Constructing a Rhetorical Figuration Ontology

Randy Harris and **Chrysanne DiMarco**[1]

Abstract. Many essential components of language charted by rhetoric, the ancient study of persuasion, remain understudied and underrepresented in current Natural Language systems. Our goal is to combine linguistic and rhetorical theories with discourse analysis and machine learning to develop formal models of computational rhetoric that may be usefully applied in real-world Computational Linguistics systems. As part of this initiative, we are building an ontology of rhetorical figures and formalizing their expression.

1 INTRODUCTION

Natural Language Processing (NLP) systems are now sufficiently advanced for use in everyday business, educational, and personal applications. Search engines, a prime example, have become an essential part of how academics do their research, how businesses follow trends, and how the average person accesses the Internet. But NLP systems are still challenged by basic problems in understanding the full significance of a text. Current systems generally deal with only restricted language or use simplified methods of analysis, such as shallow parsing. Too much attention has been placed on semantics at the expense of rhetoric (including stylistics, pragmatics, and sentiment). While computational approaches to language have occasionally deployed the word "rhetoric", even in quite central ways (such as Mann and Thompson's Rhetorical Structure Theory [13]), the deep resources of the millenia-long research tradition of rhetoric have only been tapped to a vanishingly small degree. This tradition studies three general attributes of texts that we can formalize and therefore utilize: style (including lexical choice, syntactic structure, and modes of address); purpose (such as description, persuasion, and instruction); and affect (such as trust, deference, and anger).

Our method of "stylistic patterning" is based on the idea that the meaning of an utterance is communicated through the relations among its constituents, as well as their relations with contextual and co-textual elements. We take the notion of style broadly, as "all the choices a writer [or speaker] makes in his or her words and their arrangement" [11, page 14]. Stylistic patterns may be unintended, but they are never meaningless. The question is, how can we get at stylistic meaning? Our model does it with a range of stylistic and rhetorical effects and functions at various levels of discourse organization. In the context of the discourse, stylistic choices contribute to various rhetorical elements: situational parameters or formality levels, as well as intentions, stances, social identities, or moods [17]. As well, some of the stylistic phenomena collectively referred to as "register" are known to be related to text structure and genre. In this paper, we present the first stages of our approach to building a facility for incorporating the stylistic aspects of rhetoric in computational Natural Language systems, specifically, we describe an *ontology of rhetorical figuration* for describing rhetorical patterns that can be used to create persuasive effects in discourse.

2 RELATED WORK

Rhetorical theory has become an increasingly valuable resource to researchers in Natural Language Processing as formal grounding for their computational models and systems. Crosswhite [2] puts succinctly the value of rhetorical argumentation theory to computational linguists: the repositories of formal argumentative schemata (e.g., [14]), rhetorical figures (cf. [6] for evidence of their value), and the representation of the audience, both "a particular audience (with particular values and beliefs) and a universal audience (one that is constructed by imagining away the peculiar local beliefs and attitudes of some actual audience and imagining into this audience the requisite intelligence, memory, attention, knowledge, and so on, so that the resulting audience embodies one's concept of rationality)". The second and third issues are of particular interest to us.

Within the computational community, various formal models of rhetoric have been employed. The recent series of workshops on Computational Models of Natural Argument (CMNA) attest to the growing adoption of models such as Toulmin's [18] logical model, Perelman and Olbrecht-Tyteca's [14] argumentation schemes, and Walton's [19] informal logic for analyzing and evaluating natural argumentation. The usefulness of rhetorical argumentation has so far largely been addressed at a rather abstract level of discourse representation, i.e., formal frameworks and schemata of rhetorical argumentation [8, 12, 9]. Our work is more concrete in its aim: to develop computational representations of fine-grained aspects of style and rhetoric, and apply these to problems that require linguistic expressivity, but where computational efficiency is also a key issue.

