Toward an Ontology of Rhetorical Figures

Ashley R. Kelly
University of Waterloo
Department of English
Language & Literature
arkelly@uwaterloo.ca

Nike A. Abbott
University of Waterloo
Department of English
Language & Literature
nabott@uwaterloo.ca

Randy Allen Harris
University of Waterloo
Department of English
Language & Literature
raha@uwaterloo.ca

Chrysanne DiMarco*
University of Waterloo
Department of English
David R. Cheriton
School of Computer Science
cdimardo@uwaterloo.ca

ABSTRACT

Our paper describes the Rhetorical Figure Ontology Project, a multidisciplinary research project that is presently working towards the development of a comprehensive database of rhetorical figures, an associated wiki, and, ultimately, an ontology of rhetorical figures. The database and wiki project provide the dataset and space for the conceptual development, respectively, to create an ontology. We define an ontology as a formalized taxonomy or system of classification of concepts and associated descriptions of said concepts. Here we provide an overview of the present state of the project and a discussion of the development of ontological descriptions of rhetorical figures. This work is a joint venture between Dr. Randy Allen Harris (English) and Dr. Chrysanne DiMarco (Computer Science, and English) at the University of Waterloo, Canada.

Categories and Subject Descriptors

I.2.7 [Computing Methodologies]: Artificial Intelligence—Natural Language Processing

General Terms

Theory, Design

Keywords

ontology, rhetoric, rhetorical figures

*Dr. Chrysanne DiMarco is the Director of the Inkpot Natural Language Research Group.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires specific permission and/or a fee.

SIGDOC 2010, September 27–29, 2010, S.Carlos, SP, Brazil.
Copyright 2010 ACM 978-1-4503-0403-0…$5.00.

1. INTRODUCTION

Situated in the David R. Cheriton School of Computer Science at the University of Waterloo, the Inkpot Natural Language Research Group aims to advance research in the area of natural language processing and computational discourse analysis. Inkpot has taken an innovative approach to the study of human language in computer science by incorporating the work of language scholars, from the field of rhetoric (see: Harris & DiMarco 2009; Gawryjolek 2009) [11] [7], and computational methods to analyze and manipulate human language. The Rhetorical Figure Ontology Project, a subset of the Inkpot team, seeks to combine linguistic and rhetorical theory with machine learning methods to create an ontology, a formal representation of concepts, for use in computational processing of real-world communication. The research team is comprised of experts in rhetoric, linguistics, and natural language processing. This paper describes the development and implementation of a database of rhetorical figures and the ongoing work to create an ontology of said figures.

The first stage in developing the ontology is the construction of a comprehensive database of rhetorical figures. The database is its current incarnation is a simple column-based application that allows for data to be collected, organized, and modified. Data being collected includes: definitions, examples, and bibliographic information about rhetorical figures. In addition to this data collected from historical sources, the figures are being categorized on the basis of: figure type-of, linguistic domain, and kind-of relationship—these terms will be explained below, but essentially all describe individualizing linguistic and rhetorical characteristics of the figures. Presently the database contains a list of almost one thousand entries; these entries are divided into two types of figures: primary figures and synonyms. At last count about three hundred and fifty figures would be classified as primary figures with the remaining listed as synonyms.

Many of the figures present complex and confounding historical trajectories. In order to provide a space to explore the nuances of each figure in great detail there is an associated wiki with the database. This space allows for individual figures to be treated, but, perhaps more importantly, allows for a space to discuss the details of figure-relationships and
the individualizing characteristics as they are treated in the database. In this way, we are able to create increasingly refined classes of figures, their attributes, and relationships.

We argue that the study of rhetoric offers profound insights into the understanding of human language both in its structure and use. In particular, rhetoric offers an extensive treatment of complex forms of expression, known as rhetorical figures or figures of speech. While the encyclopedic information being collected and assimilated in the database is significant, it is the relationship of the figures to one another that is of central importance to the discussion in this paper. The data being aggregated for each figure provides critical information that helps elucidate and explicate the figure as a meaning-making, purposive device.

