

A Cognitive Ontology of Rhetorical Figures

Randy Allen Harris,¹ Chrysanne Di Marco,² Ashley Rose Mehlenbacher,³ Robert Clapperton,⁴ Insun Choi,⁵ Isabel Li,⁶ Sebastian Ruan,⁷ and Cliff O'Reilly⁸

Thought is metaphoric ... and the metaphors of language derive therefrom.

—I. A. Richards, *The Philosophy of Rhetoric*

Abstract. We report on an ongoing research project that uses OWL and a corpus of figurative instances to explore the cognitive dimensions of rhetorical figures. We argue that figures beyond familiar tropes like metaphor and metonymy are deeply cognitive. We adopt the position that figures are form/function pairings which make them especially promising for Advanced NLP, taking one figure as exemplary, the reverse-repetition scheme, antitabole (e.g., “All for one and one for all”). We illustrate three different methodologies for building ontologies—top down, middle out, and bottom up. And we demonstrate the fruitfulness of cognitive ontological modelling by the way it reveals new categorical arrangements, new conceptual relationships, and important but hitherto unsuspected properties of figures, their multiplicity (figures working together) and their multiplexity (figures having multiple types of instantiations).

1. INTRODUCTION

Rhetoric is the ancient study of persuasion, with particular attention to the effects of expressive style on belief, action, and knowledge. Rhetorical figures are units of style, and their effectiveness is based on their cognitive appeal. The burgeoning work on metaphor, metonymy, and a very small handful of other tropes has established that fact for conceptual figures. But the same is true of schemes, figures based not in semantics (“meaning”) but in phonology, morphology, lexis, and syntax (“form”). Many figures, for instance, like ones we take up in this paper, are based on repetition, and we all know how effective repetition is in aiding memory, shaping belief, and supporting reason. A syllogism is impossible without repetition.

Remarkably, however, rhetorical theories of figuration have rarely been mapped against the results of cognitive science, a metadiscipline that is to the current intellectual environment what evolutionary theory was at the turn of the twentieth century. In an

earlier paper ([16]), we made the case that ontological representation was a natural way to chart the cognitive dimensions of figures, with the promise of considerable Advanced Natural Language Processing (ANLP) payoffs. While this project has connections with other work in the field, most notably suggestions by James Crosswhite, a rhetorician ([6]), and the explicit proposal by Floriana Grasso, a computer scientist ([12], [14]), of a Computational Rhetoric subfield, it is surprisingly novel. Crosswhite briefly suggests the importance of rhetorical figures for argumentation studies, a suggestion Grasso endorses, both of them in a computational context. But neither of them make any suggestions of how figures might be made computationally amenable, what functions they might serve in argument mining or other ANLP domains.

Our project is innovative in the field of rhetoric, where the cognitive dimensions of figures have seen very little attention (though, see [10], [4]), computational methods have not been used to investigate figures, and ontological research is very rare; and our project is unique, to the best of our knowledge, in computer science, artificial intelligence, and NLP, where rhetorical figures are largely unknown and rhetoric is poorly understood.⁹ An excellent recent survey of argument mining, which conspicuously mentions rhetoric many times, nevertheless does not mention rhetorical figures ([20]).

We say “surprisingly novel,” for two main reasons. Firstly, Jeanne Fahnestock’s work shows very clearly how relevant figures are to argumentation ([8], [9], [15], [28]). They epitomize lines of argument. Secondly, many figures are highly amenable to computational modelling and detection. Rhetorical figures are virtually tailor-made for ANLP.

We have been heartened to see a few others adapt our approach or pursue our goals, work we have profited considerably from. Dubremetz and Nivre ([7]) and Hromada ([18]) have had promising success detecting figures. Mladenović and Mitrović

¹ Corresponding author and Director of RhetFig Project; English Language and Literature, University of Waterloo, Waterloo ON Canada N2L 3G1; Email: {raha}@uwaterloo.ca

² Cheriton School of Computer Science, University of Waterloo, Waterloo ON Canada N2L 3G1

³ English Language and Literature, University of Waterloo, Waterloo ON Canada N2L 3G1

⁴ School of Professional Communication, Ryerson University, Toronto ON Canada M5B 2K3

⁵ Mathematics, University of Waterloo, Waterloo ON Canada N2L 3G1

⁶ Biology, University of Waterloo, Waterloo ON Canada N2L 3G1

⁷ Cheriton School of Computer Science, University of Waterloo, Waterloo ON Canada N2L 3G1

⁸ Independent scholar, London.

⁹ Mann and Thompson’s Rhetorical Structure Theory (RST; [21]) is symptomatic of this (understandable) ignorance. While it has made some valuable insights into text linguistics, it is simply incorrectly named, by scholars who appear to know little or nothing about rhetoric. RST has really to do with text *coherence* rather than with *rhetoric* as traditionally understood, as the study of suasive language.

([22]) have built an impressive ontology of rhetorical figures in Serbian.