3 OUR APPROACH

We are combining Computational Linguistics and Rhetorical Theory to develop formal computational models of style, pragmatics, and sentiment that may be applied in Natural Language systems. For instance, a central concern of rhetoric has been stylistic flaws (such as excessive repetition of terms and dysfluencies of reference), as well as stylistic merits (such as clarity and cohesion). With rhetorical diagnostics we can locate textual deficiencies; with rhetorical strategies, we can repair them. With other diagnostics and strategies we can detect and reinforce or alter the purposes and/or the sentiments of texts. We can re-engineer textual elements to clarify purposes, enhance or reduce emotional effects, and therefore reshape texts as a function of audience, genre, and context. In short, we can tailor texts to specific readers and specific needs. This capacity is especially important in an environment with vast textual reservoirs and widely discrepant

[1] University of Waterloo, Waterloo, Ontario, Canada. Email: raha@uwaterloo.ca; cdimarco@uwaterloo.ca

audiences, such as health care. Computational rhetoric will, for example, allow doctors and other medical personnel to generate patient-specific brochures, matched to criteria like gender, age, reading-level, symptomology, prognosis, collateral conditions, contra-indications, medication, and so on. Our HealthDoc Project [4, 5] has this aim of automatically generating health education tailored to a patient's individual characteristics and medical condition.

It is a difficult challenge to develop expressive, fine-grained ontologies that will lend themselves to use in computationally efficient textual analyzers. Our approach is based on adaptation of well-established rhetorical theories, representation of this theoretical information in precise computational formalisms (e.g., feature-based logics), and implementation in software systems that are computationally efficient. As first steps, we have developed a rich *stylistic ontology* of finely-detailed linguistic features at multiple levels of description [3] and we are now implementing an efficient *stylistic annotator* based on this ontology [16]. Our approach is based on the belief that surface analysis of fine-grained stylistic features, easier and more tractable than full-scale deep-semantic analysis, can yield significantly more, and more meaningful, information than current shallow parsers or statistical methods in a computationally efficient manner.

As part of this work, we have developed a *rhetorical-figure annotation tool* [7], based on traditional definitions of figures, to manually annotate rhetorical figures in various text corpora (e.g., political speeches, health educational materials). Our next step in this work, the automated annotation of rhetorical figures will however require a formal description—an ontology of figuration, in effect—that can be used to characterize and classify rhetorical patterning for use in tasks like recognizing rhetorical strategies such as persuasion and argumentation; detecting and then "repairing" stylistic dysfluencies such as repetitive or awkward text; enhancing or recalibrating redundancies and saliencies for specific aspects of the message; improving credibility and shaping emotional response; and so on. Towards these ends, we are formalizing a set of rhetorical figures in a manner that lends itself to computational representation.

4 RHETORIC

Since the Enlightenment, heavily abetted by the rise of science and technology and their concomitant theories of language as a neutral, context-free, affectless, transparent vessel of communication, rhetoric has fallen into disrepute, with rhetorical figures prominently targetted as the devices of purposive, context-laden, emotional, and opaque language. "Who can behold, without indignation," Thomas Sprat asked, mounting an early modern assault on rhetoric in his *History of the Royal Society*, "how many mists and uncertainties, these specious Tropes and Figures have brought on our Knowledg[e]?")[2]. We pass over the matter of how accurate the windowpane theory of language is, and even the spectacle of Sprat using a rhetorical question to launch his condemnation of rhetoric, to notice simply (1) that the purposive, contextual, and emotional aspects of language are precisely the ones that interest us, (2) that all language manifests these aspects, (3) that rhetoric is not so much a cause of opacity as a methodology for understanding it, and (4) that Sprat is certainly right to implicate figuration in all of these matters. So, we are building an ontology to make rhetorical figures more tractable computationally.

[2] Thomas Sprat. 1667. *The History of the Royal Society of London, for the improving of natural knowledge*. London, J. Martyn and J. Allestry, Printers to the Royal Society. Page 112.

5 FIGURES

Rhetoricians have been studying figuration for millennia, in many languages, under many different theoretical allegiances, with the result that hundreds of overlapping, inconsistent, and even contradictory taxonomies exist. But this very extensive research has produced a rich basis from which we can develop our ontology, and which we have augmented by work in computational linguistics and cognitive science. At the first level of analysis, for instance, two traditional categories are indispensible: tropes and schemes. Tropes, such as metaphors and synecdoches, are conceptual in nature. Schemes, such as alliteration and polyptoton, are formal.

Metaphors rely on the cognitive principle of comparison (*Jeff is a brick* compares a person to an object known for its solidity), synecdoche on *PartOf* representation (*All hands on deck* identifies sailors by singling out aspects of their anatomy critical for sailorly tasks). The form a trope takes is secondary to the conceptual principles at work. If there is no expression of comparison, there is no metaphor. We can say *A brick, that's Jeff.* We can say, *Jeff is the brick of that family.* And so on, in as many syntactico-lexical configurations as we have the imagination, and we still have a metaphor. But, if we say, *Jeff is a stable guy*, the metaphor is gone (strictly speaking, of course, we bring in another metaphor, since *stable* is fundamentally a physical term, and here we are using it for emotional and social purposes, but it is a subtler, more 'literal' sort of metaphor). Schemes, conversely, are formal in nature, and their conceptual operation is secondary to their structural arrangement.

