Religious Rhetoric in US Presidential Inaugural Addresses
A Meta-analysis

Joshua Fischer

Abstract

The United States political rhetoric is littered with biblical roots despite its supposed separation of Church and State. In order to mediate the cultural divide between ecclesiastic and humanistic sectors of society, political orators utilize religious diction to draw in a wider range of constituents. The paper maps the use of 90 individual words over 57 inaugural addresses to produce a visual map of this information. Aside from interpreting and providing the graphs used in the paper, there is also a section on how the study may be replicated for further expansion. The paper pays specific attention to the words “God”, “Spirit”, “Providence” and “Heaven” in order to show their use in attracting voters during years of low turnout.

The premise of a separation between church and state in US governmental affairs is emblematic of the freedoms asserted by this form of democracy, protecting religious alignment from state scrutiny and mandate. As a nation primarily founded by Calvinist refugees of religious persecution, however, the United States owes much of its identity to ideas descended from Protestant theology, and although explicit issues of religion are often avoided by the electorate, religious rhetoric is ubiquitous in its writings. This is especially true of those documents that are intended for public dissemination, where the demands of moral identification with one's constituents initiate sermonesque undertones upon social and political statements made by leaders. By examining a subset of these writings, US Presidential Inaugural Addresses, a historically inclusive and periodic diagram of the constancy of this rhetoric can be inferred. Via contemporary methods of computerized data analysis, the precise extent of the use of religious referents in these documents can be quantified and organized according to their position and frequency on an historical scale. Not only do these data provide evidence of the plenitude of religious rhetoric in
political writings in general, they also enable comparative appraisal of individual speeches in this context and, more specifically still, isolate religiously connotative terms across the corpus of Inaugural Addresses, simplifying the process of identifying when some terms fell out of favor and when others began to enter the political lexicon.

The method for compiling the data used in this analysis began with my discovery of The American Presidency Project, an online database (www.presidency.ucsb.edu) created and curated by Citrus College Professor Gerhard Peters and UCSB History Professor John Woolley and dedicated to preserving, in electronic format, all extant US Presidential documents. In all, 57 speeches were processed, from George Washington's First Inaugural Address, delivered April 30, 1789 to Barack Obama's Second Inaugural Address, delivered January 21, 2013. Each speech was individually copied from its web page and pasted into a Microsoft Word document before being converted to Rich Text Format, or .rtf, for processing by open-source text recognition software. Each reformatted document was proofread for stray characters and accuracy before moving to the first stage of analysis.

AntConc is a free, downloadable concordance program that employs text recognition software to quickly compile word lists by frequency and display their locations individually within a selected corpus. The program can also locate and display morphemes or portions of words, up to combinations of words and even phrases, along with the co-occurrence of proximal words, for basic yet powerful text analysis of large corpora. From the generated list of nearly ten thousand individual words, those that held religious connotation and appeared at least twice were selected, for a total of 90 words, combinatorially occurring 949 times. Though this may seem insignificant when compared to a calculated total of over 138,000 word "tokens" (individual occurrences), it is important to keep in mind that a word like "God," which occurs over 100 times, averages nearly 2 appearances per speech (though the actual distribution of the word is quite different; more on that later). This was preliminary evidence that corroborated a sense that religious rhetoric permeated this type of discourse. The word "God" figures prominently on the selected list, not only for being second in frequency only to "Spirit," but also because a term so religiously explicit occurs so often in the data. Even at this early stage, one can quickly surmise that the separation of church and state in political speech is tenuous at best, and that where social and political mores intersect, elected leaders employ theological
tools of argument liberally, calling upon both subtle and overt religious inferences to indoctrinate their supporters and beatify their platform.

In order to see precisely how often the use of religious reference occurs, each of the 90 words were individually tracked across the 57 speeches using AntConc and the resulting data were entered into a computer spreadsheet. Though this was a tedious and time consuming process, once this data set was compiled, the spreadsheet software's functionality made it easy to compare the religious content of each speech against one another. Clearly, the early indications were accurate: with only one exception, every Inaugural Address contains at least four of the target words, to a maximum of 57 (William H. Harrison, 1841) and an overall average of 16.6. Interestingly, George Washington's precedent-setting Inaugural Speech of 1789 squarely meets this average with 17 words, while his second, a mere 134 words long, is the lone exception. Even Thomas Jefferson, perhaps the most vociferous proponent among the Founding Fathers of the separation of church and state, speaks of "Angels" and "Providence" in his two Addresses, setting the stage, along with Washington, John Adams, James Madison and other early leaders, for the Inaugural Addresses of their political progeny.