Our RhetFig Project has continued to develop ([19]), and we report on it further in this communication. We are building a comprehensive cognitive ontology of rhetorical figures in OWL—that is, an ontology organized along known cognitive affinities (like repetition, similarity, and contrast)—for Linguistic Linked Open Data (LLOD). We envisage a wide range of potential applications, including genre detection, sentiment analysis, and argument mining. After working with both a top-down and a middle-out methodology, both having their value, and both discussed below, we have shifted to a data-centric, bottom-up methodology, developing the ontology to accommodate specific instances, rather than the abstract classes.

2. RHETORICAL FIGURES ARE COGNITIVE

The machinery of argumentation specifically, persuasion generally, is guided by the organizational principles and natural affinities of the human mind. No one disputes this for analogy and metaphor, research into which goes back decades in cognitive science, but it is demonstrably true for a wide range of other rhetorical patterns. We know, for instance, that words aid in the recall of certain other words, those which share some of their phonological, morphological, or semantic characteristics (if you hear "key" you more readily call up words like "cue" "keystone" or "door"). *Lexical priming*, the phenomenon is called, and it is often described as a "spreading activation" among related words in neural net models. We know, too, that the brain is particularly attuned to sounds of similar duration and frequency. We know that repetition is critically important for fixing sequences in memory (one repeats a phone number or an address to remember it). All of these brain-function facts are reflected in the schemes and tropes of rhetorical theory.

Consider, for instance, the implications of Kenneth Burke's remarks about antithesis (a trope of juxtaposed oppositional meaning):

[I]magine a passage built about a set of oppositions ("we do this, but they on the other hand do that; we stay here; but they go there; we look up, but they look down," etc.) Once you grasp the trend of the form, it invites participation regardless of the subject matter. Formally, you will find yourself swinging along with the succession of antitheses, even though you may not agree with the proposition that is being presented in this form. ... Thus, you are drawn to the form, not in your capacity as a partisan, but because of some "universal" appeal in it. [3]

Burke shows very clearly here how cognitive inclination (swinging along) works linguistically. In doing so, he opens the question of how it is that minds function when they are exposed to the traditional furniture of rhetoric: broad patterns of reasoning and figuration.

What, we want to know, are the sources of the "universal" appeal that Burke identifies? Or, put in the terms of a highly rhetorical, ubiquitous contemporary praxis, why does that chocolate-bar jingle stay with you endlessly, despite your lack of interest in the chocolate bar, despite even your active dislike of the jingle? The intuitive answer is that there is something in the way your brain operates that allows the jingle to exploit it, colonize it, set up shop. And the intuitive answer is surely right.

Your brain functions with affinities for rhythm, repetition, similarity, and difference (among others). The jingle is rhythmic, repetitious, and contains patterns based on similarity and difference (rhyme, assonance, alliteration, ...); indeed, rhythm itself is a function of repetition, similarity and difference. The jingle, in short, insinuates itself into your consciousness by exploiting the submerged processes from which your consciousness emanates.

Look again at the three antitheses Burke offers:

- we do this, but they on the other hand do that;
- we stay here; but they go there;
- we look up, but they look down;

Antithesis is a trope (conceptual opposition), but it is quickly clear that antithesis is not the only figure at work here (figures very frequently work in concert). Other formal patterns suffuse the passage. I will give their rhetorical names, but what is important is the patterns those names identify. The passage includes parison (parallel syntax), isocolon (parallel prosody), epanaphora (clause-initial lexical repetition), and mesodiplosis (clause-medial lexical repetition), all of which contribute to the "swinging along" Burke's example illustrates, and all of which partake of the same universal character—universal for two reasons.

First, figuration is linguistically inescapable. Figures are omnipresent in language, communicating intentions and desires, coding information and attitudes, propagating belief and knowledge. There is no degree-zero, purely literal language. Language cannot but be figured; it flows in what Edward Sapir figured as "well-worn grooves of expression" ([26]). These grooves can be used or abused, optimized or overblown—which is where formal theories of figuration come in—but there are no other grooves. So, when we look to figuration, we look to primal organizing patterns of language use.

Second, and causally related to the first, figuration reflects the way our brains percolate and process. This reflection is perhaps most evident in the overtly purposive use of figures. The traditional literary purpose, generating aesthetic pleasure, is best known. But mnemonic formulas ("i before e except after c"), proverbs ("a bird in the hand is worth two in the bush"), oral traditions ("rosy-fingered dawn"), children's literature ("I meant what I said and I said what I meant / an elephant's faithful, one hundred percent")—in short, all linguistic configurations serving purposes in which cognitive functions like attention, learnability, and recall are at a premium—take a form that rhetorical theorists in the classical and early-modern periods identified as a figure. It is this insight, coupled with the high degree of overlap between the organizing principles of cognition and the organizing principles of figuration (similarity, contrast, balance, repetition, and the like), that motivates cognitive rhetoric and argues for ontological representation.