Alliteration is the consecutive use of words with the same initial consonants (*Peter Piper picked a peck. . .*); the semantics of those words is irrelevant to the scheme. Polyptoton is the use of one word stem in a variety of morphological instantiations (*That team is the suckingest bunch of sucks that ever sucked*); again, semantics are not part of the equation. For this reason, schemes are of course the most amenable to computational detection and manipulation, and we are concentrating our early energies on them.

6 TOWARD AN ONTOLOGY OF RHETORICAL FIGURES

Despite the extensive number of rhetorical figures that have been catalogued over two millennia[3], 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 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. As an example, consider some schemes of omission, in which normally expected elements are implied rather than stated. The most familiar such figure is ellipsis, in which a lexeme is omitted (e.g., *John forgives Mary and Mary, John*). Ellipsis is a clear example simultaneously of how linguistics as a field has drawn on rhetoric (frequently without acknowledgement or even awareness), and of how figures permeate mundane language processes [10]. But there is a wide range of omission schemes. Zeugma, for instance, is a scheme in which one lexeme (usually the main verb of a sentence, sometimes a noun or adjective) governs two or more other lexemes in a series. Zeugma *IsA* ellipsis, that is. Moreover, zeugma includes multiple types, depend-

[3] Sylva Rhetoricae [http://humanities.byu.edu/rhetoric/silva.htm], a superb online resource for rhetorical figures, lists 433 distinct figural terms.

ing on the kind of governing lexeme or its placement in the clause, including:

Prozeugma: The verb in the first of a series of clauses governs the noun phrases in the remaining clauses in the series (e.g., *Her beauty pierced mine eye, her speech mine woeful heart, her presence all the powers of my discourse.*).

Hypozeugma: A verb follows a series of words or phrases that it governs (e.g., *Friends, Romans, countrymen, lend me your ears.*).

Epizeugma: The verb that completes a predicate occurs at either the very beginning or the very ending of its sentence (e.g., *Fades beauty with disease or age. Either with disease or age beauty fades.*).

Mesozeugma: The verb governing multiple subjects occurs in the middle of a construction that contains them all (e.g., *Neither his father nor his mother could persuade him; neither his friends nor his kinsmen.*).

Each of these figures, in short, *IsA* zeugma (in this case, the relationship is conveniently signalled by traditional nomenclature). Function words might also be omitted: asyndeton *IsA* ellipsis in which conjunctions between clauses are omitted (e.g., *government of the people, by the people, for the people*). At the sub-lexical level, syncope (another clear example simultaneously of how linguistics has drawn on rhetoric and on how figures deeply interpenetrate with ordinary-language processes) *IsA* medial ellipsis of phones or syllables, such as when *library* is pronounced *libary*. Apocope *IsA* terminal ellipsis, in which the final sound or syllable is omitted, as in the back-formation of *pea* from *pease*. A poetic example that includes both apocope and syncope, in that order, is Alexander Pope's *What oft was thought, but ne'er so well expressed*[4].

Figures are also related through opposition. Diazeugma, for instance, *IsOppositeOf* zeugma, since it is a scheme in which a single subject governs several verb phrases (usually arranged in parallel fashion and expressing a similar idea), as in *The Romans destroyed Numantia, razed Carthage, obliterated Corinth, overthrew Fregellae*. Polysyndeton *IsOppositeOf* asyndeton, since it involves the elaborate use of conjunctions between clauses, as in this passage from Hemingway's *After the Storm*:

> I said, "Who killed him?" and he said, "I don't know who killed him but he's dead all right," and it was dark and there was water standing in the street and no lights and windows broke and boats all up in the town and trees blown down and everything all blown and I got a skiff and went out and found my boat where I had her inside Mango Key and she was all right only she was full of water[5].

Figures are also related through relations of inclusion. Consider schemes of iteration. Ploche is simple lexical repetition (*She is tall, very tall*), so it naturally participates in schemes of complex lexical repetition, such as antimetabole (lexical repetitions in reverse order: *Ask not what your country can do for you—ask what you can do for your country*) and epistrophe (repetition at the ends of phrases: *government of the people, by the people, for the people*).