We also argue that the relationships among figures provide insight into the cognitive affinities of the figures themselves. This work has a multitude of applications, but our specific focus here is the importance of such information in understanding the role rhetorical figures play as not only meaning-making, purposive devices, but as portals to the inner-workings of the human mind. The cognitive role of figures is based in the work of such scholars as literary theorist Mark Turner, linguist George Lakoff, philosopher Mark Johnson, psychologist Raymond Gibbs, Jr., and rhetorician Jeanne Fahnestock. Though rhetorical figures have been defined, differentiated, and categorized for millennia from the Rhetorica ad Herennium to Henry Peacham’s Garden of Eloquence to the contemporary work of Gideon O. Burton on Silva Rhetoricae, many figures escape clear classification.

While our project does not claim, nor does it aim, to provide definitive categorization for rhetorical figures, it does work towards a systematized approach to the treatment of figures in computational systems. While such efforts will, at present, remain imperfect, our work will describe methods of categorization of rhetorical figures, difficulties with the categorization, and some of the methods that we might use to work around such difficulties. To this end we will treat the development, including the theoretical underpinnings for the categorization, and population of the database.

2. RELEVANT LITERATURE

Natural Language Processing (NLP) is an area of study, in the discipline of Computer Science, that treats the manipulation and generation of human-readable language. While there have been some efforts to treat “rhetoric” in the field of NLP, as Harris & DiMarco (2009) argue, these treatments fail to understand the primary scholarship and rhetoric to any significant degree, rarely going beyond a generalized notion of style derived from linguistics. Often the degree to which language scholarship is consulted is limited to the study of linguistics [11].

While linguists and computational linguists have worked together on developing models for software programs, rhetoric can more fully treat the complexity of human speech. Many linguistic schemes follow tractable characteristics of words and word groupings. Rhetoric is more difficult to schematize because it seeks to uncover the exigence and purpose of a figure or rhetorical device as well as the linguistic strategies a figure employs for effectiveness. Furthermore, many characteristics of figures (especially tropes) are conceptual and therefore more difficult to formally specify. However, encouraging progress has been made towards computational methods for detecting and manipulating rhetorical figures in texts [7] [8] [22] and [3]. This progress made by the Inkpot team is generally treated below in our discussion of the ontology’s classification of schemes, tropes and chroma, but is more fully elaborated on in the previously cited works.

Cognitive studies have taken up the study of rhetoric as a way to research cognitive processes in communication. Lakoff and Johnson treat figurative language as a basic component of everyday speech in their book Metaphors We Live By [13], while Raymond Gibbs posits that “human cognition is fundamentally shaped by various poetic or figurative processes” in his book, The Poetics of Mind [10].

What these studies show is that rhetorical figuration is used in everyday speech; what can be inferred from that is that as natural language generation progress so too must our understanding of rhetorical figures as reflections of cognitive processes. People use figurative devices to understand abstract concepts and to communicate about them to others. Instead of figuration requiring special cognitive processes, rhetorical figures reflect the way humans cognitively respond to the world around them.

3. OUR APPROACH

3.1 Method

Editors contributing to the database of a rhetorical terms include both graduate and undergraduate students, research associates, and primary investigators. Combined, this team’s disciplinary expertise includes: linguistics, computational linguistics, rhetoric, artificial intelligence, computer science, and even includes students from the faculty of science. In order to ensure that the team is adequately trained and consistent in their work we have made efforts to provide the necessary skills and understanding of our project and systems of classification to ensure a common framework from which everyone can work. All members of the team are expected to, and indeed have obtained, a general understanding of rhetoric and the role of figures in theory.

Over the year that the project has been conducted we have held many sessions, bi-weekly during school terms, where the editors come together in an attempt to normalize their practices. In some cases our meetings brought forth problematic methods of classification and organization in our work, and in others helped to train those new to the project in basic linguistic and rhetorical theory. Additionally, help documentation with definitions and examples is provided to editors in a project wiki. This documentation is constantly being revised along with the database software and conceptual classification of the database.

Despite our efforts we still have some significant inconsistencies in the editors work. This is, in part, due to the ongoing development of the database and its rules for categorization, but it is also due to the difficult nature of the material that the editors are working from and with. For this reason we have employed a review system wherein one who is better versed in rhetorical theory, typically a graduate student from the Department of English Language & Literature, reviews the work of the editors and either accepts or rejects their changes.

