text spans involved in the figure. The first row of the table provides these counts for the corpus as a whole, irrespective of the occurrence of rhetorical figures.

The automated recognisers used to identify the forms of the rhetorical figures as occurring in the text return only and all actual instances of the forms as operationalised on the basis of the definitions provided in Section 3. After qualitative analysis – and we will discuss this in individual cases later as well – it can turn out that a piece of text indeed exhibits the form of a figure, but not the associated function. Such quasi-instances can provide the backbone of a study into the form-function couplings as described in the rhetorical literature, but for our primary aim of aiding argument mining, any insight into the effect in terms of argumentative density surrounding the figural forms is useful.

In Table 2 we show the support and conflict relations associated with each of the I-nodes identified as corresponding to instances of the forms of the rhetorical figures we consider. The first column gives the number of I-nodes onto which the figural forms map. The remaining columns, relating to the functional dimension of the figures, are split into two groups of three. The first group deals with the effect on support relations and the second with conflict relations. Each group reports, first, the average number of support or conflict relations that are connected to each I-node, second, the total numbers of these relations that are incoming (i.e. from a node outside of the figure to a node inside the figure; indicating a conclusion), and third, the total number of these relations that are outgoing (i.e. from a node inside the

figure to a node outside of the figure; indicating an argument or criticism). Again, the first row of the table is dedicated to the MM2012c corpus as a whole, showing the average rate of conflict and support for I-nodes in this corpus. These averages provide a baseline to which the values for each rhetorical figure can be compared. (There are no values for incoming and outgoing connections of the entire corpus, as all the relations are internal to the corpus and do not reach any I-nodes outside the corpus and can therefore not be incoming or outgoing.)

To illustrate how these calculations work, let us consider Example (9).¹⁹ The words ‘poverty’ and ‘wealth’ are recognised as antonyms, leading to an identification of antithesis.

- (9) Well that’s very good of you, but isn’t there an element in this of Christianity fetishising **poverty**? Doesn’t Christianity have a problem with **wealth** creation?

The two I-nodes that are part of the rhetorical figure are indicated with a thick border in Fig. 1, a diagrammatic visualisation of the argumentation structure. In this case, we can see that both of the nodes that are part of the antithesis, have one relation of support (‘Default Inference’) with a node that is not part of the figure, and no conflict relations. Discounting the figure-internal support relation, the average number of support relations is 1 and the average number of conflict relations is 0. Of the external support relations, one is an incoming connection (from ‘I think Christianity has still not got its head around making money.’ to ‘Christianity has a problem with wealth creation’) and one is an outgoing connection (from ‘there is an element in this of Christianity fetishising poverty’ to ‘you famously welcomed the Occupy protesters to St Paul’s Cathedral’). From the diagram, it is clear that there is one more support relation, connecting the two nodes that are part of the antithesis (from ‘Christianity has a problem with wealth creation’ to ‘there is an element in this of Christianity fetishising poverty’). Because such figure-internal relations are only possible for figures with more than one I-node, they are not presented in the table but discussed in the subsections below where relevant.

4.1. Results on anadiplosis

Anadiplosis is one of the rhetorical figures that we least encountered in the MM2012c corpus. It is, however, striking how low the number of support connections is and how high the number of conflict connections. Whilst these numbers seem interesting, they are the result of a very small number of examples. Upon inspection of these examples, there seems to be no discernible reason for the increase in conflict and decrease in support beyond random chance. For example, the instance of anadiplosis in Example (1) (Section 3.1) is analysed as a single I-node with the statement ‘Poverty being more than income’ merely a clarification of the definition of ‘poverty’. This single node happens to be in conflict with another, but this does not appear to be due to the presence of anadiplosis. Perhaps most notable about anadiplosis, is how rare instances of the figural form are in comparison to the related figures epanaphora and epistrophe.

The particular function that anadiplosis fulfils in argumentative discourse is not well described. Fahnestock [12] only mentions it in connection with the figure *gradatio*, effectively a sequence of anadiploses. She notes that anadiplosis serves to draw “out one element from the previous element” (p. 94), which we take to mean that it selects some semantic property (or properties) from the final word(s) of the first clause to emphasise or qualify by means of the initial word(s) of the second clause. Example (1), used to illustrate anadiplosis in Section 3.1, is a clear example of this use: elaborating that poverty should be understood to extend beyond economic aspects.

¹⁹An annotated version of Example (9) is available in OVA at arg.tech/ova-5611.