[4] Alexander Pope. *An Essay on Criticism*. Whitefish, MT: Kessinger Publishing, 12, 2004.

[5] Ernest Hemingway. *The Complete Short Stories of Ernest Hemingway*. New York: Simon and Schuster, 283, 1998.

7 FORMALIZING SCHEMES

In our formalization of schemes we use the descriptive elements shown in Table 1. A portion of our current set of rhetorical figures

Table 1. Formalism for Representing Rhetorical Figures

Element	Meaning
Cl	clause
Phr	phrase
W	word
Vb	verb
N	noun
S	stem
M	morpheme
C	consonant
V	vowel
P	phone
Sy	syllable
$\oslash$	gap
X/Y	X INSTEAD OF Y
$\ldots$	arbitrary intervening material
	(possibly null, with some upper limit, shorthand is *proximal*)
$\{\ldots\}$	morpheme boundaries
$[\ldots]$	word boundaries
$< \ldots >$	phrase or clause boundaries
	(assuming clauses are just special types of phrases,
	aggregating other phrases)
Subscripts	identity (same subscripts), nonidentity (different subscripts)

and their formalizations is shown below, together with sample realizations.

Adage (apothegm, gnome, paroemia, proverb, sententia, maxim): Use of familiar, traditional expressions.

(1) $X,\ where\ X = \{x_1, x_2, \ldots x_n\}$

That is, locate any occurrence of a pattern that is stored in an adage box somewhere (*A bird in the hand is worth two in the bush*, *Never look a gift horse in the mouth*, etc.). We are not yet certain how feasible doing this type of recognition automatically would be. There are two complications that we can foresee: first, finding a collection of adages/proverbs/idioms that we could represent easily in the right computational format for the pattern-recognizer; second, specifying substrings appropriately. For instance, we might encounter something like *She preferred a bird in the hand to a speculative treatment*, which evokes the adage, but doesn't fully replicate it.

Alliteration: The repetition of consonants at the beginning of proximal words.

(2) $[C_a \ldots] \ldots [C_a \ldots]$ *Lopsided loons lull listening lovers.*

Anadiplosis: Starting a clause or phrase with the word or phrase that ended the preceding unit.

(3) $< \ldots [W]_a >< [W]_a \ldots >$ *Drake covets loons, loons with cash.*

(4) $< \ldots < \ldots >_a >< < \ldots >_a \ldots >$ *Drake covets lopsided loons, lopsided loons with cash.*

Anaphora: The repetition of a word or group of words at the beginning of successive clauses or phrases.

(5) $< [W]_a \ldots >< [W]_a \ldots >$ *Drake covets loons. Drake loves cash. Drake wants fame.*

(6) $< [W]_a[W]_b \ldots >< [W]_a[W]_b \ldots >$ *Drake covets loons. Drake covets cash. Drake covets fame.*

Antimetabole: Repetition of words in reverse order.

(7) $[W]_a \ldots [W]_b \ldots [W]_b \ldots [W]_a$ *Drake loves loons. Loons love Drake.*

Apocope: Word-terminal ellipsis, in which the final sound or syllable of a word is omitted.

(8) $[\ldots \oslash]/[\ldots P]$ *yank / yankee*

(9) $[\ldots \oslash]/[\ldots Sy]$ *doc / doctor*

Assonance: The repetition of vowels in proximal syllables or words.

(10) $\{\ldots V_a \ldots V_a \ldots\}$ *anaphora*

(11) $\{\ldots V_a \ldots\} \ldots \{\ldots V_a \ldots\}$ *Drake covets lovely cash.*

Asyndeton: The deliberate omission of conjunctions within a series of related clauses.

(12) $\{Cl_a \ldots Cl_b \ldots Cl_c \ldots\}$*I came, I saw, I conquered.*

Consonance: The repetition of consonants in proximal syllables or words.

(13) $\{\ldots C_a \ldots C_a \ldots\}$ *kakaphobia*

(14) $\{\ldots C_a \ldots\} \ldots \{\ldots C_a \ldots\}$ *Weak Dickensian plots.*

Diazeugma: A single subject governs several verb phrases.

(15) $< [N]_a \ldots >< \oslash_a Vb_b \ldots >< \oslash_a Vb_c \ldots >< \oslash_a Vb_d \ldots > \ldots$ *Drake covets loons, loves cash, wants fame.*

Ellipsis: Omitting an expected element.