3.2 Primary Texts

Part of the research into primary source texts for the database is to explore the various taxonomies that have been
created by rhetoricians, logicians, grammarians, and other scholars over the past two millennia. A goal of the Rhetorical Figure Ontology Project is to provide a comprehensive exploration of the possible, and most useful, classifications of rhetorical figures. Accordingly, one of the first considerations for the project was to choose a primary name for a figure—as many figures have multiple synonyms—in order to reach an agreement on identification. The decision was made to follow the lead of the primary source for the database's first incarnation, which is the Silva Rhetoricæ website spearheaded by Gideon O. Burton of Brigham Young University. The primary names are the Greek-based English terms and the synonyms can include the Latin, alternate spellings, or an English turn-of-phrase.

Building on Burton's major contribution, many important rhetorical scholars who created seminal and influential figuration taxonomies were consulted. Many of the primary sources from which we draw the definitions and examples for each figure are rhetorical manuals from Greek and Roman antiquity, the middle ages, early modern and modern periods. Most of the texts are English language or English translations, usually from Latin. The earliest text is the Rhetorica ad Herennium [2], in Harry Caplan's translation, and is still considered the primary source for rhetorical devices. Geoffrey of Vinsauf later combined two major classical sources, the Rhetorica ad Herennium and Horace's Ars Poetria, in his Poetria Nova [23]. This 13th-century manual was written by the English Vinsauf in Latin and the edition drawn upon for our work was translated by Margaret F. Nims into English and published in 1967. Henry Peacham's The Garden of Eloquence, a 14th-century manual, provides a guidebook of rhetorical figures, with examples drawing primarily from Biblical sources, and offers advice on the use of each figure. For our purposes we have consulted the 1593 edition available through the Perseus Digital Library. César Chesneau Du Marsais' French text, Traitè des Tropes [4], is an 18th-century rhetorical manual which treats the study of figures called tropes. John Gent Smith's The Mysterie of Rhetorique Unvail'd is a 17th-century manual which is intended both to illuminate the Bible—in an effort to avoid misinterpretation—and to teach young people how to speak and write more elegantly and persuasively. Thomas Gibbons' Rhetoric: Or, A View Of its Principal Tropes and Figures, in Their Origin and Powers (1767), John Hart's A Manual of Composition and Rhetoric (1874) [12], and John Walker Vilant Macbeth's The Might And Mirth Of Literature (1875) [14] all represent the modern end of our current collection, excluding Gideon O. Burton's [1] and Garrett Epp's [5] work.

Further texts to be incorporated into the database include, but are not limited to: James De Mille's The Elements of Rhetoric [15], Richard Sherry's A treatise of scheme and tropes [19], and Thomas Gibbons' Rhetoric [9]. Additionally, we hope to continue refining our database content; that is, what is taken from primary sources to explicate and elucidate rhetorical figures and their uses. In refining the entries for each figure we are able to gain further insight into useful methods of classification and categorization. Importantly, having a wealth of sources provides some of methods of determining a common or standard usage for terms where some historical conflict has existed.

3.3 Software

In working toward an ontology of rhetorical figures we determined to first aggregate this large collection of disperse information into database software. Our database was designed and developed specifically for our project by a software engineer external to the research group, but who continues to work closely to meet our specific needs. The software application, dubbed “Rhetfig,” is a Python-based application running on top of Google App Engine. As a platform, Google’s App Engine is a scalable, cloud-computing service that allows web applications to be stored on Google's infrastructure.

The first instantiation of the software was operationalized as a data collection method and is currently being developed to include more sophisticated features such as search and sort algorithms. Phase two development of the software, including search and sort, will be complete in late September 2010. Further development of the database will continue on App Engine and will be a completely web-based application accessible by anyone on any computing platform capable of standard web-browsing.

Ultimately we will integrate open source, web-based ontology editing software (protege), for example, but until the categories and individual entries have been further refined we are focused on the theoretical and conceptual issues, and not the programmatic concerns, of an the ontology’s development. That is, we see the theoretical concerns of classifying this extremely large and diverse group of data as preceding any attempt to implement our database work in ontology-editing software.

Fortunately (but also unfortunately) there are literally millennia of taxonomic research into figuration. Some categories are very robust. Others, however, are more controversial. A significant number of figures occupy different categories for different theorists. A significant number, too, are figures to some theorists, but not to other theorists. The task is appreciable.

