Rhetoric and the Digital Humanities

Edited by

JIM RIDOLFO AND WILLIAM HART-DAVIDSON

The University of Chicago Press Chicago and London
At the Digital Frontier of Rhetoric Studies: An Overview of Tools and Methods for Computer-Aided Textual Analysis

DAVID HOFFMAN
DON WAISANEN

Over the last few decades a sizable arsenal of “textual analysis” software has become available to scholars and researchers who work with language. NVivo, Wordstat, Linguistic Inquiry and Word Count (LIWC), DICTION, and Concordance are some examples of such textual analysis software packages. While this software has been most commonly used in the analysis of open-ended survey questions and transcribed interviews, it can also be used in the analysis of “naturally occurring” public discourse that has traditionally been the domain of rhetorical criticism, such as political speeches and pamphlets, newspaper editorials, and blogs. Although one software package, DICTION, is designed to be an aid to rhetorical criticism, on the whole computer-assisted analysis is a rare thing in the pages of mainstream rhetorical criticism journals such as the Quarterly Journal of Speech, the Rhetoric Society Quarterly, and Rhetoric and Public Affairs.

This chapter will explore the uses and limitations of textual analysis software in the criticism of contemporary
and historical rhetoric. Rather than trying to cover every existing software program—which would be a hopeless task, doomed to be dated before it was even printed—we discuss four broad functions that current programs perform and that future programs are likely to perform. We explore the capacities of the major textual analysis software packages, review published studies in which they have been used, and identify how these tools relate to other projects in the digital humanities writ large, with the goal of providing a number of suggestions concerning how textual analysis software might enhance rhetorical approaches to historical and contemporary public discourse.

Four Functions of Textual Analysis Programs

Textual analysis software comes in a variety of forms. Some software packages, like DICTION and Concordance, are built for very specific and limited purposes. Others, like NVivo and QDA Miner’s Wordstat, are multifunctional, seeking to be the only software packages that their users will ever need in the course of research. These large, multifunctional packages incorporate tools that are designed for qualitative data management and automated textual analysis. Although many researchers use programs like N-Vivo primarily for data management, we will not comment on such features as they fall outside the range of this chapter’s focus on using software for rhetorical criticism. We will also leave aside stylometric programs like Signature and the Java Graphical Authorship Attribution Program, which are designed primarily to determine the authorship of texts.

Our survey of extant textual analysis packages suggests that they have four broad functions. First, they can generate basic statistics about a text, such as word count, average sentence length, number of adjectives, the Gunning Fog Index (a basic measure of readability based on sentence length and number of complex words), and a host of others. Second, they can create indexes and concordances, quickly locating every instance of a word or word combination in a text or set of texts, cataloging and presenting them in context. Third, they can use dictionaries, either preprogrammed or user generated, to rate texts on a host of qualitative variables. DICTION, for instance, can score texts on their “certainty, activity, optimism, realism, and commonality” variables relative to other texts, using built-in dictionaries. Fourth, they can do cluster analyses, using sophisticated algorithms to determine the most important concepts in a given text or group of texts and how they are related to each other.

Basic Textual Analysis

A variety of automated rhetorical analysis tools and techniques, range from broad, global analysis to more focused, targeted searches and explorations. Such analysis can be performed using a wide variety of contemporary software packages. This chapter will briefly discuss the most common approaches to automated rhetorical analysis, and will provide a brief overview of the most popular packages available today.
they are related to each other. In the following sections we will discuss how each of these functions has been used in published work.

**Basic Textual Statistics**

A variety of basic computational outputs are available for computer-aided rhetorical studies, such as findings about the frequency of a given term or the average word length. Outputs of statistical information can range from charts or tables to visualizations that juxtapose smaller and larger terms on the basis of their frequency, as in the program Wordle. Figure 12.1 shows a Wordle “cloud” graphic derived from a simple word-frequency analysis of Lincoln’s Gettysburg Address. Throughout this chapter, we insert outputs derived from an analysis of Lincoln’s speeches to facilitate comparisons between textual analysis programs and their various functions.