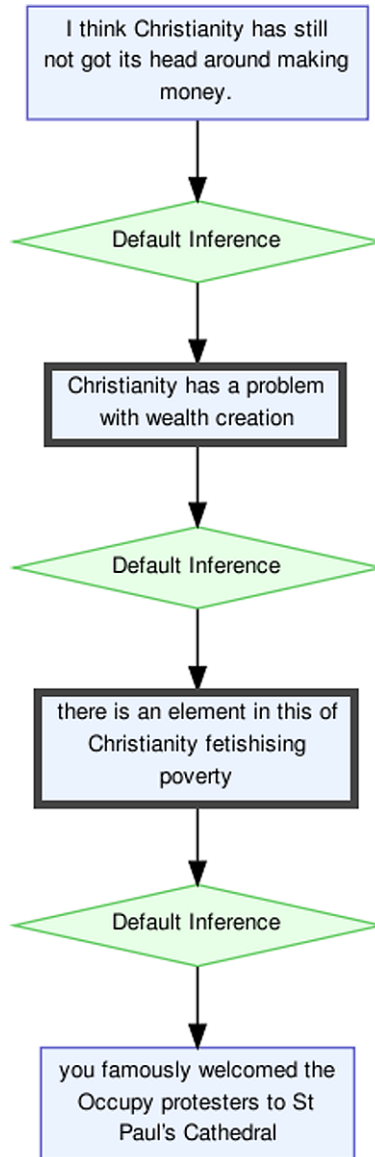


Fig. 1. Argumentation structure corresponding to Example (9) of antithesis (aifdb.org/argview/5611).

4.2. Results on epanaphora

Whilst the counts of support relations for epanaphora in Table 2 are close to those for the corpus as a whole, it is striking that there are no related instances of conflict. It is possible that this is a result of the scarcity of examples, combined with the relatively low frequency of conflict relations in the corpus as a whole. It is nevertheless something that would be useful to investigate further in future work.

A closer inspection of the argumentation structure shows that an epanaphora is a good indicator of several premises working together to support a conclusion (co-ordinative or convergent argumentation). This is illustrated by Fig. 2, which shows the argumentation structure of Example (2) (from Section 3.2)

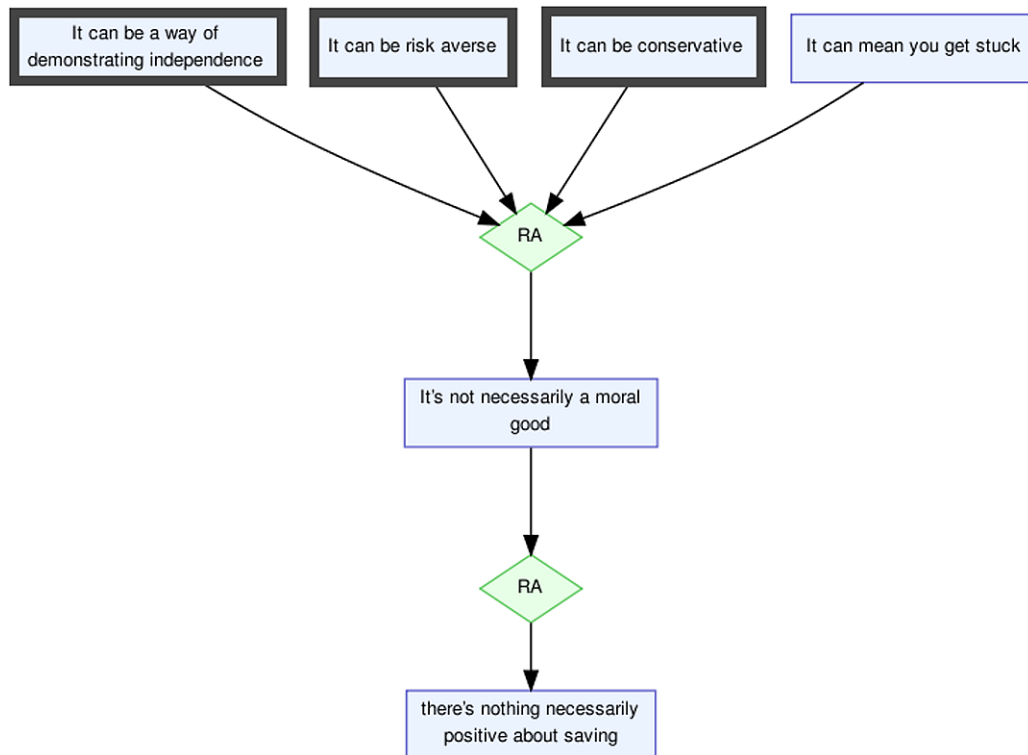


Fig. 2. Argumentation structure corresponding to Example (2) of epanaphora (aifdb.org/argview/6133).

– incidentally highlighting that the word group “It can be” was selected as constitutive of the epanaphora over the less involving repetitions of “It” and “It can”. The hypothesis that epanaphora is used to present a series of related premises is further supported by the relatively high rate of outgoing support connections, compared to those which are incoming.

Fahnestock [13] identifies epanaphora (together with epistrophe, discussed below) with “the argument form . . . comparison, induction, and eduction” (p. 231), to which we would add contrast. Figures of initial repetition frame the new material in semantically parallel ways – in English, with a default subject-verb-object sentence structure, one might conventionally think of a series of novel predicates framed in terms of the same subject). The three occurrences of ‘It can be’ in Example (2) (indicated with thick borders in Fig. 2) provide a group of alternatives, building away from the position that saving is an inherent moral good. We also note that the last two predicates have virtually synonymous interpretations, reinforcing the deontological conservatism of saving. The number of repetitions could indicate what function the epanaphora fulfils, with two occurrences of the relevant item tending toward contrast or comparison, and three or more tending toward induction or eduction.