(16) $\ldots \oslash \ldots$ *Drake likes loons, Bill geese.*

Epistrophe: Ending a series of phrases or clauses with the same word or words.

(17) $< \ldots [W]_a >< \ldots [W]_a >$ *Drake likes loons. Bill likes loons.*

Epizeugma: The verb which completes a predicate occurs at either the very beginning or the very ending of its sentence.

(18) $< Vb \ldots >$ *Surfaces daily at noon, the loon.*

(19) $< \ldots Vb >$ *The loon, daily at noon, surfaces.*

Epizeuxis (palilogia): The repetition of the same word with no others between.

(20) $[W]_a[W]_a$ *Look at the loon-loon.*

Homoioteleuton: The repetition of suffixes in proximal words.

(21) $[S_a\{M\}_a] \ldots [S_b\{M\}_a]$ *Leaping, jumping, loons.*

Hypozeugma: A verb follows a series of words or phrases that it governs.

(22) $< Na, Nb, Nc, \ldots Vb >$ *Drake, Bill, Amanda, love those waterfowl.*

Mesozeugma: The verb governing multiple subjects occurs in the middle of a construction that contains them all.

(23) $< \ldots \oslash_a \ldots >< \ldots \oslash_a \ldots >< \ldots Vb_a \ldots >< \ldots \oslash_a \ldots >< \ldots \oslash_a \ldots >$ *Husbands and wives, sons and daughters, loved that warbling loon; sisters and brothers, uncles and aunts.*

Ploche (ploce, repetitio): The repetition of the same word in a short span of text[6].

(24) $[W]_a \ldots [W]_a$ *Loons like what loons know.*

Polyptoton: The repetition of a word, but in a different form (i.e., the repetition of a stem, with a difference in affixes)[7].

(25) $[S_a\{M_a\}] \ldots [S_a\{M_b\}]$ *Looney loons.*

(26) $[\{M_a\}S_a] \ldots [S_a\{M_b\}]$ *Nonloon loons.*

(27) $[S_a\{M_a\}] \ldots [\{M_b\}S_a]$ *Looney nonloon.*

(28) $[\{M\}S_a\{M\}] \ldots [S_a]$ *Nonlooney loon.*

(29) $[S_a\{M\}] \ldots [S_a]$ *Looney with his loon.*

and so forth.

Polysyndeton: 'Excessive' use of marked conjunction[8].

(30) *and ... and ... and ... Drake like loons and swans and geese.*

Prozeugma: The verb in the first of a series of clauses governs the noun phrases in the remaining clauses in the series, which correspondingly contain no verbs.

(31) $< \ldots Vb_a \ldots >< \ldots \oslash_a \ldots >< \ldots \oslash_a \ldots >< \ldots \oslash_a \ldots >$ *Drake likes loons, Bill swans, and Amanda geese.*

Syncope: A medial ellipsis of phones or syllables.

(32) $[\ldots \oslash \ldots]/[\ldots P \ldots]$ *ev'ry / every*

(33) $[\ldots \oslash \ldots]/[\ldots Sy \ldots]$ *ma'am / madam*

8 A SEED ONTOLOGY OF SCHEMES

From the list of schemes above (and there are many more) we can begin to organize these schemes into a "seed" ontology using the specialization (*IsA*) relation, as well as other relationships such as near-synonymy and antonymy. We begin by classifying schemes according to whether there is a pattern of addition (*InclusionScheme*), deletion (*OmissionScheme*), or permutation (*PermutationScheme*). These concepts are placed at the top level of the ontology as specializations of the generic *Scheme* concept.

We can now add to the ontology, using organizing principles based on stylistic/linguistic features like repetition, placement, and lexical governance. One level down in the ontology we specialize *InclusionScheme* into the subclass *Iteration*, which is further refined into *PlacementIteration*. We can say furthermore that *Ploche* (simple repetition) *IsA Iteration*, while in turn *Antimetabole* and *Epizeuxis* (more-complex types of repetition) are both specializations (i.e., *IsA*) *Ploche*. Subclasses of *PlacementIteration* include *Anadiplosis*, *Epistrophe*, and *Anaphora*.