Let's turn to an extended example, the seemingly esoteric scheme of reverse lexical repetition, antimetabole.

3. ANTIMETABOLE

This figure (from Greek *anti* "in opposite direction" + *metabole* "turning about") is a poorly researched and largely unknown but remarkably widespread rhetorical device characterized by symmetrical lexical inversion. Its most famous modern exemplum may be John F. Kennedy's summarizing antimetabole in his inaugural address:

- (1) Ask not what your country can do for you. Ask what you can do for your country

Science

- ## Political Science

- ## Algebra

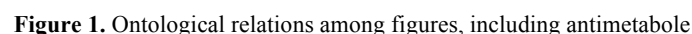
- ## Logic

- ## Advertising

- ## Ordinary Language

- Antimetaboles also show up in all the places one would normally expect to find a ‘fancy’ figure like this—fiction, poetry, philosophy, oratory, popular music (hip hop is full of antimetaboles). It is present in the ancient texts of China, India, Egypt, and Mesopotamia, and in the daily twitter feeds and status updates of the 21st century. They course through social media; Mark Zuckerberg’s recent claim about Facebook, for instance:

- Antimetaboles are aesthetically pleasing, memorable, and culturally pervasive because they stack up three cognitive affinities. Other rhetorical figures leverage these affinities (and others) in a range of similar, different, and overlapping ways. Figure 1 illustrates some of these ways, which our RhetFig Ontology systematizes.



4. FORM/FUNCTION PAIRINGS

If our work only went this far, one might fairly criticize it with a familiar line of attack against rhetoric, that it is ‘merely stylistic,’ that it tells us only about the peripheries of language, nothing about its conceptual depths. But most humanists, and all rhetoricians, realize that there is nothing *mere* about style, that style in fact offers different routes to those depths. Antimetabole charts three specific routes to a highly constrained, inter-related set of conceptual alignments ([10], [11], [25], [28]). Newton’s and Pasteur’s examples above are antimetaboles of reciprocity; the law of commutation, irrelevance-of-order; the “winners” and “place” ordinary-language examples, comprehensiveness.

What is compelling about this tight form/function coupling from a computational perspective is that once you find the form, the function is highly predictable. And the form of antimetabole, and of all schemes (tropes are a trickier matter), is computationally very tractable, in the sense that they can be modelled formally. Here is how Harris and Di Marco [16] render it as a Regular Expression (using the Waterloo Figure Representation formalism):

$$W^x \dots W^y \dots W^y \dots W^x$$

In our Newton instance, for instance, W^x = finger, W^y = stone; the triple dots loosely express intervening elements (possibly null).

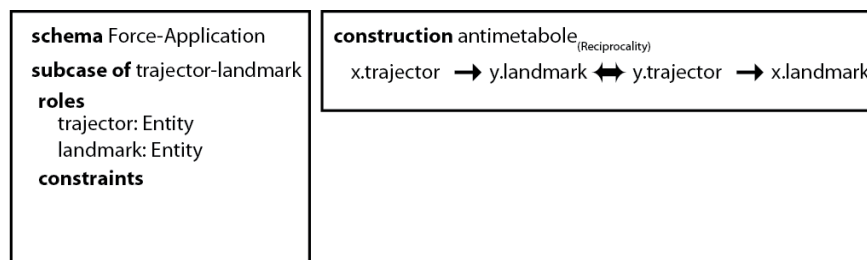


Figure 2. CxG representation of the reciprocal function of antimetabole

Figure 2 is only partially applicable, in that it captures antimetabolic instances like the Newton example (2), but fails with others, like the Pasteur instance (3) and the algebraic instance (7), in neither of which are their trajectories. But it provides a template for the sorts of CxG expressions needed for figural representation.

The abstract sequence of the antimetabole, represented in the Waterloo Figure Representation Notation above ([16], [25]), is precisely the kind of pattern a computational text analyser can find very easily; the abstract conceptual structure represented in Figure 2, and the combination of computational tractability with narrow functional range means that antimetaboles can be used to digitally diagnose major discursive themes and epitomize arguments. Here is where our work falls into the closest alliance with schools of Computational Argumentation.

Despite the hundreds of rhetorical figures that have been catalogued over two millennia, they fall into a relatively few, partially overlapping classes. While we have not worked out an exhaustive set of classes and relations, we are especially intrigued by the way in which the natural organizing principles of figures

This sort of representation is flexible and resilient—extensible to other constituents (phonemes, syllables, and phrases, as well as to words) and to all of the basic figurative operations (omission, addition, and substitution, in addition to iteration and permutation, as above), and it has proven value in computational research.