4.3. Results on epistrophe

Whilst epistrophe and epanaphora are in many respects similar figures of repetition, the results in Table 2 at first glance suggest that epistrophe is, in contrast to epanaphora, actually not a good indicator of several premises working together to support a conclusion. On closer inspection, however, this turns out not to necessarily be the case.

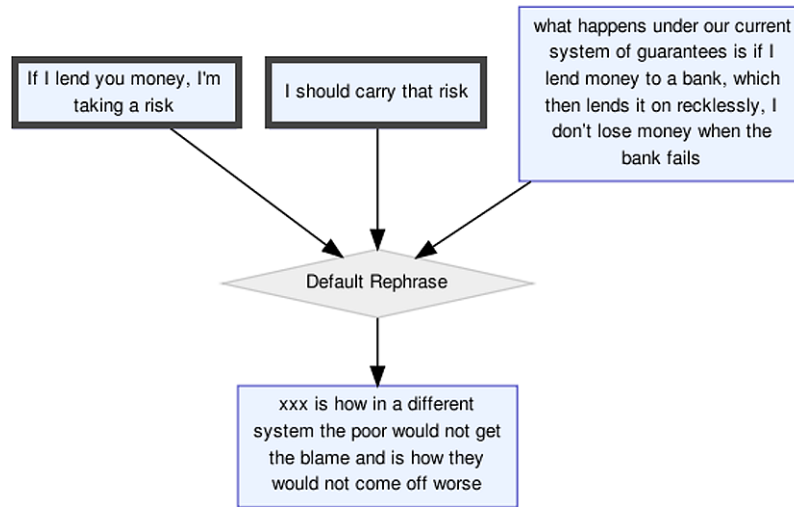


Fig. 3. Argumentation structure corresponding to Example (10) of epistrophe (aifdb.org/argview/6222).

- (10) If I lend you money, I'm taking a **risk**. And I should carry that **risk**. Now, what happens under our current system of guarantees is if I lend money to a bank, which then lends it on recklessly, I don't lose money when the bank fails.

If we examine the structural visualisation of Example (10) in Fig. 3, we can see that the thickly bordered I-nodes corresponding to the figure are once again in a sibling relationship.²⁰ While this constellation is similar to that in Example (2) of Epanaphora, the annotation shows the I-nodes as elements of a rephrase relation rather than support and they are therefore not identified as part of the support and conflict relations that we are focusing on.

In fact, for epistrophe, 6 out of the 20 nodes identified (30%) are siblings which is very comparable to the 7 out of 21 nodes (33.3%) for epanaphora. Both of these results are substantially above the rate at which siblings occur in the corpus as a whole (15.1%).

Like epanaphora, the epistrophe figure is related by Fahnestock to comparison, induction, and education, to which we added contrast (in Section 4.2). However, in Example (3), an instance of epistrophe from the corpus that we used as illustration in Section 3.3, the repetition of the word 'slavery' functions rather as part of an elaboration or explanation.

4.4. Results on epizeuxis

As in Example (4) given in Section 3.4, by far the most common lexical indicator of epizeuxis in our corpus is 'very, very' with 11 out of a total of 17 instances using this wording. It is perhaps not surprising that statements made with such vehemence, whilst often in conflict with a previous point, are themselves conflicted with less. They might either be something which the speaker feels is an unarguable point, or something which the hearer feels less inclined to dispute due to the way in which they have been presented (both of these cases are related to the potentially fallacious immunisation of standpoints or evasion of the burden of proof). This interpretation is supported by the numbers in Table 2, where we can see that, although the average number of conflict connections (0.176) is close to that for the corpus

²⁰An annotated version of Example (10) is available in OVA at arg.tech/ova-6222.

as a whole (0.159), all of these conflict relations are outgoing. This means that the identified instances of epizeuxis in the corpus only argumentatively attack other positions, without attracting any objection themselves.

For supporting connections the levels are comparable to those of the corpus as a whole. The even balance between incoming and outgoing support suggests that epizeuxis is equally used to agree with another point or to introduce a strong point that is then backed up with additional reasons.

Silva Rhetorica similarly describes the most common function of epizeuxis as adding vehemence or emphasis to an expression. Example (4) is a clear case in point of such functionality. The ‘very, very’ in (4) is an example of a rhetorical figure that has become so entrenched in everyday language that it goes nearly unnoticed – similar to, e.g., non-deliberate, dead metaphors.²¹

4.5. Results on eutrepismus

We would generally expect that the numbering and ordering of parts would often correspond to a list of reasons for adopting a particular standpoint (or, conversely, to a list of criticisms of an opponent’s standpoint). We can see that in Example (5) (Section 3.5), the utterances ‘the disappearance of so many skilled jobs’ and ‘the failure to build social housing’ are indeed being given as two independent reasons for ‘the growth of what you call ‘welfarism’’. Eutrepismus differs, we speculate, from the functions of repetition figures like epanaphora and epistrophe, which suggest open-endedness (our example of epanaphora (2) in Section 3.2, for instance, seems to imply that there are more possible alternatives), because the enumeration suggests an exhaustive itemisation (even if it only contains two elements). The use of numbers with this figure (rather than letters) can also function to imply a ranking on the ordered reasons, claims, evidence, etc.