Theoretical work is, of course, essential to complete prior to further software development as the theoretical work is the basis for software design decisions. Unlike other ontologies in the field of computer science, such as WordNet, this ontology functions as much more than a simple lexical ontology; instead, our work towards an ontology blends the notion of an ontology in computer science and a more classical definition (e.g., what features of a figure are essential, individualizing?), well-suited to rhetorical studies. To better address the quiddity of figures we then turn to rhetorical studies and will return to a discussion of software at the end of our paper.

4. RHETORICAL FIGURES

In Silva Rhetoricæ, Burton describes rhetorical figures as being like “wildflower seeds tossed on fertile ground,” which is to say that these figures “multiplied into a garden of enormous variety over time” [1]. Burton go on to provide a general introduction to rhetorical figures—an excellent introduction for those new to rhetoric—from their early classification and naming in Greek and Roman. Figures are strategies or devices that are employed in language to achieve particular stylistic effects. As we will see, there is a much richer life to these figures than is often assumed under the labeled “stylistics,” but that definition will suffice for now.
One of the best known rhetorical figures is the trope called “metaphor.” This figure is a means of implicit comparison in that the speaker (or rhetor, or author) will use one thing as a perspective on another. For example, the metaphor “Homer is a pig” is a similarity operation which suggests such features as Homer’s eating habits are similar to those of a pig. The purpose of such afiguration is also to activate other parts of the imagination, especially physical sensations. Erotema is the primary term for the figure better known as the rhetorical question. The purpose of this device is not to solicit confirmation or denial, the standard function of a sentence, but to affirm or deny a point by phrasing it as a question that does not require an answer; “Shall I compare thee to a summer’s day?” This device can be used to strengthen the rhetor’s position as expert, or can transfer agency from rhetor to audience as they are asked to provide an alternative although that alternative is believed not to exist. Anastrophe is a figure where the syntactical position of a word is changed for emphasis, such as “glistens the dew upon the morning grass.” This unexpected change is used to alert the audience to listen more attentively to what is being said and can emphasize a word in a phrase which may normally be unattended.

Rhetorical figures, or figures of speech, are part of elocutio, or rhetorical style; indeed, rhetorical figures are what have been described as the “smallest units of rhetorical stylistics” ([20] “figures of speech”). Style here should not be taken as separable from meaning or content. Style is married to meaning and content in rhetorical scholarship, and, accordingly, an important consideration when crafting a text. Understanding how to craft a text using these figures is crucial to constructing persuasive texts and it is evident that many scholars over the past two thousand years have understood this. In fact, the classification and categorization of rhetorical figures has been an ongoing project since Greek antiquity.

Jeanne Fahnestock’s seminal text, Rhetorical Figures in Science (1999), provides not only an excellent discussion of how figures are employed, but an accessible introduction to the history of the classification systems to which figures have been subjected. The initial classification system for figures has been long lost and only vague hints at its development can be found in the early Latin Rhetorica ad Herennium [6]. However, it is evident that the classification of figures had been thoroughly considered and, largely, established by Roman times. Reconfigurations of the classification systems for rhetorical figures would be developed and discussed throughout the rest of the Roman period (e.g., Quintilian’s Institutio Oratoria), through the Early Modern period (e.g., Melancthon’s De Eloquentia or Peacham’s The Garden of Eloquence), and continues today with the work of Gideon O. Burton at Silva Rhetoricae—and, plainly, with the Rhetorical Figure Ontology project [6].

In order to provide guidelines for entering information on figures into the database, certain salient features had to be drawn out and clarified. These include how to choose the primary name and the rhetorical and linguistic domain in which the figure resides. The historical texts provide challenges to these categorizations as it becomes evident that rhetoric has a long and storied past.

5. THE ONTOLOGY

The domain within which our ontology will be constructed is that of rhetorical figures of speech, as defined and described above. Included within the scope of our work are the figures, their individualizing characteristics, and the relationships among figures.

Uses of this ontology, we hope, will be multiple; the end-goal is to provide a comprehensive and robust collection of terms and a suite of tools to manipulate and explore figures and their relations to one another.

For the initial organization of the ontology we intend to devise classes into a hierarchical taxonomy, using a top-down approach of defining general concepts (e.g., linguistic domain) and subclasses (e.g., syntactic, semantic, morphological). Instances can be created from the work that we have been conducting in the construction of our database.