Basic textual statistics make simple but substantiated generalizations. Such generalizations can have a large practical impact in research and scholarship. One example of how this can work comes from the field of classical rhetoric. The loose group of fifth-century BC Greek thinkers known as the Sophists has long been associated with the beginning of formal rhetorical studies. Although the term *sophist* suggests that they had a number of adherents, based on a host of others, they quickly locating each text or set of texts, they can use dictionaries to rate texts on a commonality of variables. Fourth, they can score texts on the basis of the relationship of texts and how...
has generally been used in a derisory sense for most of the last two and a half millennia, the surviving fragments of actual Sophists—such as Protagoras, Gorgias, Lysias, Antiphon, and Isocrates—reveal thinkers who were sophisticated (pun intended) and challenging, if sometimes highly skeptical about all truth claims. In the early 1980s, in works like G. B. Kerferd’s *The Sophistic Movement* (1981), efforts to rehabilitate the Sophists in their true character began in classical studies. Such efforts were also made by those who studied rhetoric in speech communication and English departments and who saw in the Sophists a glimmer of a “postmodern”-flavored alternative to the Aristotelian orientation that had dominated thought about classical rhetoric in the previous generation. In the *Philosophy and Rhetoric* article “Toward a Sophistic Definition of Rhetoric,” John Poulakos (1983) attempted to rethink the meaning of rhetoric from a sophistic perspective. Poulakos’s sophistic perspective on rhetoric was challenged by Edward Schiappa in a series of articles and chapters (Schiappa 1990a, 1990b, 1991a, 1991b, 1992), touching off a heated debate, and, ultimately, changing accepted ideas about the relation between rhetoric and the Sophists.

Underpinning Schiappa’s criticism of Poulakos’s work was the contention that the actual Greek word for *rhetoric* did not appear in any Greek text before Plato in the first half of the fourth century BC, considerably after the Sophists’ period. The claim was supported by a search of the *Thesaurus linguae graecae*, an early database of extant Greek literature. The debates touched off by Schiappa’s challenge are typical of the kind of disputes that follow from the type of generalization made possible by simple textual statistics. For example, simply because no surviving text contains the word *rhetoric*, does that mean that it was not used anywhere? Even if *rhetoric* was not in use as a term, does that mean that the concept of rhetoric was not in play? On the other hand, if, as Schiappa claimed, Plato actually coined the term *rhetoric* in the fourth century, is that not a meaningful moment in the history of rhetoric?

The next generation of scholars learned the lesson of how powerful claims backed by exhaustive database searches could be and integrated these techniques into their work using the next generation of textual analysis tools. The Perseus Digital Library is a full, but not exhaustive, online collection of ancient Greek and Roman texts, together with some other materials like artworks. The site has textual analysis tools with the capacity to produce statistics about word frequencies across all major classical authors. In past published work, Hoffman has used this function of the Perseus Digital Library to make claims about the evolution of some rhetoric, not just the original transformed, *logos*, which of words and meant “to be to what is existing can allow keywords be provided a tool to change over time.

Few, if any, they are accurate are often even verbs typically to computer a dictionary, ters long, th words,” amo and emotion

Indexes and

The creation of comprehensive alphabetical lists derives its place in a teor of usage” (M searches of e textual analysis manipulate cordance, fo cordances. (display (see)