Table 2 also shows that eutrepismus is a strong indicator of support, with an overall rate of 1.334 and, even more notably, four outgoing support relations from only five I-nodes. Whilst these numbers are interesting and justify an intuitive sense of the function of eutrepismus, it should be noted that the number of examples in this case is small and further investigation on a larger corpus would be required before this claim could be advanced reliably.

4.6. Results on polyptoton

Table 2 shows that the frequency of support relations connected to polyptotonic nodes is strikingly high (1.119). Returning to Example (6) from Section 3.6, we can inspect the corresponding argumentation structure more closely. In Fig. 4, we can see that there are multiple I-nodes which constitute the structure, and that the majority of the support is occurring between these nodes. A similar point is worth noting for conflict relations. Although the conflict rate (0.202) is generally only slightly higher than that for the corpus as a whole (0.159), none of this conflict occurs between the I-nodes inside the rhetorical figure.

Of all the figures which we consider in the current pilot study, polyptoton is the most spread out across the I-nodes in a turn: 24 I-nodes originating from just 7 turns. In all cases these nodes are being used to build an argument, which is strongly supported internally, even though it may as a whole be in conflict with external nodes that are not part of the rhetorical figure.

²¹In general, distinguishing between accidental and deliberate uses of figures of speech goes beyond the scope of the current paper. The topic is taken up in the introduction to the issue of *Argument & Computation* that the paper appears in.

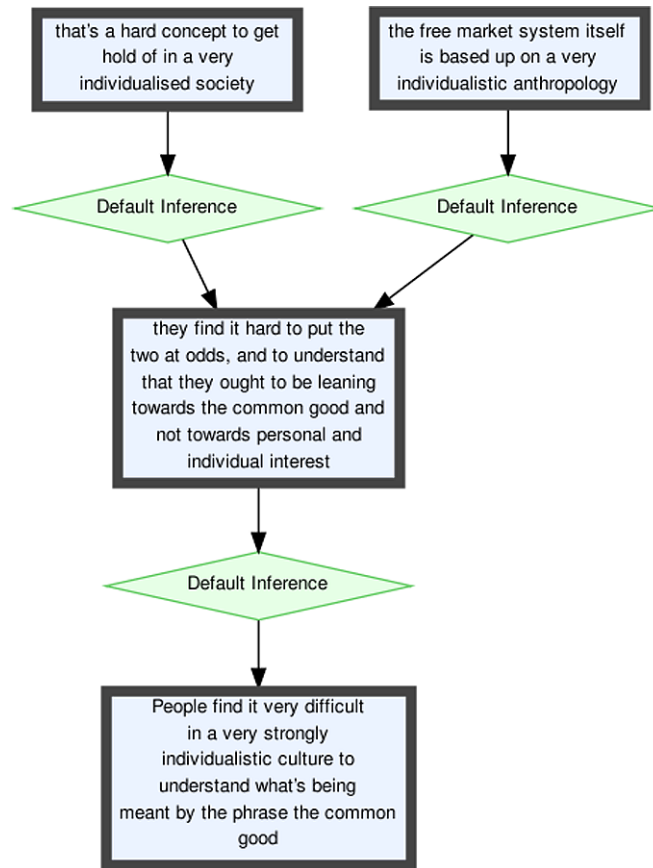


Fig. 4. Argumentation structure corresponding to Example (6) of polyptoton (aifdb.org/argview/5641).

Polyptoton features prominently in Fahnestock’s [12] work. She describes as one of the main functions of polyptoton the transferring claims attached to one form of a word to another by means of association. The rhetorical figure is also found to function as emphasis of complexity, simplicity, irony, or even paradox with respect to a concept evoked by the root. Example (6) does not, however, show the transference that Fahnestock suggests, nor even offer any variation on the theme of individuality – possibly as a result of the redundancy of much English morphology, also identified by Fahnestock. All the variations on ‘individual’ in (6) are adjectives, and synonymous, applied to clearly related nouns (‘culture’, ‘interest’, ‘society’, and ‘anthropology’). Because of the invariability of both the linguistic and the conceptual aspect, no transfer of claims is observed. Since Example (6) does instantiate the form of polyptoton, but lacks its recognised figural function, some would analyse it as a non-figural instance of the form of this particular rhetorical figure (or revert to a classification as plocé).

4.7. Results on antithesis

Like polyptoton, antithesis is one of the core examples discussed by Fahnestock [12]. Following Aristotle [2], she identifies arguing from opposites as one of the primary functions of antithesis. The juxtaposition of ‘top’ and ‘bottom’ in Example (7) from Section 3.7 realises this function. Although we might expect that the introduction of a contrast by means of the juxtaposition of two antonyms in antithesis

would commonly indicate the presence of conflict between two positions, the results in Table 2 show that the frequency of conflict connections is actually slightly below that for the corpus as a whole. If we look more closely at some of the forms of antithesis identified in the corpus, the potential reasons for the result become clearer.