5.1 Historical & Contextual Classification

Each figure features a wealth of historical and contextual information in the database entry. Such information is comprised of basic figure-related information, such as: the figures’ name, its original source, synonyms, etymology, related figures, and examples. Information that might be better described as historical would include the definition, which often demonstrates an evolution of the figure’s meaning, and a list of sources who have discussed the figure in question. Of course, distinguishing between the historical and contextual information is only helpful in attempting to identify the purpose of each category in the database. All of the contextual information is also historical and all historical information provides context. However, the historical aspects of the figures become somewhat secondary to context as our purpose is to classify the figures in some kind of consistent and comprehensive taxonomy. As noted, a figure’s past often only confuses the definition and use of a figure. Accordingly, we see our classification schemes as acknowledging the rich history of a given figure, its inconsistencies, and conflicts while, ultimately, working towards an understanding of the figure as a rhetorical strategy, or device, that can be broadly defined and generalized such that we are able to generalize its nature well-enough to begin the creation of our ontology.

5.2 Linguistic & Rhetorical Classification

Our initial ontology is envisioned as treating specific aspects of rhetorical figures as classes. That is, certain characteristics of rhetorical figures will be broadly defined and organized based on certain linguistic or rhetorical characteristics which will, in turn, provide a governing set of rules and constraints with which to organize our ontology. For the purposes of this paper we will explicate three of the key classes from which we will be able to create a formal taxonomy from our database of rhetorical figures. These three conceptual classifications are: “type of” (scheme, trope or, chroma) distinctions, linguistic domain, and “kind of” relationships. Each of the classification schema are described, defined, and illustrated below.

5.2.1 Schemes, Tropes, & Chroma

The highest-level distinction among the figures in our database is a trisection, into tropes, chroma and schemes:
• **Tropes** are figures whose most salient features are conceptual—such as metaphor (e.g., “my love is a red, red rose”), where the literal falsity draws attention to conceptual similarities between two terms, and synecdoche (e.g., “all hands on deck”), where a representative part is conceptually equivalent to the whole.

• **Chroma** are figures whose most salient features are intentional—such as erotema (rhetorical question; “Do you think I’m an idiot?”), where the question is intended to suggest a proposition not solicit an answer, and apostrophe (addressing someone/something which is not part of the audience; “O, pardon me, thou bleeding piece of earth [Mark Antony to Caesar’s corpse, in order to rouse the mob]”), where the remarks are intended not to move the addressee but the ‘overhearing’ audience.

• **Schemes** are figures whose most salient features are formal—such as rhyme (e.g., “quick flick”), where the sound calls attention to the word and antimetabole (e.g., “I said what I meant, I meant what I said”), where the symmetrical inversion calls attention to the syntax.

Attempting to formalize features of each category has proven a challenging task, but tropes and chroma both prove to be significantly more difficult to treat than schemes [11]. Schemes are described in our system as “formal” because there are rather straightforward linguistic markers that we are able to note and use in our work. For example, anaphora is a rhetorical figure that we have classified as a scheme. This figure describes the repetition of a word or phrase construction at the beginning of contiguous clauses (e.g., “You whom vertue hath made the Princess of felicity, be not the Minister of ruine; you whom my choyce hath made the Load-star of all my sublunary comfort, be not the rock of my shipwrack.” [21]).

Using natural language processing methods, fellow Inkpot researcher Claus Strommer has been able to develop a software application that can mark instances of enanaphora in a text [22]. While somewhat distinct from the ontology itself, Strommer’s work has provided valuable lessons in the way that we can implement the efforts of our conceptual work on rhetorical figures into a computational system. Most importantly, it has demonstrated to us that in terms of formal–such as rhyme (e.g., “quick flick”), where the sound calls attention to the word and antimetabole is preserved by lawes” [16]. In this case it is relatively safe, based on our system of categorization, to place hypoezguna in the syntactic category.

Some figures are less amenable to a single classification. For example, antimetabole has been classified as operating in both the syntactic and lexical domains in our classification system. An antimetabole is a “[r]epetition of words, in successive clauses, in reverse grammatical order” [1]; an example of this figure: “I said what I meant, I meant what I said” [18]. It should be clear from the example that the structure of the phrase is important, as is the repetition of the same words. Accordingly, this figure’s primary or defining linguistic domain is necessarily both lexical and syntactic. A related figure, chiasmus, is defined as “[r]epetition of grammatical structures in inverted order” which is not to be confused with antimetabole; an example of this figure from Silva Rhetoricae: “It is boring to eat; to sleep is fulfilling” [1]. This figure features defining characteristics from both the syntactic and semantic linguistic domains.

