

An Annotation Tool for Automatically Detecting Rhetorical Figures SYSTEM DEMONSTRATION

Jakub Gawryjolek

Chrysanne DiMarco

Randy Harris

University of Waterloo

Waterloo, Ontario, Canada

jjgawryj@uwaterloo.ca; cdimarco@uwaterloo.ca; raha@uwaterloo.ca

1 Introduction

Linguistic annotation provides additional information asserted with a particular purpose in a document or other piece of information. It is widely used in various fields, from computing and bioinformatics, through imaging, to law and linguistics. There is also a clear distinction between *what* is communicated through the written/spoken natural language and *how* this is passed on. We are studying a new problem in linguistic annotation—annotation of classical rhetorical figures. We have developed a *rhetorical annotation tool* that can automatically detect patterns of text in which a characteristic syntactic form modifies the standard meanings of words, leading to a change or an extension of meaning. The annotation of rhetorical figures is important not only from a purely linguistic analysis, but also for discovering different styles of writing, purpose, and effect in documents, and for more-sophisticated Natural Language Understanding systems in general.

2 The Formalization of Figures

Despite the extensive number of rhetorical figures that have been catalogued over two millennia¹, they fall into a relatively few, partially overlapping classes. In formalizing a subset of these figures (several shown below) [Harris and DiMarco, 2009], 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.

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

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

Antimetabole: Repetition of words in reverse order.

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

¹Sylva Rhetoricae [<http://humanities.byu.edu/rhetoric/silva.htm>], an online resource for rhetorical figures, lists 433 distinct figural terms.

Isocolon: A series of similarly structured phrases.

Ploche (ploce, repetitio): The repetition of the same word in a short span of text².

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

3 JANTOR: A Rhetorical Annotation Tool

We have developed a rhetorical-figure annotation tool dubbed JANTOR (Java ANnotation Tool Of Rhetoric) [Gawryjolek, 2009], which enables manual and automated annotation of files in HTML format. We have applied a lexicalized probabilistic context-free grammar parser for the recognition of the figures of repetition. We use a simple parse-tree distance for calculating the difference between similarly structured phrases, which is necessary for the recognition of some of the figures of parallelism. Moreover, we have applied the semantic relationships contained in the WordNet lexical database and extended the Porter stemmer algorithm for finding derivationally related words. Lastly, we have developed a method for finding pairs of words which are ordinarily contradictory, which is crucial for detecting the interesting figure of speech, *oxymoron*. For this purpose, typed dependency grammars together with WordNet were used. The experiments we have so far conducted on the automated detection of a selected subset of rhetorical figures have yielded promising results.

We created a small collection of eleven HTML documents for testing the precision and recall of our rhetorical-figure detection system. Each of the files consists of the examples of one particular type of figure. The examples were taken from various sources—excerpts from political speeches, commercials, the Bible, poetry, prose, lyrics, film dialogues, etc. As sources we used the Wikipedia, About.com, American Rhetoric, and Silva Rhetoricae websites, as well as Corbett's [Corbett, 1990] well-known textbook on classical rhetoric.

In its current stage of development, the JANTOR rhetorical annotation tool supports the automated detection of 11 figures of speech. Apart from the figures of repetition, more specifically the repetition of words, the annotation tool also

²There is some potential for confusion with polyptoton, following, given the mushiness of the notion 'same word'.

enables the detection of isocolon, a figure of parallelism, and oxymoron, a figure of semantic contradiction between words. Our results for precision and recall were very promising as far as figures of repetition are concerned. The results for the detection of the other forms of rhetorical devices were satisfactory, but suggest that improvements are still necessary for these more difficult figures.

4 Demonstration: Annotation and Visualization of Presidential Speeches

We applied JANTOR to the annotation of figures of speech in the inaugural addresses of 14 American Presidents. In our live demonstration of JANTOR, we will show the visualizations of the rhetorical annotations of the speeches. Here we summarize the highlights of our findings.

The inaugural address, an example of political speech, shares many characteristics with this genre in general, but an inaugural speech is special as it is the first ‘dialogue’ between the new president and the citizens of the country. The communicative function as well as the rhetorical strategies have to be well-developed and thought through in detail in order to successfully have the intended effect on the audience, and to convey the president’s message to the nation.