Indexing managing k in the research
or most of the last two and of actual Sophists—such as Isocrates—reveal thinkers challenging, if sometimes early 1980s, in works like, efforts to rehabilitate the accal studies. Such efforts in speech communication in the Sophists a glimmer he Aristotelian orientational rhetoric in the previous article “Toward a Sophistic 3) attempted to rethink the effective. Poulakos’s sophistic Edward Schiappa in a series 990b, 1991a, 1991b, 1992), changing accepted ideas Sophists. Poulakos’s work was the rhetoric did not appear in of the fourth century BC, he claim was supported by an early database of extant by Schiappa’s challenge are from the type of generalizations. For example, simply be-rhetoric, does that mean that as not in use as a term, does it not in play? On the other hand, he coined the term rhetoric in all moment in the history of the lesson of how powerful these could be and integrated next generation of textual a full, but not exhaustive, roman texts, together with has textual analysis tools. word frequencies across all work, Hoffman has used this take claims about the evolution of some key terms in the technical vocabulary of ancient Greek rhetoric, namely, logos and eikos (see Hoffman 2003, 2008a, 2008b). Such work demonstrated that logos began its life as the verb legein, with the original sense of “to gather,” as in “to gather firewood,” and was transformed, largely in early philosophical discourse, into the noun logos, which can be understood to mean “a gathering or collection of words and phrases—a composition.” Eikos began life as a verb that meant “to be similar” and evolved into a participle that meant “similar to what is expected or likely.” “By-author” and “by-text” term frequencies can allow researchers in any field to see where and when particular keywords began to be used. The ability to locate all keyword usages provides a foundation for an analysis of how purposes and meanings change over time, to which informed speculation about the forces that drove changes in meaning can be added.

Few, if any, programs currently offer only basic textual statistics. But they are accessible in most textual analysis packages, and the basics are often even provided by word-processing programs. Textual statistics typically form one component of more sophisticated approaches to computer-aided textual analysis. LICW, in addition to employing a dictionary-based approach, uses basic counts of words over six letters long, the number of personal pronouns, and counts of “function words,” among other features, to make assessments of the complexity and emotional tone of a given text’s rhetoric.

Indexes and Concordances

The creation of indexes and concordances constitutes a second function of computer-aided textual analysis. Traditionally referring to alphabetical lists of key terms contained within a text, “a concordance derives its power for analysis from the fact that it allows us to see every place in a text where a particular word is used, and so to detect patterns of usage” (McCarty 2007, “The Concordance,” par. 2). Although Google searches of e-books can produce a very basic sort of concordance, some textual analysis software packages make it far easier to produce and manipulate concordances. The primary function of the program Concordance, for instance, is to index all words in texts and produce concordances. One output of such programs is a “keywords in context” display (see table 12.1).

Indexing and concordancing software may assist in finding and managing keywords in context but very much leaves these processes in the researcher’s hands. For instance, to group similar terms together,
potential search words typically need to be inputted manually. In constructing a concordance of key terms, such lemmatization describes an author’s grouping of certain close words under a single category, as in the words go and goes (McCarty 2007).

Researchers have used Concordance and similar programs to enhance content-analysis procedures by ensuring their accuracy and thoroughness and to cut down on the tedious human labor involved. Hansen and Benoit (2002) used Concordance to generate an exhaustive list of issues in all presidential television advertising between 1952 and 2000, as part of a project to track how closely issues in television ads are associated with public priorities (at least as revealed through opinion polls). Using this methodology, the researchers were able to produce a table comparing public priorities with the issues addressed in George H. W. Bush and Michael Dukakis’s campaign ads in 1988 (see table 12.1). Where the numbers in the column “Public” represent the percentage of the public that thought the issue was “important,” the numbers in the columns “Bush” and “Dukakis” represent the number of mentions of the issue in the advertising sample.

The methodology used to produce this table has some limitations, but it does open up some interesting avenues for “rhetoric and reality” types of studies. On the limitations side, the percentage of the public thinking an issue is important is not strictly comparable to the number of mentions by a candidate. Also, the way in which the advertising was sampled does not take into account that some ads about specific issues might have run more frequently on television than other ads. The first, but not the second, limitation is addressed when Hansen and Benoit move to a comparison of the rank ordering of priorities between candidates and the public. A similar use of Concordance in content analysis is made by Potnis (2010), who analyzed data from e-Government Readiness Assessments.

In general, Concordance is a powerful tool for finding discrete bits of content that can be manually classified by researchers. Yet dictionary-based textual need for the n

Dictionary Cc

Dictionary coined rhetoricaries to assess texts. The deg might be asse:tionary of “an degree any to causal words, .net/#, par. 1)’ own dictionari grams, see Per

One of the tioned above, ments of verb proximately ti software comp erages constru public discour in Verbal Style paign Talk (20r

Figure 12.2
be inputted manually. In contrast, lemmatization describes an output than can be produced by computer programs to ensure their accuracy and reduce human labor involved, further generating an exhaustion advertising between 1952 and 1954. Issues in television (at least as revealed through the researchers were able to) with the issues addressed in s's campaign ads in 1988 see column "Public" represent the degree issue was "important," the ukakis" represent the number of sample.