In an example of an automatically recognised instance of the figural form of antithesis, Example (11), the words ‘ethical’ and ‘unethical’ are correctly identified as being antonyms.²² In this example, however, the addition of ‘problems’ to the first part of the *prima facie* antithesis leads to a conceptual inversion of polarity. The result is rather a non-figurative instance of ‘pseudo-antithesis’, as both parts refer to the unethical character of investment banks rather than introducing any contrast. This is also reflected in the corresponding argumentation structure, in which the two I-nodes identified are in a support relationship.

- (11) Investment banks that have highly public **ethical** problems don’t lose their clients. So people are still going to them, probably because they are **unethical** rather than despite it.

Example (12) is another instance fitting the figural form of antithesis adding to the possible explanation for the surprisingly low frequency of conflict relations associated with this rhetorical figure.²³ In (12) we find the antonyms ‘external’ and ‘internal’, but in terms of argumentation no real connection is made between these two concepts: they are neither supporting nor conflicting with each other, but rather express a notion of comprehensiveness.

- (12) I think there has to be an **external** push as well as an **internal** code.

4.8. Results on *dirimens copulatio*

While both antithesis and *dirimens copulatio* function by means of juxtaposition, the effect of the latter is an increase in detail or specificity, rather than an foregrounding of opposition. This is realised through the negation of the absence of detail (signalled by ‘not only’) and the subsequent rectification by supplying the conceptual omission (signalled by ‘but also’). The function of *dirimens copulatio* is to provide balance or division (increasing the salience of the quality that is singled out). In Example (8) from Section 3.8, the argument is that while moral integrity is a necessary condition for trust (in the banking sector), on its own it is not a sufficient condition for trust, as it should be supplemented by competence (the omitted conceptual detail).²⁴

Instances of *dirimens copulatio* in the MM2012c corpus follow one of two general patterns. The figure is often used either to introduce a new concept, often in contrast to an existing point (such as in Example (8)), or to amplify an existing point, making it appear more substantive and striking.

- (13) Acquiring debt **not only** lost the stigma that it had before **but** became actually the thing to do.

In Example (13), the speaker is amplifying the first point rather than seeking to introduce a new concept.²⁵ If the speaker is introducing a new point then this may subsequently be supported, or it may rather be attacked by the hearer. If the function of *dirimens copulatio* is amplification, then the speaker

²²An annotated version of Example (11) is available in OVA at arg.tech/ova-5601.

²³An annotated version of Example (12) is available in OVA at arg.tech/ova-5608.

²⁴We were alerted to an interesting aside in relation to Example (8) by Randy Harris: the speaker is, perhaps unwittingly, identifying two of the three attributes Aristotle [2] insists on for ethotic credibility, virtue and competence (*arête* and *phronesis*) (1378^a).

²⁵An annotated version of Example (13) is available in OVA at arg.tech/ova-6122.

usually supports or criticises some existing standpoint which has been introduced before. Indeed, we can see from Table 2 that, for this figure, both the support rate (0.929) and (especially) the conflict rate (0.571) are substantially above that of the corpus as a whole (0.795 and 0.159, respectively).

Furthermore, we can see that this is the only rhetorical figure, out of those that we take into account as part of this pilot study, where the number of incoming support connections is greater than that of outgoing support connections. This could reflect that in case *dirimens copulatio* is indeed used to forcefully introduce a claim, the speaker anticipates this claim to be met with doubt, and therefore supports it with argumentation.²⁶ The high levels of support and conflict align with the idea that *dirimens copulatio* demands attention, and suggests that where this figure occurs, not only will it likely contribute to the argumentation structure, it will also likely be a relatively central point in the dialogue as a whole – potentially a useful characteristic for automated text summarisation [1].

5. Conclusion

Rhetorical figures have traditionally been studied mainly in conventionalised monological contexts (chiefly public orations, literature and poetry); in texts that often are prepared in advance and go through extensive revision. This focus on ‘static’ monological discourse is also prevalent in the limited existing computational work on rhetorical figures (see, e.g., Java’s [23] figure mining approach to authorship attribution of prose). The discourse in our corpus, however, is constructed in a dynamic dialogical setting involving multiple participants. While we can expect the pundits and professional debaters participating in the radio programs to have prepared some of their contributions in advance, these will not have undergone the elaborate crafting and editing that is characteristic of argumentative monologue in many communicative domains (e.g., political speeches).

Despite the presumed lack of pre-planning in this type of argumentative activity, our pilot study shows that (the forms of) rhetorical figures are still prevalent. Although no single figure discussed here comprises a substantial proportion of the entire text, combined these eight figures (selected out of the hundreds catalogued in the rhetorical tradition) cover almost fifteen percent of the entire corpus. Rather than used intentionally to engender some particular rhetorical effect, we can in part see the figures as reflective of cognitive predispositions and functional imperatives [20] – extending the suggested distinction between deliberate and non-deliberate metaphors to all figures [42]. The kind of corpus and approach we use in this pilot study provide an important testing ground in this respect. Considering that we are currently only taking into account eight of the nearly 500 rhetorical figures on the *Silva Rhetoricae* list, the observed prevalence alone is sufficient warrant for our pilot study and successive investigations.