There are several reasons why rhetoric of this sort is not only used but also effective in this arena. The influential power of strong oration is alive and well in pulpits and altars across the world, and the human evolution of the sociopolitical paradigm from divine, ordained rulers to elected officials preserves in the latter many of the traditional features of the former, including discourse that elevates
the official and his platform to an honored, quasi-sanctified rank. In addition, such discourse, according to Christopher Chapp's Religious Rhetoric and American Politics, is emotionally evocative as well as fundamental to ideas of identity, making it a powerful tool in the hands of charismatic and influential persons: "American political rhetoric has been saturated with claims of collective religious identity...[which] is almost always multidenominational and aims to infuse the founding of America with collective religious purpose, and no single candidate, party, or interest group has had a monopoly on the rhetorical use of collective religious identity to build movement solidarity and assert a sense of divine purpose" (Chapp, 23). These speeches evoke, according to Chapp, a civil religion, a "broad religious identity [that] unites virtually the entire nation" and are "a solution to the challenge of appealing to a religious constituency that is both committed and diverse" (Chapp, 7). In a broad sense, then, religious rhetoric in electorate discourse is a unifying and galvanizing implement for the amalgamation of otherwise disparate American interests.

Though target words occur in nearly every Inaugural Address, there is wide variation in total target word occurrence on an individual, speech-by-speech basis. Moreover, as the graph above illustrates, there is a vaguely cyclical pattern to the target words' frequency that is more pronounced in the earlier years and levels off somewhat in the last 10 speeches, or past 40 years. William H Harrison's 1841 Inaugural Address has the most occurrences, though that may be attributable to the fact that, with over 8,000 words, it is the longest Inaugural Address to date. Examining the total word count of all speeches, Address length plainly plays a role in Harrison's speech's high number of target words; the others, however, do not necessarily mimic this correspondence. Take, for example, William Taft's 1909 Inaugural Address with 5,426 total words but only 5 occurrences of target words. Though a pattern is difficult to detect, it is interesting to note that Addresses with markedly higher target word frequency tend to follow one or more consecutive Addresses with much lower-than-average target word frequency, as if a periodic and strong re-establishment of religious conviction is necessary following periods of low observance.

Having quantitatively confirmed the initial supposition that religious referent is a conventional device in US Presidential Inaugural Addresses, we will now turn our attention toward some behaviors of the target words themselves. With 90 words in the target group, a complete analysis is both beyond the scope of this project as well as potentially inappropriately skewing for its purposes; it should, however, suffice to examine a handful of words that adhere to certain general
patterns of frequency. Of these, words that have either fallen out of the lexicon, appeared only in later Addresses, or periodically recur will be given attention.