Our classification system allows for multiple linguistic domains to be selected by the editors; that is, this is not an exclusive class that only allows for the selection of one domain. Given the proclivity for linguistic domains to be blurred together, with overlapping and mixed boundaries, it is necessary to consider multiple defining characteristics for each figure.

5.2.2 **Linguistic Domain**

The “linguistic domain” describes a system of classification that indicates the primary area within which the figure operates. That is, a distinction is made between several linguistic lines and categories of linguistic boundaries are operationalized (e.g., phonological, syntactic, semantic). Within those boundaries we attempt to classify which rhetorical figures are operating within a respective domain. Our complete class for linguistic domain includes:

- **phonological** refers to speech sounds and sound patterns;
- **morphological** refers to word constructions and forms and variations of those forms (e.g., suffixes, prefixes, co-occurrence);
- **syntactic** refers to clauses and phrases;
- **lexical** refers to words and word relations;
- **semantic** refers to meaning;
- **orthographic** refers to lettering or spelling.

For example, the figure called hypoezguna is classified as syntactic because it is defined as placing essential information at the end of a clause. Peacham provides an example of this construction: “The foundation of freedome, the fountain of equity, the safegard of wealth, and custodie of life, is preserved by lawes” [16]. In this case it is relatively safe, based on our system of categorization, to place hypoezguna in the syntactic category.

Some figures are less amenable to a single classification. For example, antimetabole has been classified as operating in both the syntactic and lexical domains in our classification system. An antimetabole is a “[r]epetition of words, in successive clauses, in reverse grammatical order” [1]; an example of this figure: “I said what I meant, I meant what I said” [18]. It should be clear from the example that the structure of the phrase is important, as is the repetition of the same words. Accordingly, this figure’s primary or defining linguistic domain is necessarily both lexical and syntactic. A related figure, chiasmus, is defined as “[r]epetition of grammatical structures in inverted order” which is not to be confused with antimetabole; an example of this figure from Silva Rhetoricae: “It is boring to eat; to sleep is fulfilling” [1]. This figure features defining characteristics from both the syntactic and semantic linguistic domains.

Our classification system allows for multiple linguistic domains to be selected by the editors; that is, this is not an exclusive class that only allows for the selection of one domain. Given the proclivity for linguistic domains to be blurred together, with overlapping and mixed boundaries, it is necessary to consider multiple defining characteristics for each figure.

5.2.3 **“Kind of” Classification**

The “kind of” category is the least refined category, but provides one of the more interesting systems of classification. Here the categories can describe figures’ operations at multiple linguistic levels. It is the strategies that figures...
use for their effectiveness that are defined in this classification system. Unlike the previous method of classification, linguistic domain, there are no subclasses within this taxonomy. Rather, we have constructed a list of common strategies employed by rhetorical figures as a method to group them based broadly on their techniques. It is noteworthy that this system is currently under going significant revision, but the general structure and intentions will remain in subsequent versions.

- **Repetition** refers to when a figure uses repetition of sounds (consonants, vowels, or syllables), words, syntax (phrases or clauses) or semantics (concepts).
  e.g., Rhyme–Hickory Dickory Dock. The mouse ran up the clock.

- **Symmetry** refers to when a figure pairs two constructions in an inverse (“mirrored”) way.
  e.g. Antimetabole: When the going gets tough, the tough get going.

- **Opposition** refers to when a figure oppose two structures or concepts.
  e.g., Oxymoron: A wanton modesty. Proud humility.

- **Identity** refers to when a figure uses two or more identical elements.
  e.g., Place: O villain, villain, smiling, damned villain! (Hamlet [5]).

- **Similarity (partial-identity)** refers to when a figure uses resemblance of a concept.
  e.g., Simile: My love is like a red, red rose.

- **Omission** refers to when a figure omits expected elements.
  e.g., Asyndeton (which omits the expected conjunction, “and”): I came, I saw, I conquered.

- **Series** refers to when a figure establishes a series (through words or concepts).
  e.g., Abecedarian: Adorable, beautiful, charming, delightful, exciting, fantastic–you run the gamut from A to Z [1].