This representation is amenable to coding, and has been rendered in Perl, for instance ([17]). It has now been used for such purposes as text summarization ([1], [2]) and figure detection ([7], **Error! Reference source not found.**, [11], [17], [19], [24] [27]). But it was never meant as anything more than a sketch of possibilities; our project is regularizing the formalism, axiomatizing it with definitions of necessary concepts like identity, proximity, and sequence; and enforcing an overall rigour on the computational representations. Regular Expressions, too, only do half the job, by representing the pattern of figures; we need a parallel formalism for the functions they serve.

We are exploring how the necessary formalism may be found in Construction Grammar (CxG), a new morphosyntactic architecture developed over the last decade in close conjunction with Frame Semantics ([18]). In the Newton instance above (2), the function is reciprocity; Newton utilizes the antimetabole to express the utter balance of reciprocal forces. To the precise extent that the finger presses on the stone, the stone presses on the finger. Figure 2 captures this conceptual structure in CxG terms.

manifest well known cognitive affinities, like comparison, contrast, and symmetry, and by the interplay of well-known linguistic operations in the patterning of figures, like addition, deletion, and permutation. Our work applies these insights to the construction of a cognitive ontology of rhetorical figures.

5. ONTOLOGY DEVELOPMENT METHODOLOGIES

Rhetorical figures are prime candidates for a cognitive ontology. They are both richly, demonstrably cognitive, and richly, demonstrably valuable for NLP tasks like argument mining. The rhetorical tradition also suggests they hold considerable promise for a range of other automated text processing tasks, such as authorship attribution, genre detection, sentiment analysis, credibility assessment, pathology diagnosis and monitoring, voice interaction systems ... virtually any activity that implicates pragmatics or style. Ontological representation seemed so obvious to us, so natural, so inevitable.

We started with a top-down approach. We used the categories: Cognitive Affinities (such as CONTRAST, SIMILARITY, SEQUENCE, REPETITION, and POSITION); Linguistic Domains (such as PHONEMES, SYLLABLES, MORPHEMES, WORDS, and PHRASES); and Figure Kind (such as TROPES and SCHEMES, as well as two somewhat novel categories, CHROMA and MOVES—see [4], [11]). Antimetabole, for instance, is a scheme of words utilizing affinities for repetition, sequence, and contrast, or <SCHEME; WORD; REPETITION, SEQUENCE, CONTRAST>.¹⁰

We followed this approach to considerable profit. We quickly realized that we would do better with multiple linked ontologies than trying to incorporate everything into one ontology, so we built a Cognitive Affinities ontology and an abbreviated Linguistic Domains ontology (we are hopeful that we will find existing ontologies that we can link to our RhetFig Ontology, and are aware in particular of multiple existing linguistic ontologies, but we have not had the chance to pursue this very far; we are still in the sandbox phase). Since OWL uses subsumption as its guiding organizational principle, we capitalized on the features of inheritance, which led us to reanalyse the conventional taxonomy of figures. So, for instance, there is figure known simply as *ploce* in the rhetorical tradition, a figure of unconstrained lexical repetition (that is, <SCHEME; WORD; REPETITION>), as in these instances:

- (16) The best surprise is no surprise at all. (Holiday Inn marketing slogan)
- (17) Villain, villain, smiling, damned villain! (Shakespeare, *Hamlet*)

But antimetabole, as well as a host of other lexical repetition figures, all of which are constrained in various ways, also satisfy the description <SCHEME; WORD; REPETITION>, except that they are constrained by other affinities, such as sequence or position. Antimetabole, for instance, is constrained by relative sequence (the repetitions must be in reverse order). Epanaphora is constrained by position. We developed a hierarchy corresponding to this realization, partially illustrated in Figure 3.

This new hierarchy provides us with a different understanding of figures in terms of their mutual relations based on cognitively principled categorization.

But the welter of rhetorical figures that have been catalogued over the millennia since their discovery soon made the top-down approach rather unwieldy and we opted for a middle-out development methodology, focussing only on figures of repetition. These include lexical repetitions (such as epanaphora and antimetabole), but also repetitions of phonemes (as in alliteration, repeating word-initial consonants), syllables (as in rhyme, repeating word-final syllables), morphemes (as in homoioputon, repeating affixes), and so on. It also includes tropes such as synonymia (repeating concepts in different words), but we have restricted ourselves to schemes.

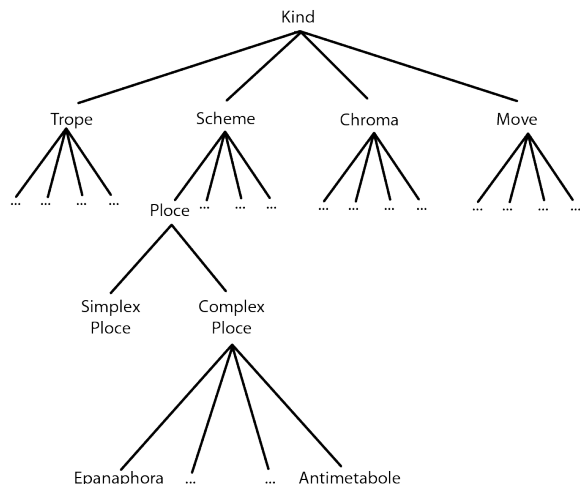


Figure 3. A new understanding of rhetorical figures of lexical repetition, based on the inheritance characteristics of ontologies