The second main category of schemes, *OmissionScheme*, has a rich network of subclasses. An *Ellipsis IsA OmissionScheme* and has two main subclasses, *PlacementEllipsis* and *GoverningEllipsis*, or *Zeugma*. *PlacementEllipsis* has three specializations, *InitialEllipsis*, *MedialEllipsis*, and *TerminalEllipsis*. At the next level down in the

[6] There is some potential for confusion with polyptoton, following, given the mushiness of the notion 'same word'.

[7] The M could be null in any expression. Commonly, this shows up in cases like *My new friend is quite friendly.* But, in cases where both Ms are null, there is confusion with ploche, as in *Most of us use margarine, but he actually uses butter to butter his toast*, where the first *butter* is a noun, the second a verb.

[8] Here we've just stipulated three, assuming that greater-than-three comes along "for the ride".

ontology, *Asyndeton IsA MedialEllipsis* (at the syntactic level), *Syncope IsA MedialEllipsis* (at the sub-lexical level), and *Apocope IsA TerminalEllipsis*. *GoverningEllipsis*, the other form of *Ellipsis*, has subclasses *Prozeugma*, *Hypozeugma*, *Epizeugma*, and *Mesozeugma*.

In addition to the *IsA* relation, we can define relationships between *Scheme* concepts based on near-synonymy, part-whole (*PartOf*), and antonymy (*IsOppositeOf*). For example, *Diazeugma IsOppositeOf Zeugma*. The organization of these basic *Scheme* concepts to form our seed ontology is shown in Figure 1.

9 APPLICATION: RHETORICAL FIGURATION IN POLITICAL SPEECHES

Political rhetoric is a particularly interesting topic for investigating formal models of persuasion, e.g., [1, 15]. Our work takes a distinctive approach in piecing together the stylistic choices made across multiple levels of linguistic description through textual analysis using stylistic and rhetorical ontologies based on an integrated model of meaning. Barack Obama's inaugural address offers a particularly rich site for our computational rhetorical approach. Beginning literally from the moment of its completion, it attracted hosts of commentators. Most notably, a controversy quickly broke out between literary critic Stanley Fish and linguist Mark Liberman. Fish, in his opinion piece for the *New York Times*, noted that the oratorical leanness was, at least in part, meant to serve a literary richness. That is, *oral* eloquence was sacrificed for *written* complexity. The speech worked best as a literate artifact, "a framework on which a succession of verbal ornaments [were] hung", so that the audience was not being "invited… to move forward" so much as "to stop and ponder significances", as in an art gallery. An oral performance, of course, moves through time, largely independent of the hearer, whereas a written document is arrayed in space, allowing the readers to top and ponder at their leisure. The main stylistic feature Fish correlates with this stop-and-ponder literate dimension is parataxis (Gk., "arrange side by side"), the contiguous arrangement of equivalent clauses, such as Julius Caesar's "I came, I saw, I conquered." Parataxis is frequently defined in opposition to hypotaxis (Gk., "arrange under"), the syntax of subordination, and Fish argues that "the prose of Obama's inauguration is surely more paratactic than hypotactic".

Our role here is not to adjudicate between Fish and Liberman, beyond that effective oratory can work on the page *and* on the stage, and work quite distinctly in the two modes. But their exchange forms a useful backdrop to the sort of interpretation of Obama's rhetoric that our ontology supports. We can, for instance, get at the 'rhetorical austerity' issue by way of our features, polysyndeton (Gk., "many connectives") and asyndeton (Gk., "no connectives"). An instance of the former is "these men and women struggled and sacrificed and worked", with its abundance of *and*s; of the latter, "it has been the risk-takers, the doers, the makers of things", with its absence of the expected *and*. In large part, Fish is responding to the high asyndeton/polysyndeton ratio, such that there is a relatively low number of explicit connectives in Obama's inaugural address. We can see this all the more clearly if we compare another speech of Obama's, one that does have the kind of familiar eloquence that was expected on January 20, 2009. In the inaugural address, we have a ratio of 19/4 asyndeton over polysyndeton. In Obama's famous speech on winning the South Carolina Democratic presidential primary, commonly known as the *Yes We Can* speech, the ratio is far different: 6/8. The high proportion of asyndeton in *The Inaugural* goes a long way to explaining most interpreters' sense of terseness in the speech and especially Fish's stop-and-ponder opinion. The address has several paratactic passages, but the effect of these are heightened substantially by asyndeton. From these results, we can detect differences between the speeches in both register (level of formality) and ethos (variance in roles of both speaker and audience): *Yes We Can* is much less formal than *The Inaugural*, and the speaker positions himself differently towards the audience in these two speeches, an individual candidate trying to win over the electors versus the institution of the President issuing a call-to-action and responsibility.