This table has some limitations, such as for "rhetoric and reality", the percentage of the public is comparable to the number of times mentioned above, which are original and specific to other ads. The first issue is that when Hansen and Benoit (2000) mentioned priorities between canonicity and concordance in content analysis, it was concordance data from e-Government tools for finding discrete bits of information. Yet dictionary-based textual analysis programs, treated in the next section, remove a need for the manual classification of content.

### Dictionary Comparisons

Dictionary comparison programs constitute a third area for computer-aided rhetorical criticism. These programs use preconstructed dictionaries to assess the degree to which particular qualities are present in texts. The degree to which a given text expresses anger, for instance, might be assessed on the basis of how many of its words match a dictionary of "angry words." LIWC allows researchers to determine "the degree any text uses positive or negative emotions, self-references, causal words, and 70 other language dimensions" (http://www.liwc.net/#, par. 1) while also allowing scholars the ability to construct their own dictionaries (for a detailed comparison of LIWC with other programs, see Pennebaker, Mehl and Niederhoffer 2003). One of the best examples of this type of software is DICTION, mentioned above, which uses preprogrammed dictionaries to make assessments of verbal style and linguistic habits (Hart 2001). Through approximately three decades of iterations and updates, current DICTION software compares a text or group of texts with thirty-five variable averages constructed from an extensive database of texts collected from public discourse. Vigorously championed by its creator, Roderick Hart, in *Verbal Style and the Presidency* (1984) and ensuing works like *Campaign Talk* (2000), DICTION has been employed in a large number of political discourse studies, including Ballotti and Kaid (2000), Crew and Lewis (2011), and Hart and Lind (2010). It has also been used to examine discursive trends in education (Graddy 2004), diplomacy (Bashor 2004), religion (Eldemüller 2002), management (Finkelstein 1997), and even stand-up comedy (Waisanen 2011a, 2011b, in press).

Table 12.2. 1988 Public policy priorities and issues addressed

<table>
<thead>
<tr>
<th>Issue</th>
<th>Public</th>
<th>Dukakis</th>
<th>Bush</th>
</tr>
</thead>
<tbody>
<tr>
<td>Budget deficit</td>
<td>12</td>
<td>12</td>
<td>9</td>
</tr>
<tr>
<td>Economy (general)</td>
<td>12</td>
<td>5</td>
<td>11</td>
</tr>
<tr>
<td>Drugs</td>
<td>11</td>
<td>26</td>
<td>13</td>
</tr>
<tr>
<td>Unemployment</td>
<td>9</td>
<td>30</td>
<td>13</td>
</tr>
<tr>
<td>War/international</td>
<td>9</td>
<td>26</td>
<td>54</td>
</tr>
<tr>
<td>Poverty</td>
<td>7</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Lincoln's First Inaugural, Gettysburg Address, and Second Inaugural using three of its five master variables: optimism, certainty, and commonality. According to DICTION, Lincoln gets steadily more optimistic as he moves through this set of speeches; however, his certainty peaks at Gettysburg even as he uses language with less commonality than in either of his inaugurals.

Through such outputs, dictionary programs provide researchers with the ability to do efficient comparative work both within and between texts. These programs make arithmetic judgments that can be a valuable supplement to contextualist claims (Hart 2000) and have an "ability to deal with great quantities of verbal information" that is useful in a "culture and era supersaturated with political messages" (Hart 1985, 103). But moving beyond—while also incorporating—basic statistical techniques, the comparisons between a text or texts and preloaded dictionaries move researchers from descriptive to more analytic and conceptual findings. By using dictionaries, one can gain an understanding of how common or deviant a text's language is in comparison with other texts, letting a researcher "quickly distinguish between idiosyncratic and normative behavior" (Hart and Lind 2010, 357).