Again the approach was profitable, but again we ran into complications. On the profit side of the ledger, we found efficient ways to characterize the similarities and differences of figures. For instance, polyptoton is a figure in which the lemma is repeated but with different morphology, while in homoioputon morphemes are repeated but with different lemmata, and (as you recall) ploce is a figure in which lexemes are repeated (i.e., both the lemma and the morpheme are repeated). Our use of linguistic domains in our categorization also allowed us to make further distinctions that had not been observed in the rhetorical tradition. Consider the following instances (all from James Baldwin's *Go Tell it on the Mountain*):

- (18) It was later to become his proud testimony that he hated his sins—even as he ran towards sin. (Ploce)
- (19) He wanted to be with these boys in the street, heedless and thoughtless. (Homoioputon)
- (20) You know honey, the Word tell us to hate the sin but not the sinner. (Polyptoton)

The formal structure of instances (18) and (20) superficially look the same. Both have the repeated lemmata with different morphemes. But (18) we categorize as ploce, because the morphology is inflectional, with negligible syntactic or semantic effects (*sins*, *sin*), while (20) is polyptoton because the morphology is derivational, with marked syntactic and/or semantic effects (*sin*, *sinner*). By categorizing these figures in the directions that OWL led us, we end up with a very efficient system that reminds one of the Principles and Parameters programme of Chomskyan linguistics:

Ploce <SCHEME; WORD-LEXEME; REPETITION>

Homoioputon <SCHEME; WORD-LEMMA_x, MORPHEME_y; REPETITION_y, CONTRAST_x>

Polyptoton <SCHEME; WORD-LEMMA_x, MORPHEME_y; REPETITION_x, CONTRAST_x>

¹⁰ The formalism is as simple as it looks. The expressions are delimited with angled brackets. The main classes of information (figure type(s), linguistic domain(s), and cognitive affinity(/ies)), all mandatory, are separated by semi-colons. The specific

type(s), domain(s), and affinity(/ies), when there are more than one, are separated by commas.

(Subscripts are markers of affiliation: for homoiptoton, MORPHEME goes with REPETITION; WORD-LEMMA goes with CONTRAST; vice versa for polyptoton.) Homoiptoton is a figure in which lemmata contrast, but a morpheme repeats. Polyptoton is a figure in which a lemma repeats, but morphemes contrast.

Ploce is, in these terms, a figure of simple lexical repetition, in which morphology goes unmentioned because it is inflectional. Homoiptoton and polyptoton both share the affinities of repetition and contrast, and both share the linguistic elements of word-lemma and morpheme. They differ only in terms of affinity-element assignments. The similarity between this treatment and the “switchbox” of Principles and Parameters ([5]) is suggestive on this level. Set the switches one way, you get polyptoton, set them another, homoiptoton. When the cognitive charting of figures advances to become fully cross-linguistic—as any cognitive attention to data must—we feel that there may be value in a framework of this type (though, of course, we can’t know until the data is in *what* approach will be of value, and there is not even an initiative to collect such data). Free word order languages and fixed word order languages, for instance, will have different propensities toward figures implicating lexical sequence and figures implicating morphology.

The middle-out method soon hit a wall of diminishing returns as well—not, in this case, because of the great variety of figures but because of the limitations of OWL, and of description logics more generally, which present difficulties with capturing important notions like sequencing and indexing (since we need to represent relative and absolute order). We are familiar with some work that addresses some of these limitations through a tight and customized linking of ontologies. O’Reilly’s work in particular ([23], [24]) modelled concepts and relationships such as *hasFirstWord*, *hasLastWord*, *hasNextWord*, and so on, in one ontology. Rhetorical Figures like epanaphora and epistrophe (clause-initial and clause-final lexical repetitions, respectively; they are types of complex ploce in our terms) went into another ontology, with Semantic Web Rule Language (SWRL) logic rules defining figures. While we feel these methods are productive, and we encourage O’Reilly to continue this work, others to emulate it, we are hopeful of more direct ways to express cognitive affinities like sequence and position. More importantly, we increasingly found that our instances were not falling into neat categories, and we changed to a data-centric approach, building the ontology from the bottom up.

We began by somewhat arbitrarily choosing a single instance of antimetabole, an exchange from Hemingway’s *For whom the Bell Tolls*:

(21) “Coward,” Pablo said bitterly. “You treat a man as a coward because he has a tactical sense. Because he can see the results of an idiocy in advance. It is not cowardly to know what is foolish.”

“Neither is it foolish to know what is cowardly,” said Anselmo, unable to resist making the phrase.