As our source we used the archives of The American Presidency Project website³, which contains 86,001 documents related to the study of the Presidency. The Presidential speeches we examined included those by Barack Obama (2009), George W. Bush (2005), William Jefferson Clinton (both 1993 and 1997), Ronald Reagan (1981), Jimmy Carter (1977), Richard Milhous Nixon (1969), John Fitzgerald Kennedy (1961), Harry Truman (1949), Franklin Delano Roosevelt (1933), Woodrow Wilson (1913), Theodore Roosevelt (1905), Abraham Lincoln (1861), Thomas Jefferson (1801), and George Washington (1789). Altogether, we analyzed and compared 15 speeches.

4.1 Intensity—Washington versus Obama

Our first analysis concentrated on the intensity of usage of various rhetorical figures. We do not take into account exactly where in the text the figures occur, but only count the number of occurrences. This examination is meant to give just a general overview of the ‘rhetorical maturity’ of the speakers⁴. We consider first the inaugural addresses of the current and first presidents, Barack Obama and George Washington. Table 1 presents the number of selected figures of speech in their respective addresses. We have set the minimum length of a repeating-word sequence to two for anaphora and epistrophe.

As we can observe, Obama used much more anaphoras and polysyndetons in his address. Also, because the minimum length of a repeating sequence was set to two, the number of his epistrophes is one. The system correctly identified the following epistrophe:

- (1) Our challenges **may be new**. The instruments with which we meet them **may be new**.

³<http://www.presidency.ucsb.edu>

⁴The numbers of figures presented in this section are estimates. Due to various reasons concerning precision of the annotation tool, we cannot as yet determine exactly the number of figures.

When we changed the minimum length of a sequence to one, JANTOR also found the following memorable expression:

- (2) All this we can **do**. All this we will **do**.

The lengths of the speeches measured in the number of words were 2406 and 1435 for Obama and Washington, respectively (using the Linux “wc” command). Even though the Obama speech is roughly 1000 words longer, we can definitely conclude that Obama uses the power of anaphora to a larger extent. He also uses more polysyndetons, so the pace of his speech is sometimes slower, but certain fragments are more emphasized. We also observe that the sentences used by Washington were significantly longer on average.

One last observation concerns polyptotons—both presidents used words in different forms almost the same number of times. Keeping in mind that Washington’s speech is significantly shorter, we might say he ‘wins’ in this category. However, the individual words in this figure were much further apart than the ones in polyptotons used by Obama. The example below presents one of Obama’s uses of polyptoton.

- (3) The Nation cannot **prosper** long when it favors only the **prosperous**. The success of our economy has always depended not just on the size of our gross domestic product, but on the reach of our **prosperity**, on our ability to extend opportunity to every willing heart, not out of charity, but because it is the surest route to our common good.

Even though Obama’s speech is considered to be very pragmatic, not poetic like his campaign speeches⁵, and deliberately not flowery⁶, rhetorically it seems well-developed. Reiterations allow the audience to thoroughly think over and understand his words. The usage of many stylistic devices helps make his audience truly understand the assertiveness of his language.

4.2 Placement

Our first observation about the nature of placement figures in the two speeches is that some parts of polyptotons in Obama’s text are missing. This is because certain words such as *our*, *ours* were too short with respect to the length of the text to be captured in the visualization. Our tests showed though that Washington’s parts of polyptoton seem to be more far apart.

Another conclusion we were able to draw pertains to the distribution of placement figures. Mainly due to the excessive use of (rhetorical) anaphora, the figures in Obama’s speech occur more uniformly. When we excluded anaphora, the distributions did not vary so much. However, what is noticeable is the stylistic difference between the two speakers in their usage of figures in their opening paragraph—Obama did not use any rhetorical devices, while Washington seems to have bombarded the audience with rhetorical figures in his opening.

We next compared the use of isocolon between the addresses of Barack Obama and another accomplished political rhetorician, John Fitzgerald Kennedy. We also included

⁵Frank Rich, “No Time for Poetry”, *New York Times*, January 25, 2009.

⁶Jeff Greenfield, CBS News video on Obama Inaugural Address, <http://www.cbsnews.com/video/watch/?id=4738702n>.

Table 1: Number of figures in the inaugural addresses

Figure	Barack Obama	George Washington
anaphora	40	10
epistrophe	1	0
antimetabole	6	1
epanalepsis	0	0
anadiplosis	0	0
polyptoton	10	9
polysyndeton	15	6

Ronald Reagan and Richard Nixon for additional comparison. The examination and visualization of the placement of isocolon led us to a provocative, but perhaps expected, conclusion. Thoroughly structured texts, which inaugural addresses usually are, are rhythmic not only within the scope of individual paragraphs, but also throughout the whole speech. In our visualizations of isocolon placement, we could easily spot the areas of increased intensity of usage. Kennedy’s address is exceptional in this regard. Not only are the areas of increased usage clearly evident, but they also repeat in roughly the same intervals.