The study presented is clearly of a preliminary nature. Our goal in the pilot study has been to highlight the value and importance of the area of rhetorical figures for argument mining, and to explore the connection between the mining of forms of rhetorical figures and their function in realising argumentation structure. Where the primary function of the figure is claimed by rhetoricians to be affecting the argumentation structure, the automated corpus-based methods we present can be employed to test the established form-function pairings of the figures on quantitative empirical grounds. For functions that reach beyond argumentative structure, the methods can be used to generate large sets of examples of forms of a particular figure from corpora of persuasive texts, ready for qualitative manual analysis with respect to rhetorical function. Additionally, the corpus-based method we outline can be used to

²⁶In this respect, it might be interesting to explore the relation between *dirimens copulatio* and *prolepsis*, as discussed in Mehlenbacher’s contribution to the present issue of *Argument & Computation*.

quantitatively study the comparative predominance of specific types or classes of rhetorical figures in particular communicative contexts – comparing monological to dialogical contexts, or highly institutionalised to less strictly conventionalised contexts, or investigating any other characteristic distinctions between different activity types.

Our primary objective, however, has been exploring the use of figural forms as a resource for argument mining. On the assumption that the figures can play a functional role in establishing argumentation structure, the formal dimension of rhetorical figures can be used, e.g., as a feature in machine learning approaches to argument mining. With a particular focus on concerted approaches to argument mining – in which diverse features and techniques, and insights from various disciplines and perspectives are combined to achieve the best results [26] – the first step would be to show that the forms of rhetorical figures can be identified automatically. As we have demonstrated, indeed, the very definition of various figures serves as the algorithm for their identification. This leads to an interesting corollary: the standard approach to measuring algorithm success – comparison with a human-annotated gold standard – will not work here. In contrast, we would expect machine identification of the formal dimension of rhetorical figures to be *more* reliable than human annotation (an observation also made by Strommer [43]).

As a second step, we have shown that the consideration of rhetorical figures in argument mining enables the formulation of new and intriguing hypotheses. Does polyptoton co-occur with argumentative support and thereby act as a strong indicator of inference? Might antithesis (surprisingly) act as a weak contra-indicator of conflict? Etc.

The final step is to substantiate or repudiate such hypotheses. While our pilot study explores this direction, addressing this step has not been our primary goal here. Such testing of rhetorically grounded hypotheses represents a vast research project. Any one of the types of rhetorical figure could be interestingly challenging to identify based on its form and highly correlated functionally with some aspect of argumentative structure. With some 500 figures to choose from, this represents an extremely rich seam for argument mining.

Acknowledgements

This research was supported by the Engineering and Physical Sciences Research Council (EPSRC) in the UK under grant EP/N014871/1. We would like to thank Randy Harris for prior discussions and suggestions, and two anonymous reviewers for their commentary, all of which greatly improved the paper.

References

- [1] M. Alliheedi and C. Di Marco, Rhetorical figuration as a metric in text summarization, in: *Advances in Artificial Intelligence*, Lecture Notes in Computer Science, Vol. 8436, 2014, pp. 13–22.
- [2] Aristotle and G.A. Kennedy, *On Rhetoric: A Theory of Civic Discourse*, Oxford University Press, 1991.
- [3] B. Bowden, M.-A. Pilon and R. Harris, The Rhetorical Figure Ontology Project, <https://artsresearch.uwaterloo.ca/rhetfig/>.
- [4] K. Budzynska, M. Janier, C. Reed, P. Saint-Dizier, M. Stede and O. Yaskorska, A model for processing illocutionary structures and argumentation in debates, in: *Proceedings of the 9th Edition of the Language Resources and Evaluation Conference (LREC)*, 2014, pp. 917–924.
- [5] K. Budzynska and C. Reed, Whence inference, Technical report, University of Dundee, 2011.
- [6] G. Burton, Silva Rhetoricae: The forest of rhetoric, <http://rhetoric.byu.edu>.
- [7] C. Chesñevar, S. Modgil, I. Rahwan, C. Reed, G. Simari, M. South, G. Vreeswijk, S. Willmott et al., Towards an argument interchange format, *The Knowledge Engineering Review* **21**(4) (2006), 293–316. doi:10.1017/S0269888906001044.