The use of the word "God" is an appropriate starting point for several reasons, not least of which is, as mentioned earlier, its frequency and explicit religious connotation. Its frequent appearance in Inaugural Addresses contradicts the overarching premise of separating church and state affairs; moreover, its highly explicit religious significance firmly entrenches religious rhetoric as a viable if not necessary component of these types of documents. Note, however, that the word "God," with the one exception of James Monroe's Second Inaugural Address from 1821, does not appearing the first 16 speeches. It is only with Franklin Pierce's Inaugural Address of 1853 that the word begins to occur with any regularity. Monroe concludes his second Address with these words: "...with a firm reliance on the protection of Almighty God, I shall forthwith commence the duties of the high trust to which you have called me." Compare this to Pierce's Address, which states: "It must be felt that there is no national security but in the nation's humble, acknowledged dependence upon God and His overruling providence." Plainly, a difference in semantics exists between the two. The tone of Monroe's statement suggests that God is mentioned in a sort of prayer of thanks. By contrast, and drawing again from Chapp's analysis, the development of a civil religion can be ascertained from Pierce's words. God is evoked, not as a personal supplication, but rather as a unifying force that protects the nation's citizenry as a whole. It is this shift in semantic that allows the use of explicit religious terms in an allegedly nonreligious forum.
Perhaps for the same reasons that "God" came to prominence, some target words, following early popularity, have faded into obscurity among Inaugural Addresses. Antiquated terms like "Beneficent" understandably occur less frequently as they are dropped from the common lexicon, but words like "Providence" and "Heaven," still in regular use today in the religious arena, also have all but vanished from US Inaugural Speeches. Like "God," these words are explicit in their religious connotation but have fallen out of favor. The reason for this may be that, unlike "God," these words are more resistant to extrication from their original application; it is more difficult to adjust the semantic nuance and place these words within the framework of Chapp's civil religion. Reframed, it seems as if "God" is inherently inclusive while "Heaven" and "Providence" are exclusive. The idea of God is made available to all who seek it, while Heaven and Providence are reserved for a select, devout few. Any divisiveness in the rhetoric of one who seeks to represent an entire country or society must be abandoned.
Though initially intended to be an analysis of specific trends in presidential discourse using digital text analysis tools, this project consumed a considerable amount of time in the aggregation of data and subsequent visualizations thereof. The rest of this analysis will discuss the process by which US Presidential Inaugural Addresses were collated, quantized and reorganized. In exploring textual data manipulation and graphical visualization at the present state of technology, a logical and linear approach seems of utmost importance, for the software and processing engines that perform these tasks are largely nonintuitive and omnidirectional. The relative infancy of the development of these tools holds promise that more effective systems will be devised in the future, but for now the bulk of text visualizations tend to be either generally uninformative on the one hand or downright mystifying on the other; middling examples that exhibit the opposites of both are rare indeed. Many eschew informativity for aestheticism, and a sense that the source text has been somehow reduced often invades an image's premise. In this light, it may be useful to consider the way a qualitatively regarded text is first rendered quantitatively before being restructured into what is (hoped) a new, remediated, qualitative item. The crucial step here is the reorganization of narrative into data, or the conversion of qualitative into quantitative. This in many ways signals the destruction of the original document, yet it is through this destruction or better yet dissection that we hope to create new methods of literary analysis via the translational medium of graphical visualization.

The architectures of visualization engines tend to run toward the extremes of a continuum; one end is highly specific yet equally highly restrictive (low interaction) while the other is more "open format," nonspecific with seemingly endless variations and little in the way of guidance or instruction (high interaction). Take, as examples, a word cloud generator (like the ones found on wordle.net or the ManyEyes website) and Gephi, the free visualization software
downloadable from gephi.org. Word clouds represent the "low interaction" end of the spectrum: data entered is organized in only one way (by frequency) which is correspondingly represented visually in only one way (by size). Gephi, by contrast, offers many ways to organize and represent data. The spreadsheet format of the data set promises variety by supporting many subclassifications per item and the several visualization engines provide vivid animations that play out the software's algorithms (themselves adjustable) on the computer screen like miniature, 2-dimensional Disneyland light parades. This "high interaction" platform gives the user plenty of control, but like the word cloud, there is precious little information organized around traditional literary analyses. Sheer word frequency alone would have difficulty representing something like theme, and subtext would be lost altogether. Similarly, without an accompanying explanation, map, or legend, Gephi visualizations are little more than interesting (some, admittedly, beautiful) yet ultimately meaningless and undirected computer-generated images of lines, colors and shapes. In short, the divide between literary criticism and computer-generated text visualizations is still wide and in many ways unspanned; the process by which a text is separated, quantified and then restructured as a qualitative image for dissemination as critical analysis is tedious, uncertain and generally untested.