To cite several telling examples of the types of claims that can be supported by dictionary comparisons, studies have showed how the presidential candidate Bob Dole's declaration of being the "most optimistic man in America" belied how he "used less verbal optimism in his campaign speeches than any Republican since Tom Dewey with one exception" (Hart 2000, 4). A speech by John F. Kennedy—which much of the American press described as quite typical—was actually "massively uncharacteristic": "Kennedy was much more pragmatic on
of optimism, and commonality. This occasion than he normally was" (Hart 1985, 124). Furthermore, a comparison of President Carter's inaugural address with his State of the Union speech showed that the speeches were in many respects not similar, contrary to press accounts (Hart 1984, 240–41).

Dictionary software searches for what it is designed to search for, no less and no more (Hart and Daughton 2005, 168). Although a wide variety of dictionary-based programs for textual analysis are available—Wordstat has a list of ten that can be imported, including the whole of LIWC—if the dictionaries are not designed to answer the questions a researcher is interested in, they may not be valuable. For instance, if a scholar wanted to get an analysis of how much medical terminology was used in a set of speeches, DICTION would not be helpful—unless a dictionary specific to that purpose were constructed by the researcher. At the same time, not all dictionaries may be equal in terms of their validity. Although established programs like LIWC and DICTION have gone through decades of refinement, researchers are well advised to scrutinize the lists of search terms employed by dictionary programs to make sure that they are suitable for their purposes. Overall, like the other functions outlined in this chapter, dictionary comparisons can be used as a central or merely supplemental method for rhetorical work.

Cluster Analyses

While dictionary-based programs can rate and compare texts with reference to various qualities like optimism and certainty, they do not have the capacity to tell researchers how the actual locations of various terms relate and link with other terms. For such an approach, one needs to turn to cluster analysis programs, which can provide both written and more advanced visual outputs. These programs, such as NVivo, Wordstat, T-Lab, and Catpac, have the capacity to discover relations between key terms within and across texts. Where Concordance can give one an instance-by-instance view of how a key term is used, cluster analysis gives the researcher a statistically driven overview of what terms tend to be collocated with each other throughout a text.

One major use of cluster analysis programs has been to conduct frame analyses across large sets of texts. Yan Tian and Concetta Stewart (2005) used Catpac to analyze 332 CNN and 408 BBC reports about the SARS outbreak in 2003. With the aid of Catpac, the researchers were able to identify seven clusters of associated terms in the CNN reports and six in the BBC reports. The tables in which they reported their results are reproduced in tables 12.3 and 12.4.
CHAPTER TWELVE

Table 12.3 Clusters from the CNN text

<table>
<thead>
<tr>
<th>Cluster number</th>
<th>Cluster theme</th>
<th>Keywords</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Beijing</td>
<td>Beijing, millions, last, and cases</td>
</tr>
<tr>
<td>2</td>
<td>Public health</td>
<td>Disease, health, SARS WHO, Hong Kong, people</td>
</tr>
<tr>
<td>3</td>
<td>Symptoms, statistics, and effects on travel</td>
<td>Symptoms, patient, virus, outbreak, officials, reported, China, travel, number</td>
</tr>
<tr>
<td>4</td>
<td>Chinese government</td>
<td>Chinese, Taiwan, first, full, government</td>
</tr>
<tr>
<td>5</td>
<td>Toronto</td>
<td>Story, Toronto</td>
</tr>
<tr>
<td>6</td>
<td>Economic impact</td>
<td>Case, countries, death, hit, Yen, down, higher</td>
</tr>
<tr>
<td>7</td>
<td>Treatment and control</td>
<td>World, control, city, hospital, Singapore, infected, spread</td>
</tr>
</tbody>
</table>

Source: Tian and Stewart (2005), table 1.