This seemed like an unassuming little antimetabole in which one character uses the reverse order of the figure in a kind of refutation of another character’s utterance. We like it especially for the phrase that Hemingway uses to summarize Anselmo’s motivation, “unable to resist making the phrase.” Just as the first-one-order-then-the-other-order is the most natural and iconic way to represent the commutative principle in math and logic—i.e., (7) and (8), which exemplify the irrelevance-of-order antimetabolic function—a reversal of someone’s utterance seems the most natural

and iconic way to represent its repudiation. Hemingway, one of the most resolutely anti-rhetorical authors (an anti-style stylist, if you will), found the venerable antimetabole irresistible here to express the antipathy between his two characters, ascribing the compulsion to one of those characters.

Approaching categorization from the bottom had two unexpected benefits, discovering the multiplicity of figures and the multiplexity of figures.

The Hemingway instance (21), we soon realized, contains a multiplicity of figures, so that it was an instance not just of antimetabole, but of other figures as well, and if we were going to use our data to learn about the functional role of figures in language, we would have to be more thorough in our classifications. Fortunately, OWL intrinsically allows assigning one instance to multiple classes. Some of these other figures seemed somewhat incidental to the refutational function of the antimetabole—polyptoton, for instance (*coward*, *cowardly*), and epanaphora (*because he ...*, *because he ...*). Others are more integral—in particular, mesodiplosis (clause-medial lexical repetition; here *... to know what is ...*, *... to know what is ...*). Indeed, as we looked through our data, we came to realize that most of the most clearly functional antimetabolic instances were not just reverse lexical repetitions but reverse lexical repetitions around another stable lexical repetition. That is, they are more fully represented in the following terms (where the W^x and W^z constituents represent the antimetabole, the W^y constituents represent the mesodiplosis):

$$W^x \dots W^y \dots W^z \dots W^z \dots W^y \dots W^x$$

If you look back at the antimetabolic instances in this paper, armed with this insight, the structure is impossible to miss. For (2) the mesodiplosis constituent is *press*, for (4) it is *moves in opposite circles*, for (5) it is *that determines their*, for (6) it is *fear*, for (7) it is the addition and division signs, and so on.

Moreover, some figures have particularly strong bonds with one another. Antimetabole manifests a syntactic (sequential, formal) opposition. Antithesis manifests semantic (conceptual, ideational) opposition. So, the two frequently co-occur, reinforcing each other. A huge proportion of antimetaboles are also antitheses, like the famous Kennedy aphorism (1), as well as the Marx (5) and Zuckerberg (15) instances. More compellingly, the combination of figures usually doubles down on the functional constraints. All of the antimetabole-antithesis expressions we have found combine the comprehensiveness function of the first figure with the rejection function of the second figure. Adding antithesis to antimetabole, that is, always produces an utter rejection of one proposition, concomitant with its utter replacement by the other, either assertively (Kennedy, Marx) or implicationally (Zuckerberg).

Rhetorical functions, in short, are very often the result of specific compounding effects. Figures work in multiplicities.

The data, we also discovered through the careful attention that ontological considerations forced on us, is multiplex. In particular, we found as we worked through the instances, especially in our focus on linguistic domains, that what we thought was one figure, antimetabole (and what the rhetorical tradition often treated as one figure, sometimes two) was in fact at least seven distinct patterns, which we identify as follows (some of the instances are repeated from above, under this new categorization; some of them are original):

Antimetathesis

(22) She sells sea shells by the seashore.

Antimetabole

- (23) It's Trump calling real journalism "fake news" and holding up "fake news" as real journalism. (tweeted by (@ExportedFromMI))

Morphological Chiasmus

- (24) Friendly Americans with America friends.

Syntactic Chiasmus

- (25) Despised, if ugly; if she's fair, betrayed. (Leapor. "Essay on Woman")

Antimetalepsis

- (26) [T]he new metaphor — the brain as a calculating machine — is ... fraught with the danger of anthropomorphizing the machine and mechanizing the man.

Pseudo-Antimetabole

- (27) StarKist doesn't want a tuna with good taste. StarKist wants a tuna that tastes good.

Implied Antimetabole

- (28) Britannia waives the rules [evoking the familiar inverse phrase, "Britannia rules the waves."].

We call this grouping of figures *The Chiastic Suite*, after one of the related terms for *antimetabole* from the rhetorical tradition, *chiasmus*, itself named after the Greek letter chi (χ , the repeating elements are said to 'cross' each other).

There's more going on here than just a new data story. Identifying this cluster of figures from the instances (each and every one of these configurations has been held to exemplify antimetabole, or one of its synonyms, in the tradition) allows us to 'purify' antimetabole for analytic terms. This is most obvious with pseudo-antimetabole. It is a *pseudo*-antimetabole because the two occurrences of *taste* in our example evoke different meanings (therefore, different words). The first word is *taste* in cultured appreciation of the arts sense, the second is *taste* as in the gustatory sense. There is, in other words, no repetition of *taste*; rather, there is a pun (formally, a paronomasia), a 'repetition' of the semiotic vehicle without the semiotic interpretant, along with a repetition of *good*. If one builds an automatic detector for antimetabole in order to do argument mining and comes across an instance like (27), the algorithm needs to be able to filter it out.