A strong logical and/or linear basis to visualizations of this sort, then, is demanded by those who wish to produce an item that serves to enhance, inflect, or expose an aspect of the targeted text corpus (or corpora) that may otherwise be obscured by, or difficult to prize away from, the corpus itself. Because a word cloud visualization is restricted to just one approach (word frequency), its application as a tool of critical analysis is severely limited. Instead, this discussion will focus on the "high interaction" platforms, specifically, Gephi. Gephi allows the user to organize their data in any way they choose under two main type headings: nodes and edges. The sizes, shapes and colors of the nodes can be adjusted, as can the edges' thicknesses, colors and lengths, to represent whatever the user desires. Though this open and option-laden format suggests a promising framework for creative exemplification, by lacking any specific organizing "templates," Gephi underscores the very issue plaguing all text visualizations, including word clouds, to wit, the remediation of textual criticism into an image is by no means simple nor obvious. It thus becomes crucial to establish a regular and familiar pattern of recognition within the visualization from the outset. This logical foundation is a tether, establishing the terra firma upon which an observer must stand in order to make any sense of the text-to-image representation before him.
Because analysis visualizations are direct descendants of graphs and charts, one of the clearest such bases is the historical paradigm. Our passage through time is unquestionably linear and the universally accepted concept is easily graspable and relatable. Graphical representations of this sort are an excellent way of analyzing past trends and helpful for predicting future ones, but whether this is a productive approach to studying texts is less clear. One successful approach to the marriage of text and graphical representation has been via the creation of a network. This network may revolve around the characters in a story or an item or artifact that recurs throughout a narrative. The approach to be described in this paper is a combination of the two. The selection of the corpora of US Presidential Inaugural Speeches provides a historical context that extends over 200 years into the past and the network is created by individual, repeated words; in this case, words that are religious in definition or connote religious rhetoric. The overall aim of the visualization is to expose religious rhetoric as a networked form in inaugural speeches and chart the frequencies of specific words used in this context through the history of the United States government. This visualization was predicted to identify (1) religious rhetoric as pervasive despite a purported separation of church and state, (2) speeches that tended to rely more or less heavily on religious rhetoric, and (3) the tenor of religious rhetoric (with specific regard to word choice over the years) as subject to change, that words will lose and gain favor, some will disappear altogether and some will appear only in later years as trends and oratory styles change. In this regard, the visualization is not intended to be the end of the exercise, either, but also a tool with which a researcher may quickly and precisely track the occurrence of a given word to a corresponding set of speeches and then, returing to the speeches themselves, analyze how that word's usage may have erupted, changed or faded away entirely, perhaps, ultimately, providing a reflection on the way American sociopolitical sentiment (and its accompanying religious undertone) has evolved.

In order to create a text visualization from scratch with Gephi, one must already possess some proficiency with text recognition and analysis software, as well as some familiarity with computer spreadsheets. So-called text analysis (data mining) software such as AntConc, a concordance tool available for free online, does the preliminary job of converting the target text from qualitative speeches to quantitative groupings of words. This is a fairly simple process, consisting of locating a digitized version of the target texts, copying and pasting them into Microsoft Word documents, deleting any extraneous characters, and finally converting them to .txt format to be read by AntConc. Finding a digital
rendering of a text can be difficult, especially in older documents; in my case this problem evaporated when I discovered The American Presidency Project (www.presidency.ucsb.edu), a public website co-created by UCSB History Professor John Wooley that seeks to archive every American Presidential document in digital format. Though it does take some time to collate and prepare the 57 individual inaugural speeches, delivered once every four years beginning in 1789, the concordance processor itself works extremely quickly and it is interesting to note how much easier it is to break down a body of text than to build one up. The concordance software was used to create a hierarchical word list arranged by frequency; going down the list, a sublist of words with religious inference that appeared two or more times was collected. The result was a list of 90 words, ranging in frequency from twice to, in the case of the word "spirit," 138 individual occurrences. Once this was completed, the difficult and time devouring task of creating a spreadsheet that mapped each word's occurrence to its respective speech and calculated totals in each direction (i.e., number of words per speech and total number of each word across all speeches) was undertaken. This was done using Microsoft Excel, a powerful computer spreadsheet platform that nevertheless presented its own challenges when used to provide the data sets read by Gephi.