Table 12.4 Clusters from the BBC text

<table>
<thead>
<tr>
<th>Cluster number</th>
<th>Cluster theme</th>
<th>Keywords</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Public health</td>
<td>Affected, public, country, death, authorities</td>
</tr>
<tr>
<td>2</td>
<td>World</td>
<td>Canada, Asia, world, countries</td>
</tr>
<tr>
<td>3</td>
<td>Singapore</td>
<td>Quarantine, Singapore, spread</td>
</tr>
<tr>
<td>4</td>
<td>China and Toronto</td>
<td>Beijing, China's, Chinese, government, reported, far, officials, patients, Toronto, city, hospital, number</td>
</tr>
<tr>
<td>5</td>
<td>Hong Kong and WHO</td>
<td>Cases, Hong Kong, people, WHO, SARS, virus</td>
</tr>
<tr>
<td>6</td>
<td>Outbreak and impacts</td>
<td>Disease, China, infected, illness, outbreak, first, health, died, travel</td>
</tr>
</tbody>
</table>

Source: Tian and Stewart (2005), table 2.

By comparing the two sets of clusters, the researchers were able to make a number of general observations about how the two networks framed the SARS outbreak. They observed that, while both networks focused on the impact of the outbreak on the travel industry, CNN reported on the situation in Taiwan far more frequently than did the BBC. Other instances of cluster analysis include Stewart, Gil-Egui, Tian, and Pileggi (2006) and Stephen (1999), who used the WordStat program in a similar manner.

Conclusion

Although the usefulness of textual analysis programs in rhetorical studies should be obvious from the examples discussed above, it is still relatively rare for those engaged in rhetorical criticism to make use of the tools they provide. There is, to our knowledge, at present no systematic treatment of conducted to intelligent cho

Although is great, cert noted. O ne I large numberings must con that you cou by itself, tell researcher co conservative tween the tw need to com in context is analysis in t

A second; allow same precedented meaning the metric progr to use certain the specific n. For example,

1. Once upon some young
2. In a far, far once upon :
3. Once upon far, far, awa

A human of the fi duction to a fairy tale tha traditional fs sentences co
the researchers were able to about how the two networks that, while both networks on the travel industry, CNN more frequently than did the include Stewart, Gil-Egui, Tian, used the WordStat program

is programs in rhetorical stud-discussed above, it is still rela-
l criticism to make use of the edge, at present no systematic

treatment of how to use these tools in rhetorical scholarship. We have conducted this analysis to aid researchers and students in making intelligent choices about the appropriate place of textual analysis software in their work.

Although the potential contribution of textual analysis programs is great, certain global limitations of the available processes should be noted. One limitation is that, although these programs can compare large numbers of texts, any capacity to understand context-related meanings must come from the human reader. For example, there is no program that you could plug a transcript of The Colbert Report into that could, by itself, tell you that Colbert is a parody of a conservative pundit. A researcher could conduct a comparison between Colbert and an actual conservative pundit that could identify some of the parodic features between the two—but the initial idea that a parody was intended would need to come from the researcher. This inability to understand texts in context is likely to remain a limitation of computer-assisted textual analysis in the foreseeable future.

A second global limitation of these programs is that, although they allow some of the semantic features of text to be analyzed at an unprecedented scale, they are unable to deal with features of style and meaning that depend on sentence-level syntax. Although some stylo-metric programs, such as Signature, can register an author’s tendency to use certain grammatical patterns, they cannot make any guess about the specific meanings that audiences may infer from such articulations. For example, consider the following three sentences:

1. Once upon a time, far, far, away, in a kingdom by the sea, there lived a handsome young prince.

2. In a far, far away kingdom by the sea, there lived a handsome young prince, once upon a time.

3. Once upon a kingdom, by the young prince, there lived a handsome time, in a far, far, away sea.

A human reader with sufficient cultural background can easily recognize the first sentence as a traditional, if somewhat clichéd, introduction to a fairy tale. The second sentence is the introduction to a fairy tale that is trying to be somewhat fresh while still alluding to the traditional formula. The third sentence is pure nonsense. Yet all three sentences contain exactly the same words and are grammatically cor-
rect. They differ only in syntax. Dictionary and clustering programs would find the same keywords in roughly the same proximity, with basic textual statistics unaffected. We do not know of any program that can analyze sense and style at the sentence level in all their syntax-dependent glory. No program could pick up the difference between

Ask not what your country can do for you, but what you can do for your country and

Ask not what your country can do for you, but what your country can do for you.