- [8] L. Chien and R.A. Harris, Scheme trope chroma chengyu: Figuration in Chinese four-character idioms, *Cognitive Semiotics* **10**(6) (2011), 155–178. doi:10.3726/81610_155.
- [9] J. Crosswhite, J. Fox, C. Reed, T. Scaltsas and S. Stumpf, Computational models of rhetorical argument, in: *Argumentation Machines: New Frontiers in Argument and Computation*, C. Reed and T. Norman, eds, Kluwer, 2003, pp. 175–209.
- [10] M. Dubremetz and J. Nivre, Rhetorical figure detection: The case of chiasmus, in: *Proceedings of NAACL-HLT Fourth Workshop on Computational Linguistics for Literature*, ACL, 2015, pp. 23–31.
- [11] R. Duthie, K. Budzynska and C. Reed, Mining ethos in political debate, in: *Proceedings of the Sixth International Conference on Computational Models of Argument (COMMA 2016)*, Potsdam, Germany, P. Baroni, T.F. Gordon, T. Scheffler and M. Stede, eds, IOS Press, 2016, pp. 299–310.
- [12] J. Fahnestock, *Rhetorical Figures in Science*, Oxford University Press, 1999.
- [13] J. Fahnestock, *Rhetorical Style: The Uses of Language in Persuasion*, Oxford University Press, 2011. doi:10.1093/acprof:oso/9780199764129.001.0001.
- [14] V.W. Feng and G. Hirst, Classifying arguments by scheme, in: *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, Vol. 1, Association for Computational Linguistics, 2011, pp. 987–996.
- [15] J. Gawryjolek, C. Di Marco and R.A. Harris, An annotation tool for automatically detecting rhetorical figures, in: *CMNA (Computational Models of Natural Argument)*, 2009.
- [16] O. Gladkova, C. Di Marco and R. Harris, Argumentative meanings and their stylistic configurations in clinical research publications, *Argument & Computation* **6**(3) (2015), 310–346. doi:10.1080/19462166.2016.1171255.
- [17] F. Grasso, Towards computational rhetoric, *Informal Logic* **22**(3) (2002), 225–259. doi:10.22329/il.v22i3.2589.
- [18] R. Harris and C. Di Marco, Constructing a rhetorical figuration ontology, in: *Symposium on Persuasive Technology and Digital Behaviour Intervention, Convention of the Society for the Study of Artificial Intelligence and Simulation of Behaviour (AISB)*, 2009.
- [19] R.A. Harris, Figural logic in Mendel’s experiments on plant hybrids, *Philosophy & Rhetoric* **46**(4) (2013), 570–602. doi:10.5325/philrhet.46.4.0570.
- [20] R.A. Harris, C. Di Marco, A.R. Mehlenbacher, R. Clapperton, I. Choi, I. Li, S. Ruan and C. O’Reilly, A cognitive ontology of rhetorical figures, in: *Cognition and Ontologies*, 18–21 April, 2017.
- [21] D.H. Hromada, 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.
- [22] M. Janier, J. Lawrence and C. Reed, OVA+: An argument analysis interface, in: *Proceedings of the Fifth International Conference on Computational Models of Argument (COMMA 2014)*, Pitlochry, Scotland, S. Parsons, N. Oren, C. Reed and F. Cerutti, eds, IOS Press, 2014, pp. 463–464.
- [23] J. Java, Characterization of prose by rhetorical structure for machine learning classification, PhD thesis, College of Engineering and Computing, Nova Southeastern University, 2015.
- [24] A.R. Kelly, N.A. Abbott, R.A. Harris, C. Di Marco and D.R. Cheriton, Toward an ontology of rhetorical figures, in: *Proceedings of the 28th ACM International Conference on Design of Communication, SIGDOC ’10*, ACM, New York, NY, 2010, pp. 123–130. ISBN 978-1-4503-0403-0. doi:10.1145/1878450.1878471.
- [25] J. Lawrence, F. Bex, C. Reed and M. Snaith, AIFdb: Infrastructure for the argument web, in: *Proceedings of the Fourth International Conference on Computational Models of Argument (COMMA 2012)*, 2012, pp. 515–516.
- [26] J. Lawrence and C. Reed, Combining argument mining techniques, in: *Proceedings of the 2nd Workshop on Argumentation Mining*, Denver, CO, Association for Computational Linguistics, 2015, pp. 127–136.
- [27] B. Liu, Sentiment analysis and subjectivity, *Handbook of natural language processing* **2** (2010), 627–666.
- [28] M. Mladenović and J. Mitrović, Ontology of rhetorical figures for Serbian, I. Habernal and V. Matoušek, eds, Springer, Berlin, 2013, pp. 386–393.
- [29] R.M. Palau and M.-F. Moens, Argumentation mining: The detection, classification and structure of arguments in text, in: *Proceedings of the 12th International Conference on Artificial Intelligence and Law*, ACM, 2009, pp. 98–107.
- [30] V. Pallotta and R. Delmonte, Automatic argumentative analysis for interaction mining, *Argument & Computation* **2**(2–3) (2011), 77–106. doi:10.1080/19462166.2011.608225.
- [31] B. Pang and L. Lee, Opinion mining and sentiment analysis, *Foundations and Trends in Information Retrieval* **2** (2008), 1–135. doi:10.1561/15000000011.
- [32] A. Peldszus and M. Stede, From argument diagrams to argumentation mining in texts: A survey, *International Journal of Cognitive Informatics and Natural Intelligence (IJCINI)* **7**(1) (2013), 1–31. doi:10.4018/jcini.2013010101.
- [33] C. Perelman and L. Olbrechts-Tyteca, *The New Rhetoric: A Treatise on Argumentation*, University of Notre Dame Press, 1971. doi:10.1007/978-94-010-1713-8_8.
- [34] C. Reed, K. Budzynska, R. Duthie, M. Janier, B. Konat, J. Lawrence, A. Pease and M. Snaith, The argument web: An online ecosystem of tools, systems and services for argumentation, *Philosophy & Technology* **30**(2) (2017), 137–160. doi:10.1007/s13347-017-0260-8.