Microsoft Excel employs a proprietary file format (delineated as .xls or .xlsx) that is not always recognized by other software platforms. In the case of Gephi, Excel files must be converted to Comma Separated Values (.csv), a relatively simpler format (similar to .txt), before being imported. Moreover, the original spreadsheet data set, which plots rows of individuated words against columns of individual speeches (or vice-versa) must be reorganized in a way that is appropriately recognizable by Gephi, i.e. the "nodes" and "edges" format. Individual attributes that differentiate the nodes and edges from each other also should be carefully labeled as such in the new spreadsheets. Though this information can be individually entered directly into Gephi one at a time, the formatting can be inconsistent, and a robust and reliable program like Excel is better for "cut and paste" features and consistent saving of files in progress. Using the original spreadsheet data set, two new spreadsheets were created. The first, labeled "Nodes," consisted of one entry for every word and one entry for every speech, leaving a total of 147 separate entries. Edges, which illustrate how many of each word appear in which speeches, comprised the second new spreadsheet and contained 612 separate entries, corresponding to one edge for every time a targeted word appeared in a speech.
Not surprisingly, the first few trial runs at visualizing the data sets produced an immense, complicated web. Though the data was faithfully reproduced as entered, there was simply no way to tell what was being communicated. The intricacy of the web produced an aesthetically challenging and interesting image that merely suggested rather than declared the presence of scholarship and critical analysis. Both aesthetic and analysis were crucial to the proper execution of this exercise, and it was clear that more work was necessary to achieve this. Returning to the two new spreadsheets, more information was added to further differentiate the entries. For the nodes, each entry was given an ID number, a Label (for speeches, the year the speech was delivered; for words, the word itself), a Weight value (for speeches, the total number of words; for words, the total number of occurrences), and an Attribute (for speeches, the Attribute differentiated between first and reelected inaugural speeches; for words, the Attribute was simply "Term"). Similarly, the edges spreadsheet was also expanded with a Weight column, which corresponded to the number of times a word appeared in a given speech. After uploading the improved spreadsheets into Gephi, nodes were marked with different colors according to the Attributes column and different sizes according to the Weight column. For the edges, line thicknesses were marked to vary according to the Weight column. When a visualization was initiated, what appeared on the screen was another complex and tangled image, much more vivid than the previous one. Clearly the aesthetic had taken a step forward, but what about the analysis? Toggling on the Labels display showed that, despite the wealth of information, there was still very little to guide the viewer through it. (Figure 1) Using the several layout options available tended to either collapse or expand the graph; none seemed useful for organizing the data. The nagging problem of aesthetic over analysis seemed in danger of nullifying many hours of work to produce something of substance. Something was needed to ground the visualization in a realm that was familiar to an observer viewing an unfamiliar rendition of text. It was at this point realized that, when faced with a large data set of information (especially textual information) converted to graphical form, the alienation from original context that accompanies the rendered data must be held together by a logical and/or linear arrangement that is relatable across both media. The choice here was between the words (perhaps alphabetically) and the speeches, arranged historically. The latter was chosen, given that the original intent of the project was to take a historical approach to an analysis of the frequency and type of religious rhetoric used. 57 new edges were added, one from each speech to the next in order of year of delivery. The weight of these edges was made consistent and higher than any other edge weight value, creating a highly visible
"backbone" to the visualization. This modification, along with a few hastily learned techniques applied to the graph layout feature called Force Atlas, led to the final image seen in Figure 2, below.

The graphical visualization in Figure 2, while still far from a satisfactorily informative image, nevertheless begins the task of pictorially uncovering qualitative information concerning religious rhetoric in US Presidential inaugural speeches that may have been previously undetectable or, at least, much less obvious, in the collected speeches themselves. It is something of a marvel that texts can be qualitatively disemboweled, then quantitatively collated and finally reconfigured and remediated in order to produce a qualitatively cogent and relevant consideration of the original, source text. Doubtless, the day will come when generating an informative graphical representation of any text requires nothing more than uploading the document and selecting a few processing protocols and visualization styles from a drop down menu, everything automated, simplified and clear. At this point in the technology, however, problems and complications abound, as attested by the many long hours spent troubleshooting, tweaking, and fine-tuning the visualizations on the following pages. And though many of these were eventually surmounted, the overarching question of whether these visualizations are ultimately reductive remains. Can a text that is broken down into its smallest recognizable bits and rebuilt, piece by piece, into a visual representation satisfactorily recreate or aptly represent the original text? At this stage, the answer seems to be no, for visualizations like the one discussed here cannot stand by themselves, without reference to their predecessor. Still, these visualizations, while not autonomous, can and do represent something, and as some argue, where there is a something, there is the potential for everything.
Figure 1
Figure 2

Words are marked with red icons. Speeches are blue; dark blue corresponds to first inaugural speeches, light blue to second (or more). Notice the most frequently used words that appear in the center and those that hover toward the ends of the historical range. Words that appear on the outer edges and cluster in a particular period of time draw that part of the range of speeches away from the center.
Works Cited