Isocolon is most excessively used by Nixon and Obama, and although zones of increased usage are distributed over the entire text, they are still clearly evident. Reagan’s speech does not seem as strongly structured in terms of the usage of this figure, but zones of usage can also be observed in his address. Finally, all four presidents emphasized the end of their speeches by the use of parallelism.

Below we present several memorable examples of isocolon in presidential speeches, which were all detected by JAN-TOR.

- (4) Our security emanates from **the justness of our cause, the force of our example**, the tempering qualities of humility and restraint. (Barack Obama)
- (5) If a free society **cannot help the many who are poor, it cannot save the few who are rich**. (...) And so, my fellow Americans: *ask not what your country can do for you—ask what you can do for your country*. (John Kennedy)
- (6) Government can and must **provide opportunity, not smother it; foster productivity, not stifle it**. (Ronald Reagan)

4.3 Rhetorical Evolution?

The final stage in our study of presidential speeches was to search for any rhetorical trends within the inaugural speeches of the same president. For this purpose we analyzed the two addresses delivered by President Bill Clinton. The figures we included in this comparison were: anadiplosis, anaphora, antimetabole, epanalepsis, epistrophe, epizeuxis, polyptoton, and polysyndeton. Below we describe several regularities we were able to observe.

First, the figures of epanalepsis, epizeuxis, and anadiplosis were very rare in both speeches. For anaphora, the first trend we can observe is the absence, followed by the accumulation, of this figure at the end of both speeches. It is

well-established in rhetoric that the closing paragraph of a speech plays an important role in affecting the audience. In most of the addresses we examined, speakers emphasized the concluding sentences, and the two Clinton speeches are good examples of this. As Trosborg observes [Trosborg, 2000], the typical rhetorical feature of Clinton’s 1993 speech is *iconic linkage*. What she is referring to is a repetitive pattern that attracts attention and joins the parts of the text. This linkage, established by excessive use of anaphora, is not only very characteristic in Clinton’s first speech, but in his second as well.

In terms of the other remaining figures, the 1993 Clinton speech seems to be more rhetorically developed. First of all, antimetaboles, polysyndetons and polyptotons tend to form groups, mostly in the first half of the speech, which sets up a certain rhythm in this part of the speech. Trosborg also highlights the dominant cohesive features of Clinton’s speech. These features concern the notions of *change* and *renewal*. These words in many variations occur quite often in the text, and are depicted by many polyptotons. Polyptotons are also present in Clinton’s 1997 address, but have a different nature. Only once does Clinton use *new*, *renew*, and interestingly he then refers to his previous (1993) inauguration speech. Overall, we can conclude that, although different in nature, rhetorically both speeches do not vary that much. The words used in individual figures of speech are different, but in general both addresses set a good rhetorical standard.

5 Conclusion

We have developed an annotation tool that enables both the manual and automated markup of rhetorical figures. This first step towards the automated annotation of figures of speech can be used to characterize and classify rhetorical patterning. Other tasks might involve the recognition of rhetorical strategies such as persuasion and argumentation. The system we have created combines the ancient theory of rhetoric with modern computational linguistic technology to facilitate more efficient and accurate identification of pragmatic evidence in natural language texts. In general, the automated annotation of figures provides a computationally efficient method of pragmatic analysis which can help in understanding how to speak and write effectively, compose messages in the most informative way, and engage audiences in the most appealing way.

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.

References

- [Corbett, 1990] E.P.J Corbett. *Classical Rhetoric for the Modern Student*. Oxford University Press, New York, 1990.
- [Gawryjolek, 2009] Jakub Gawryjolek. Automated annotation and visualization of rhetorical figures. Master's thesis, University of Waterloo, 2009.
- [Harris and DiMarco, 2009] Randy Harris and Chrysanne DiMarco. Constructing a rhetorical figuration ontology. Symposium on Persuasive Technology and Digital Behaviour Intervention (AISB 2009), Edinburgh, Scotland, 2009.
- [Trosborg, 2000] Anna Trosborg. *Analysing Professional Genres*. John Benjamins, 2000.