Even in the area of semantics, where these programs generally excel, they can leave the researcher with a rather flattened view of the world. The statistical observation that Lincoln tends to use the words constitution and government in proximity to each other does not provide a “street-level” view of a passage like the following: “This country, with its institutions, belongs to the people who inhabit it. Whenever they shall grow weary of the existing Government, they can exercise their constitutional right of amending it, or their revolutionary right to dismember or overthrow it” (Lincoln, First Inaugural). In computer-aided textual analysis one is given a view equivalent to that of a reconnaissance satellite that strains to make out any features that are smaller than a meter in size—useful for detecting large-scale patterns, but not the proper instrument to photograph a fine painting. These tools are emphatically not a substitute for close reading.

Although these programs might be characterized as less than intelligent in some ways, they are superhuman in others. This kind of work sits on a nexus between quantitative and qualitative inquiry, holding the possibility of both bridging methodological divides and adding to researchers’ repertoires. Overall, we find that computer-aided textual analysis can contribute to the study of rhetoric in at least three ways. (1) Programs that produce textual statistics and concordances can help us map the synchronic and diachronic distribution of ideographs and memes in a way that would fulfill many of the aspirations of Michael McGee’s (1980) seminal article “The ‘Ideograph,’” which were hitherto unrealizable. Huge numbers of texts can be analyzed to discover how the meaning and usage of key terms like liberalism and conservatism have changed over time and how catchphrases and choice bits of jargon—paradigm shift, for instance—have spread. (2) Programs like Diction and LIWC can do objective and systematic comparisons of
try and clustering programs yield the same proximity, with no knowledge of any program that ice level in all their syntax, p the difference between

at you can do for your country

at your country can do for you.

These programs generally extract a rather flattened view of the icon tends to use the words to each other does not propose the following: "This country, le who inhabit it. Whenever vermin, they can exercise or their revolutionary right first Inaugural). In computer-equivalent to that of a recognizable features that are smaller; large-scale patterns, but not the painting. These tools are being characterized as less than intelligent in others. This kind of work qualitative inquiry, holding logical divides and adding to that computer-aided textual analysis of rhetoric in at least three statistics and concordances chronic distribution of ideological fulfill many of the aspirations cle "The 'Ideograph," which is of texts can be analyzed to key terms like liberalism and how catchphrases and choice have spread. (2) Programs and systematic comparisons of broad features of textual style and tone. (3) Cluster analyses are useful in large-scale studies of framing.

These tools for textual analysis might be useful across a broader range of digital humanities efforts than are covered in detail here. Dictionary-based software might complement efforts to map the "literary genome" that result in charts showing stylistic and thematic similarities (Jockers 2012). Rather than generating information about the overall relatedness of works, dictionary-based textual analysis could show how works in a particular literary corpus stack up against each other in terms of specific qualities: optimism, introversion, passion, and others. This would enable researchers to ask questions like, Who was the angriest author of the nineteenth century? in addition to, Which author is most typical of nineteenth century literature?

Furthermore, there is no reason that such techniques should be confined to literature. They could be equally well applied in philosophy and history or in more professional applications. Systematic key-term analysis aided by programs like Concordance could provide further depth by showing how particular concepts—whether key terms of a disciplinary lexicon or the "ideography" of a political speech—have shifted in meaning according to the time and place in which they were used. Finally, cluster analysis techniques can point out recurring sets of terms in literature, philosophy, and history just as well as in contemporary public discourse. The Department of Homeland Security even recently began a program to use several textual analysis techniques, including LIWC, to identify which terrorist groups are the most dangerous (Dempsey 2011). There is no doubt that they might also be used for happier purposes.

It should be remembered that one of the most ardent humanities scholars, Kenneth Burke, assigned a role for statistical procedures and associational clusterings in works of rhetorical criticism. He admonished critics to trace words and images across a range of works and to look for "what goes with what" and "what is vs. what" (1974, 20, 69). While the scholar's ability to read and reflexively interpret texts from a close perspective should always remain central to rhetorical studies, it is also clear from our foregoing survey that computers can do much to supplement and forge new paths of inquiry for twenty-first-century discourse analysis.

References