On the other hand, antimetalepsis pattern conceptually very much like antimetaboles, which suggests one should widen the search pattern to include instances of 'crossed' synonymia, for argument mining. Instance (26), for example, serves a comprehensiveness function; the claim of the sentence is that the mind-as-machine metaphor that was so prevalent in early cognitive science is completely dangerous, because it misconceives minds and it misconceives machines. We don't have the space to map out all of the chiastic suite—and, to be frank, we haven't fully done so; the interactions are tricky, calling for considerably more investigation—but they demonstrate the productivity of cognitive ontologies for research into linguistic patterns.

6. CONCLUSION

Cognitive ontologies are highly productive, and vastly under-utilized, research tools for ANLP, especially in matters of style, pragmatics, and argumentation. Our research project demonstrates their virtues in multiple ways. Figure ontologies provide mechanisms for plumbing texts. And they reveal unsuspected features of linguistic configurations. But this research is in very early days. Our future plans for the project include (i) exploring other ontological representations, (ii) developing our annotation

scheme, (iii) linking other ontologies, (iv) applying our bottom-up methodology in other figurative domains, (v) expanding our research to include other categories of rhetorical figures, beyond schemes, and (vi) expanding and refining our detection engine.

- i. OWL has been particularly restrictive in its limitation to the *IsA* relation, which is insufficient both for the relations among many figures (including, for instance, various kinds of meronymy) and for the kind of information we foresee necessary for applications. So, we are looking at ways to incorporate other relations in OWL, and to augment OWL's expressiveness in other ways. But we are not committed to OWL. So we are also looking for other representational languages.
- ii. The annotation scheme has multiple challenges. We only annotate the *form* of the figures at this stage, but annotating their functions is equally important, perhaps more so for activities like argument mining and text summarization. The tendency for figures to cluster in the same stretches of also makes for very messy annotation, so we are exploring stand-off markup. In both instances, we will ensure standardization as best we can, to increase the usefulness of our work, by following Apache UIMA (Unstructured Information Management) guidelines.
- iii. We have been unsuccessful at finding other ontologies we can link to ours, and plan to expand our efforts at finding them, especially linguistic and cognitive ontologies.
- iv. The bottom-up strategy has its limitations, but has been the most productive methodology we have adopted. We began with a single instance, but it was not arbitrarily chosen. It was an antimetabole, which we knew had various interesting implications and it has led to the discovery of the chiastic suite. We plan to begin another 'mini-ontology,' bottom-up with instances of the figures *gradatio*, because of its relation to the scheme *anadiplosis* (*anadiplosis* is lexical repetition at the end of one phrase and the beginning of the following phrase; *gradatio* is two or more successive *anadiploses*, so *anadiplosis* is a *PART-OF gradatio*), such tropes as *incrementum* and *decrementum* (successive words with semantic inclines or declines), and the compound figure (part scheme, part trope), *climax* (*gradatio* + *incrementum*). We apologize for the necessarily enigmatic description of these plans. Adequately defining and exemplifying the figures here would require too much space. But we invite interested readers to consult Fahnestock's ([8]) chapter on *incrementum* and *gradatio* for more detail about some of these figures, many examples, and an account of how they function in argumentation.
- v. Work on a *gradatio* mini-ontology, as we've just suggested, will involve the inclusion of tropes and compound figures, but there is room for expansion into other tropes, into *chroma* as well, and other figures as well. There has been considerable work on Argument Schemes in argument mining, for instance, a nice overview of which can be found in a recent paper by Argument Scheme guru, Douglas Walton, with Fabrizio Macagno ([29]). There are many figures, chiefly the ones we call *Moves*, that are strikingly similar to Argument Schemes (for instance, *prolepsis*, which is the projection and refutation of a counterargument, and *paralipsis*, the insinuation of premises

while decrying their inclusion in an argument), and we envision a convergence of our research with Walton's line, starting with the incorporation of moves into our ontology.

- vi. Jakub Gawryjolek developed JANTOR (Java ANnotation Tool Of Rhetoric) in 2009 ([11], [12]) at the outset of our project, an ambitious and accomplished but preliminary tool for the detection and annotation of rhetorical figures in HTML files. We plan to update and enhance JANTOR and make it available for other research groups