In this system is it evident that there are classical influences; Peacham, for example, created multiple divisions in this classification of rhetorical figures, including “figures of repetition” and “figures of omission.” We hope that our work here builds on, rather than rehashes, previous work on classification to provide a useful category for broadly defining the strategies employed by rhetorical figures for their effectiveness. Furthermore, we hope that this system can be paired with the linguistic domain classification to reference the levels at which figures may be functioning in both kind and linguistic-level.

6. FUTURE WORK

Presently we are continuing to refine the entries for each rhetorical figure, drawing on the primary texts noted above, as well as expert knowledge from language-related disciplines (i.e., rhetoric, linguistics, and literary studies). In refining each figure’s entry we hope to ensure that its subsequent placement in the ontology, when populated, will accurately reflect strategies and intentions employed by the figure. That is, each figure will have a clear placement in terms of the primary linguistic domain within which it operates, what “kind of” strategies it uses, and so on with the categories described earlier in the paper.

Prior to populating an ontology we hope to ensure that there is a high inter-annotator agreement on each figure’s classification in the fields that we hope to use as classes for the ontology (e.g., type of, linguistic domain, and kind of). To do this we would have annotators independently select the classifications for a set of figures and then determine Cohen’s kappa coefficient for statistical validation of our taxonomy.

Ontology tools that we will create will be integrated into the web-accessible database that has already been implemented. In this way we hope that our work will be included in broader research contexts than our own and, additionally, provide a useful tool for other researchers. Our application has yet to be developed as we are still working towards the completion of version of our database of rhetorical figures.

Additionally, as we populate and refine the database we are engaged in an iterative design process wherein the categories and systems of classification that will come to shape our ontology are continually being revised and refined.

7. FINAL REMARKS

While the work described above is highly theoretical and may first appear to be situated firmly in the domain of rhetorical scholarship alone, the impact of rhetorical figures reaches much further. Rhetorical figures are present in all communication and, as Lakoff & Johnson have famously described in their seminal text *Metaphors We Live By*, figures shape how we think about and through the world. Though Lakoff & Johnson seem unaware of the many precedents, rhetoricians like Giambattista Vico, I.A. Richards, and Kenneth Burke had all advanced similarly compelling arguments about figuration and thought.

In related Inkpot projects we have explored the context of health communications and as many other scholars studying health communications have demonstrated [17], the design of health communication has a significant impact on patients and rhetorical figures are inescapable components of the communicative design of language, especially in those areas concerned with persuasion.

Health care communication is only one example of where rhetorical figures are crucial elements of a persuasive text. Indeed, the presence or absence of carefully chosen rhetorical figures has a profound impact on all genres of communication. Technical writers may find particularly strategies of repetition or ellipsis present in technical communications, for example. Researchers situated more firmly in computer science may find our initial classification schemes to be valuable in detecting figures in a corpus and subsequently diagnosing, modifying, or mining that corpus. With further development of the database and ontology software researchers will be able to sort and search through figures based on a multitude of classifications.

The end goals of the Rhetorical Figure Ontology Project are to support researchers in rhetoric, poetics, cognitive science, human-languages technologies, and any of the other fields (and subfields) for which rhetorical figures have either been traditionally important or in which rhetorical figures are more recently becoming important. As an ongoing project, community feedback will certainly inform our work and looking to form partnerships with others working in re-
lated fields.

Figures are extraordinarily pervasive in language, and extraordinarily fluid; a figure for one generation is literal for the next (becoming, e.g., a ‘dead metaphor’). This situation makes them very challenging to explore. But it also makes them utterly essential for an adequate and responsible account of discourse, and for any investigations, of any sort, into style. They cluster in certain ways relative to affect (anger, joy, sadness), for instance, and relative to genre (newspaper articles, scientific reports, blogs), and relative to functions (persuasion, description, argumentation). Computationally detecting, or reproducing, or modifying any of these phenomena will be greatly facilitated by the ontology we are building.

8. ACKNOWLEDGMENTS

We wish to thank *Silva Rhetoricæ* (that is, its mastermind, Gideon O. Burton), which not only served as the inspiration for this database, but as the primary source for its phase-one population, and as an invaluable training tool for its contributors. We would also like to thank the entire Inkpot Natural Language Research Group for their input and support in the development of this project, Garret Kelly for the design and development of our software, and the University of Waterloo Computer Science Computing Facility staff for their ongoing support. This research is funded by the Social Sciences and Humanities Research Council of Canada (SSHRC).

9. REFERENCES


