Networks of the Great Awakenings: Classification of Puritan Sermons by Word Usage Statistics

Michał Choinski
michalchoinski@gmail.com
 Jagiellonian University, Poland

Jan Rybicki
jkrybicki@gmail.com
Jagiellonian University, Poland

Revivalism has always played a significant role in the social functioning of religion in the US. As argued by McClymond, "religious revivals are as American as baseball, blues music, and the stars and stripes" (2010, 306). This strong presence of revivalism in the American religious landscape translates into the considerable significance of revival preaching not only for the American pulpit practice, but also for American culture in general. One could argue that a certain continuity is discernible in the American revival tradition, and preachers of consecutive "Awakenings", starting with the Great Awakening, have utilized similar communicative strategies, analogous sets of cultural references, as well as persuasive ploys to forward the "New Birth" to their hearers and to spread the revival zeal. Billy Graham, the most celebrated televangelist of the 20th century, testified to the importance of this tradition when in 1949 in Los Angeles, during the "Canvas Cathedral" Crusade, he delivered to a contemporary audience the most notorious sermon of the Great Awakening (and, perhaps, of America's entire pulpit oratory), Jonathan Edwards's *Sinners in the Hands of an Angry God*. Graham was by no means a pioneer in this respect, as the 19th century American revivalists, like Charles Finney, eagerly fell back on the rhetorical heritage of the first Great Awakening sermons.

Crowds of thousands of people, the emotional reactions of the audiences which often bordered on mass hysteria, fervent theological debates and a surplus of publications played a significant role in the shaping of the early American ecclesiastical order. Similarly, the American rhetorical tradition was strongly informed by the revival developments in pulpit oratory, especially in the context of both the oratory of the American Revolution and the Civil War. New forms of preaching manifested the power of the spoken word, and propelled the dynamics of public debate in a period which was to prove vital for the development of American identity.

The study of American preaching tradition from the diachronic perspective seems particularly important. Different groups of preachers in consecutive Great Awakenings, (First 1735-1750, Second 1790-1840, Third 1850-1900, Fourth 1960-1980) appropriated the rhetorical models employed by the previous generations and built upon their output.

However, because of the sheer size of the available corpus, the comparative study of the preachers of all Great Awakening has been so far impossible. Stylometry based on word usage makes it possible to highlight the connections between particular groups of preachers, as well as to demonstrate the evolution of the American revival preaching tradition; and to confront these findings with the existing attempts at classification/chronology (Stout 1986).

In our study, 42 collections of sermons by individual preachers (one text file per preacher) have been collected, digitized and modernized to avoid an excessive chronological bias due to spelling differences; of course, purely linguistic bias could not be entirely eliminated. Of these, 13 are traditionally classified as First Awakening; 9 as Second; 19 as Third; the most traditionally controversial group, Fourth, is represented by Billy Graham alone.

Classification was made by comparing distances, or differences, between most-frequent-word frequencies of the texts using Burrows's Delta procedure (2002); the distance measure applied was "Cosine Delta" as proposed by Jannidis et al. (2015), implemented in the *stylo* package (Eder et al. 2015, 2016) for R (). The word frequencies were submitted to a consensus procedure of cluster analysis at word frequency vectors of 100 to 2000 most frequent words, and these results then served to produce network diagrams in *Gephi* (Bastian et al. 2009) using the gravitational Force Atlas 2 algorithm (Jacomy et al. 2014). This produces a network or a "map" of data points for the individual preachers' sermons; the closer and the thicker the links between them, the more similar they are.

Figure 1 presents such a network graph for the 42 preachers. The color coding follows the traditional division into four "Awakenings" (from First, green, through Second, yellow, Third, red, and Fourth, purple). A very strong evolutionary pattern emerges;

apart from minor imperfections, the network shows a clear evolution from the earliest preachers to the most modern one, Graham. At the same time, the grouping of the data points suggest that a different classification might also exist in the dataset. This is why another algorithm available in *Gephi*, "Modularity," was used to discover these "communities," or groups (Blondel et al. 2008).
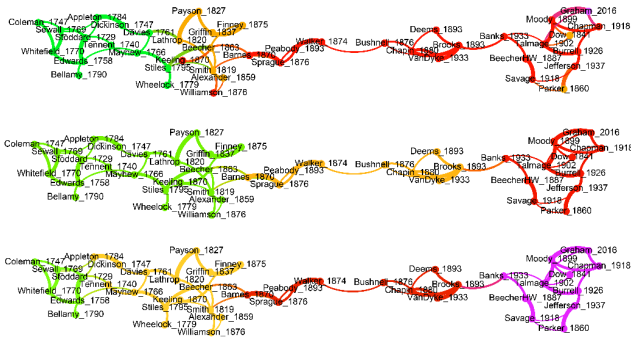

Figure 1. The four Great Awakenings in traditional classification (top); divided by modularity into 3 (center) and 4 (bottom) groups.

This produces two alternatives to the traditional division of the Revivalist movements(s). In the first of these (center), what is usually referred to as the First and the Second Great Awakening would be merged into a single group, while the traditional Third Awakening splits into early and late phases; the latter of which now also included the single representative of the Fourth. Perhaps more interestingly, computer-generated four Awakening communities (bottom) suggest a First only limited to some of Edwards's contemporaries, a Second that extends a little more than traditionally into the past, and again a limited early Third with an expanded Fourth.

This is of course not to say that the above results invalidate the traditional, or historical, division of American revivalist writing. But the stylistic (or, at least, stylometric) divisions are no less valid. Most-frequent-word usage is a significant element of distant reading; and such a distant reading of the Revivalism opens new avenues for close reading of the American homiletic tradition.

Acknowledgements

Bibliography

Bastian, M., Heymann, S., Jacomy, M. (2009). *Gephi: an open source software for exploring and manipulating networks*. International AAAI Conference on Weblogs and Social Media.

Blondel, V., Guillaume J.-L., Lambiotte, R., Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* (10), P10008 (12pp) doi: 10.1088/1742-5468/2008/10/P10008.

Burrows, J. (2002). 'Delta': A measure of stylistic difference and a guide to likely authorship. *Literary and Linguistic Computing*, 17: 267-287.

Eder, M., Kestemont, M. and Rybicki, J. (2013). Stylometry with R: a suite of tools. *Digital Humanities 2013: Conference abstracts*, University of Nebraska-Lincoln, 487-489.

Eder, M., Rybicki, J., Kestemont, M. (2016). "Stylometry with R: A Package for Computational Text Analysis," *R Journal* 8 (1): 107-121.

Jacomy, M., Venturini, T., Heymann, S. and Bastian, M. (2014). ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software. *PLoS ONE*, 9(6): e98679. doi:10.1371/journal.pone.0098679.

McClymond, M. (2010). Revivals, in P. Goff (ed.), *The Blackwell Companion to Religion in America*, Chichester: Wiley-Blackwell.

R Core Team (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Wien, http://www.R-project.org/.

Stout, S. (1986). *The New England Soul. Preaching and Religious Culture in Colonial New England*. New York: Oxford University Press.