10 CONCLUSION

We have taken the first steps towards constructing and formalizing an ontology of rhetorical figures for use in Natural Language systems such as automated stylistic annotators and text generation systems. Many very difficult problems remain to be solved. For example, because of the complexities of English spelling, we will need somewhat elaborate notions of consonants and vowels, not just simple letter repetition. Word, phrase, and morpheme are not especially simple notions, either, but working with text we should be able to leverage all kinds of cues (blanks, hyphens, punctuation generally).

Tropes provide a wealth of semantic difficulties and even many schemes provide obstacles to formalization. Formalizing isocolon (a series of similarly structured phrases), for instance, requires figuring out how to formalize—and recognize—the notion of similarly structured phrases. In principle, the problem is pretty much the same as polyptoton, in which we have the 'same' word, in different forms (*suck*, *suckingest*, *sucked*). In isocolon, we have structurally the 'same' phrase, with at least some different constituents (*I came, I saw, I conquered.*). We might also represent the parallels between tropes and schemes using formal ontological relationships: for example, synecdoche expresses a conceptual *PartOf* operation; ploche is formally *PartOf* antimetabole. Our ultimate goal is to work out a hierarchy of rhetorical figures in the form of a formal ontology, since so many figures obviously implicate others.

Our project has far-reaching implications for Natural Language Processing, as does the general turn toward persuasion in NLP (which we might call the Rhetorical Turn, borrowing a phrase from Richard Rorty in science studies[9]. As two examples, consider the long tradition of studying ethotic figures (that is, figures which support the credibility of the speaker) and of studying pathotic figures (figures which affect the emotional quotient of the discourse). The former can be very valuable for developing or locating or managing discourses in which trust is critical. The latter can be equally valuable for sentiment analysis and management. Conversely, ineffective, inappropriate, or awkward uses of figures can be detected and ameliorated. We are at the outset of a long, and we forecast, very rich collaboration between the ancient discourse technologies of rhetorical theory and the contemporary discourse technologies of Natural Language Processing.

ACKNOWLEDGEMENTS

We would like to thank the members of the Inkpot Natural Language Research Group at the University of Waterloo for many discussions and valuable input. We are grateful to Fred Kroon for his helpful assistance in preparing the figure.

[9] Herbert W. Simons. *The Rhetorical Turn.* London: Sage, 1990.