- [35] C. Reed and T. Norman, *Argumentation Machines: New Frontiers in Argument and Computation*, Argumentation Library, Vol. 9, Kluwer, 2003.
- [36] S. Ruan, C. Di Marco and R. Harris, Rhetorical figure annotation with XML (2016).
- [37] A.F. Snoeck Henkemans, What's in a name? The use of the stylistic device metonymy as a strategic manoeuvre in the confrontation and argumentation stages of a discussion, in: *The Uses of Argument. Proceedings of a Conference at McMaster University*, 18–21 May, D.L. Hitchcock, ed., Ontario Society for the Study of Argumentation, 2005, pp. 433–441.
- [38] A.F. Snoeck Henkemans, Manoeuvring strategically with rhetorical questions, in: *Pondering on Problems of Argumentation. Twenty Essays on Theoretical Issues*, F.H. van Eemeren and B. Garssen, eds, Springer, 2009, pp. 15–23.
- [39] A.F. Snoeck Henkemans, The contribution of praeteritio to arguers' confrontational strategic manoeuvres, in: *Examining Argumentation in Context. Fifteen Studies on Strategic Maneuvering*, F.H. van Eemeren, ed., John Benjamins, 2009, pp. 241–255.
- [40] A.F. Snoeck Henkemans, The use of hyperbole in the argumentation stage, in: *Virtues of Argumentation: Proceedings of the 10th International Conference of the Ontario Society for the Study of Argumentation (OSSA)*, 22–26 May, 2013, D. Mohammed and M. Lewiński, eds, Ontario Society for the Study of Argumentation, 2013.
- [41] C. Stab and I. Gurevych, Identifying argumentative discourse structures in persuasive essays, in: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar, Association for Computational Linguistics, 2014, pp. 46–56.
- [42] G. Steen, Deliberate metaphor theory: Basic assumptions, main tenets, urgent issues, *Intercultural Pragmatics* **14**(1) (2017), 1–24. doi:[10.1515/ip-2017-0001](https://doi.org/10.1515/ip-2017-0001).
- [43] C. Strommer, Using rhetorical figures and shallow attributes as a metric of intent in text, PhD thesis, David Cheriton School of Computing, University of Waterloo, 2011.
- [44] C.W. Tindale, *Acts of Arguing: A Rhetorical Model of Argument*, State University of New York Press, 1999.
- [45] C.W. Tindale, *Rhetorical Argumentation: Principles of Theory and Practice*, Sage, 2004.
- [46] F.H. van Eemeren, *Strategic Maneuvering in Argumentative Discourse: Extending the Pragma-Dialectical Theory of Argumentation*, John Benjamins, 2010.
- [47] F.H. van Eemeren, B. Garssen, E.C.W. Krabbe, A.F. Snoeck Henkemans, B. Verheij and J.H.M. Wagemans, *Handbook of Argumentation Theory*, Springer, 2014.
- [48] F.H. van Eemeren and P. Houtlosser, Strategic maneuvering: Maintaining a delicate balance, in: *Dialectic and Rhetoric: The Warp and Woof of Argumentation Analysis*, F.H. van Eemeren and P. Houtlosser, eds, Kluwer Academic, 2002, pp. 131–159.
- [49] F.H. van Eemeren, P. Houtlosser and A.F. Snoeck Henkemans, *Argumentative Indicators in Discourse: A Pragma-Dialectical Study*, Springer, Dordrecht, 2007.
- [50] L. van Poppel, Strategisch manoeuvreren met litotes [Strategic manoeuvring with litotes], in: *De Macht van de Taal: Taalbeheersingsonderzoek in Nederland en Vlaanderen*, D. Van De Mieroop, L. Buysse, R. Coesemans and P. Gillaerts, Uitgeverij ACCO, 2016, pp. 219–232.
- [51] M.P.G. Villalba and P. Saint-Dizier, Some facets of argument mining for opinion analysis, in: *Proceedings of the Fourth International Conference on Computational Models of Argument (COMMA 2012)*, 2012, pp. 23–34.
- [52] M.A. Walker, J.E.F. Tree, P. Anand, R. Abbott and J. King, A corpus for research on deliberation and debate, in: *Proceedings of the 8th Edition of the Language Resources and Evaluation Conference (LREC)*, 2012, pp. 812–817.
- [53] B. Webber, M. Egg and V. Kordoni, Discourse structure and language technology, *Natural Language Engineering* **18**(4) (2011), 437–490. doi:[10.1017/S1351324911000337](https://doi.org/10.1017/S1351324911000337).
- [54] J.W. Wenzel, Perspectives on argument, in: *Readings in Argumentation*, W.L. Benoit, D. Hample and P.J. Benoit, eds, Foris, 1992, pp. 121–143.
- [55] A. Wyner, J. Schneider, K. Atkinson and T. Bench-Capon, Semi-automated argumentative analysis of online product reviews, in: *Proceedings of the Fourth International Conference on Computational Models of Argument (COMMA 2012)*, 2012, pp. 43–50.
- [56] D. Zarefsky, *Political Argumentation in the United States*, John Benjamins, 2014. doi:[10.1075/aic.7](https://doi.org/10.1075/aic.7).