REFERENCES

- [1] M. Alliheedi and C. Di Marco. Rhetorical Figuration as a Metric in Text Summarization. *Advances in Artificial Intelligence, Lecture Notes in Computer Science* 8436: 13-22 (2014).
- [2] M. Alliheedi. *Multi-document Summarization System using Rhetorical Information*. M.S. thesis David Cheriton School of Computing, University of Waterloo, Waterloo ON Canada. (2012).
- [3] K. Burke. *Rhetoric of Motives*. University of California Press, Berkeley, CA (1969).
- [4] L. Chien and R. A. Harris. Scheme Trope Chroma Chengyu: Figuration in Chinese Four-Character Idioms. *Cognitive Semiotics* 10(6):155-178. (2011).
- [5] A. N. Chomsky. *New Horizons in the Study of Language and Mind*. Cambridge University Press, Cambridge, UK. (2000).
- [6] J. Crosswhite. Rhetoric and computation, Symposium on Argument and Computation. Bonskeid House, Perthshire, Scotland. June 27, 2000.
- [7] M. Dubremetz and J. Nivre. Rhetorical Figure Detection: The Case of Chiasmus. *Proc. of NAACL-HLT Fourth Workshop on Computational Linguistics for Literature, Denver, CO, June 4*. (2015).
- [8] J. Fahnestock. *Rhetorical Figures in Science*. Oxford University Press, New York (1999).
- [9] J. Fahnestock. Figures of Argument. *Informal Logic* 24.2:115-135. (2004).
- [10] J. Fahnestock. Rhetoric in the Age of Cognitive Science. In: *The Viability of Rhetoric*. Richard Graff (Ed). State University of New York Press, New York, 159-180 (2005).
- [11] J. Gawryjolek. *Automated Annotation and Visualization of Rhetorical Figures*. MSc thesis. David Cheriton School of Computing, University of Waterloo, Waterloo ON Canada. (2009).
- [12] J. Gawryjolek, C. Di Marco, and R. A. Harris. An Annotation Tool for Automatically Detecting Rhetorical Figures. *CMNA (Computational Models of Natural Argument)*, 13 July 09, Pasadena, CA. (2009).
- [13] F. Grasso. Towards Computational Rhetoric. *Informal Logic* 22: 195-229. (2002).
- [14] F. Grasso. Towards a Framework for Rhetorical Argumentation. *EDIALOG 2002 - Proceedings of the 6th workshop on the semantics and pragmatics of dialogue*, Edinburgh, UK, 4-6 September 2002, p. 53-60.
- [15] R. A. Harris. Figural Logic in Mendel's Experiments on Plant Hybrids. *Philosophy and Rhetoric* 46.4: 570-602. (2013).
- [16] R. A. Harris and C. Di Marco. Constructing a Rhetorical Figuration Ontology. *AISB (Artificial Intelligence and Simulation of Behaviour)*, 8 April 09, Edinburgh, Scotland. (2009).
- [17] D. H. Hromada. Initial Experiments with Multilingual Extraction of Rhetoric Figures by Means of PERL-Compatible Regular Expressions. *Proc. of the Second Student Research Workshop associated with RANLP*, Hissar, Bulgaria. (2011).
- [18] T. Hoffmann and G. Trousdale (Eds). *The Oxford Handbook of Construction Grammar*. Oxford University Press, New York (2013).
- [19] A. R. Kelly [now Mehlenbacher], N. Abbott, R. A. Harris and C. Di Marco. Toward an Ontology of Rhetorical Figures. *Proc. of the 28th ACM International Conference on Design of Communication*, 123-130. (2010).
- [20] M. Lippi and P. Torroni. Argumentation Mining: State of the Art and Emerging Trends. *ACM Transactions on Internet Technology* 16.2:1-25
- [21] W. C. Mann and S. A. Thompson. Rhetorical Structure Theory: Toward a functional theory of text organization. *Text* 8.3:243-281. (1988).
- [22] M. Mladenović and J. Mitrović. Ontology of Rhetorical Figures for Serbian. *Text, Speech, and Dialogue*. Volume 8082 of the series *Lecture Notes in Computer Science*, 386-393. (2013).
- [23] C. O'Reilly. *Lassoing Rhetoric with OWL and SWRL*. MSc thesis, Computer Science, University of Westminster, London. (2010).
- [24] C. O'Reilly and S. Paurobally. Lassoing Rhetoric with OWL and SWRL. Unpublished. Available http://www.academia.edu/2095469/Lassoing_Rhetoric_with_OWL_and_SWRL (2010).
- [25] S. Ruan, C. Di Marco, and R. A. Harris. *CMNA XVI - Computational Models of Natural Argument*. New York, NY, 9 July. (2016).
- [26] E. Sapir. *Language: An Introduction to the Study of Speech*. Harcourt, Brace, New York (1921).
- [27] C. Strommer. *Using Rhetorical Figures and Shallow Attributes as a Metric of Intent in Text*. PhD dissertation, David Cheriton School of Computing University of Waterloo, Waterloo ON Canada. (2011).
- [28] C. W. Tindale. *Acts of Arguing: A Rhetorical Model of Argument*, State University of New York Press, Albany, NY (2000).
- [29] D. Walton and F. Macagnob. A Classification System for Argumentation Schemes. *Argument and Computation* 6.3:219-245.