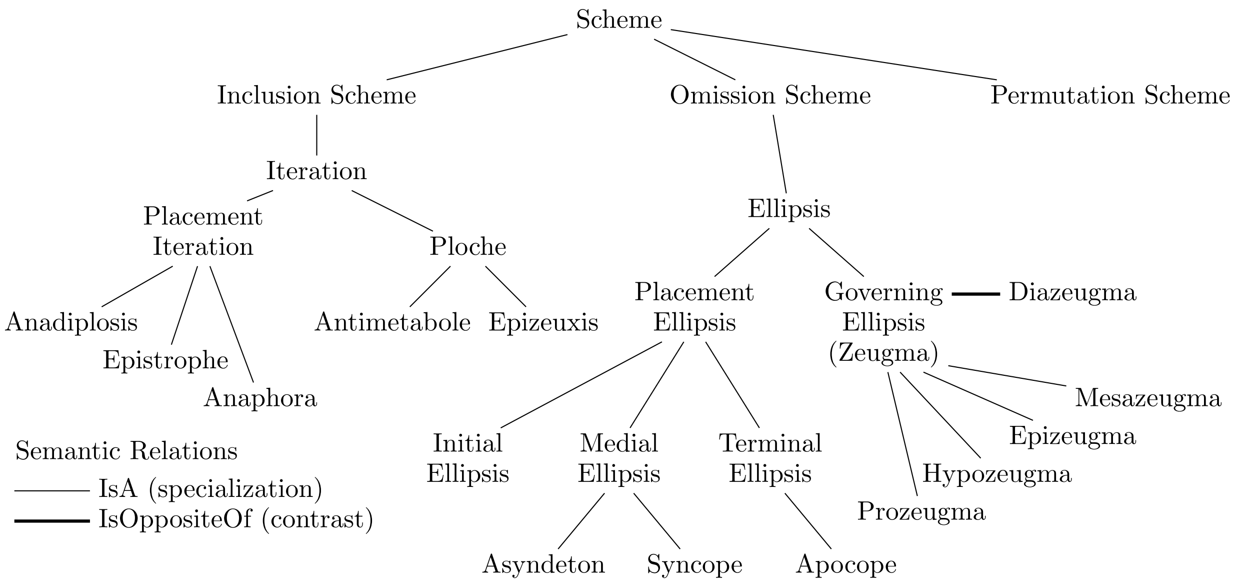

Figure 1. A seed ontology of rhetorical schemes

REFERENCES

[1] Katie Atkinson, Trevor Bench-Capon, and Peter McBurney, 'Persuasive political argument', Proceedings of the 2005 Workshop on Computational Models of Natural Argument (CMNA V), (2005).

[2] James Crosswhite, 'Rhetoric and computation', In: C.A. Reed and T. Norman (eds.) Symposium on Argument and Computation: position papers, `http://www.csd.abdn.ac.uk/~tnorman/sac/`, (2000).

[3] Chrysanne DiMarco, Steven Banks, Olga Gladkova, and Matthew Skala, 'A fine-grained multi-level ontology for computational stylistics', Submitted to 2009 Conference of the North American Association for Computational Linguistics (NAACL), (February 2009).

[4] Chrysanne DiMarco, Peter Bray, H. Dominic Covvey, Donald D. Cowan, Vic DiCiccio, Eduard Hovy, Joan Lipa, and Cathy Yang, 'Authoring and generation of individualized patient education materials', *Journal on Information Technology in Healthcare*, (February 2008).

[5] Chrysanne DiMarco, H. Dominic Covvey, Peter Bray, Donald Cowan, Vic DiCiccio, Eduard Hovy, Joan Lipa, and Doug Mulholland, 'The development of a natural language generation system for personalized e-health information', 12th International Health (Medical) Informatics Congress, Medinfo 2007, Brisbane, Australia, (2007).

[6] Jeanne Fahnestock, *Rhetorical figures in science*, Oxford University Press, USA, 2002.

[7] Jakub Gawryjolek, *Automated Annotation of Rhetorical Figures (working title)*, Master's thesis, University of Waterloo, expected 2009.

[8] Floriana Grasso, 'Towards a framework for rhetorical argumentation', 53–60, In: J. Bos, M.E. Foster and C. Matheson (eds.) Proceedings of the 6th Workshop on the Semantics and Pragmatics of Dialogue (EDILOG 2002), Edinburgh, UK, 4-6 September 2002, (2002).

[9] Nancy L. Green, 'Dialectical argumentation in causal domains', Proceedings of the 2008 Workshop on Computational Models of Natural Argument (CMNA VIII), (2008).

[10] Kyle Johnson, *Topics in ellipsis*, Cambridge University Press, Cambridge, 2008.

[11] Thomas S. Kane, *The Oxford guide to writing*, Oxford University Press, 1983.

[12] Fabrizio Macagno and Doug Walton, 'Argumentative reasoning patterns', Proceedings of the 2006 Workshop on Computational Models of Natural Argument (CMNA VI), (2006).

[13] William C. Mann and Sandra A. Thompson, 'Rhetorical structure theory: Toward a functional theory of text organization', *Text*, **8**(3), (1988).

[14] Chaïm Perelman and Lucie Olbrechts-Tyteca, *The new rhetoric: A treatise on argumentation*, University of Notre Dame Press, Notre Dame, Indiana, 1969.

[15] Isabella Poggi and Laura Vincze, 'Showing disinterest: A persuasive strategy to win the electors' trust', Proceedings of the 2008 Workshop on Computational Models of Natural Argument (CMNA VIII), (2008).

[16] Matthew Skala, Chrysanne DiMarco, Olga Gladkova, Elena Afros, Steven Banks, Randy Harris, and Frederick W. Kroon, 'A parser-agnostic stylistic annotator', To be submitted to *Computational Linguistics*, (February 2009).

[17] S.M. Sotillo and J. J. Wang-Gempp, 'Using corpus linguistics to investigate class, ideology, and discursive practices in online political discussions', 91–122, In: U. Connor and T.A. Upton, *Applied Corpus Linguistics*, 2004, (2004).

[18] Stephen Toulmin, *The uses of argument*, Cambridge University Press, 1958.

[19] Douglas Walton, *Informal logic: A handbook for critical argumentation*, Cambridge University Press, 1989.