

City University of New York (CUNY)

CUNY Academic Works

Dissertations, Theses, and Capstone Projects

CUNY Graduate Center

6-2020

Tracking Keywords in American Studies

Thomas Cleary

The Graduate Center, City University of New York

[How does access to this work benefit you? Let us know!](#)

More information about this work at: https://academicworks.cuny.edu/gc_etds/3778

Discover additional works at: <https://academicworks.cuny.edu>

This work is made publicly available by the City University of New York (CUNY).
Contact: AcademicWorks@cuny.edu

TRACKING KEYWORDS IN AMERICAN STUDIES

by

THOMAS J. CLEARY

A master's capstone project submitted to the Graduate Faculty in Liberal Studies in partial

fulfillment of the requirements for the degree of Master of Arts,

The City University of New York

2020

Copyright 2020

by Thomas J. Cleary

This work is licensed under a

Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

Tracking Keywords in American Studies

by

Thomas Cleary

This manuscript has been read and accepted for the Graduate Faculty in Liberal Studies in satisfaction of the capstone project requirement for the degree of Master of Arts

Date

Justin Rogers-Cooper

Capstone Project Advisor

Date

Elizabeth Macaulay-Lewis

Executive Officer

Abstract

Tracking Keywords in American Studies

by

Thomas J. Cleary

Advisor: Justin Rogers-Cooper

Tracking Keywords in American Studies is a computational text analysis project that provides interactive data visualizations which allow for deeper analysis of issues and discourses in the American Studies journal *American Quarterly*. Taking inspiration from *Keywords for American Cultural Studies*, which provides a list of keywords to serve as an entry point into understanding American Studies, my project takes these keywords and plots their occurrence throughout *American Quarterly*. Complimentary to the word counts, I chart algorithmically generated topics produced using topic modeling to look at discourses in the journal. This offers a different perspective on American Studies than what is provided through *Keywords for American Cultural Studies*, showing topics which might not be represented in the keywords. Like *Keywords for American Cultural Studies*, the project is for classroom use, offering an introduction to computational text analysis. It uses data as a starting point to critique issues and disciplines in American Studies. For example, my analysis emphasizes and shows a correlation between the keywords “White” and “Race,” and show that whiteness is the center of conversations about race in *American Quarterly*. I point to how trends in the keyword chart show an increasing interdisciplinary influence in American Studies, too, which supports

anecdotal impressions of the field. Further, I propose the creation of “Art” and “Education” as new keywords based on statistical evidence from the topic models.

Table of Contents

List of Figures	vii
Project Manifest	viii
Introduction	1
Context on Field Formation	3
Keyword Background	5
Methods - Overview	11
Methods - Data Analysis	15
Development - Data Collection & Transformation	17
Development - Data Visualization	26
Findings	28
Future of the Project	33
Works Cited	36

List of Figures

1. Flowchart describing data collection, processing and transformation. 18
2. A snapshot of the keyword count spreadsheet that collects how many times each keyword is present in a single document. The COUNTIF() function is shown in the cell. 22
3. A snapshot of the keyword count spreadsheets that shows how many times a keyword is present in a given year. The SUMIF() function is shown in the cell. 22
4. An example of a MALLET keys file. First column is the key number, second is the topic number added in myself, third is the weight of each topic in all of *American Quarterly*, and the fourth are the words that comprise each topic. 24
5. Example of a MALLET composition file. The second column is the filename and the following ones are the weight of each topic within the file and correspond with the keys file. Headers are my own. 25
6. A snapshot of how the topic model CSV is organized. 26
7. Chart showing keyword counts for Asian (brown), Black (pink), Latino (cyan), Race (blue), and White (purple). Notice the dip in Black as Asian and Latino rise after 2010. 30
8. Topic model chart showing “art artists modernism”(pink) and “art painting paintings” (grey). 31
9. Topic modeling showing “school students education” (grey), “university press state” (red), “american studies scholars” (green). 32
10. Topic model chart showing “time fact point” (cyan) and “cultural american discourse” (gold). 33

Project Manifest

1. Whitepaper (.PDF File)
Tracking Keywords in American Studies. Includes background and project summary.
2. Data sets (.CSV and .XLSX files)
Data files used for visualizations and files used to transform the data.
3. Visualization code (.HTML files)
Plotly javascript coded files used to present and interact with the data.
4. Website (WARC file)
An archival version of the online component of *Tracking Keywords in American Studies*. Located at <http://trackingkeywords.net/>.

Introduction

American Quarterly, the journal of the American Studies Association (ASA), has been the key academic journal in the field of American Studies since its creation in 1949. The journal has represented the field and at times served as an outlet for examining the purpose and future of American Studies. Henry Nash Smith, Leo Marx, and Amy Kaplan form a general consensus throughout the latter half of the 20th Century that American Studies has continually shifted as a field, transitioning from its myth making stage between the 1920's and 1950's, where academics sought to conceptualize an American identity, to focusing on social history and ethnic studies in American in the 1960's and 1970's.¹ In the 1960's and 1970's, the field began to examine issues within the context of "America," rather than to develop a concept of America.² In 1997, Mary Helen Washington looked at this new orientation, stating how it only occurred when American Studies became increasingly interdisciplinary, incorporating fields like African American Studies, Asian Studies, Hispanic Studies, Native American Studies, LGBT Studies, and Women's Studies.³ In 1998, Janice Radway criticized the usage of the word "American" in the name of "American Studies," and how it might restrict the field of study possible, regardless of how interdisciplinary it had become, to only look at the United States; it might erase "the fact that other nations, groups, territories, had already staked their own quite distinctive claim to the concept and name, 'American.'"⁴ Currently, Russ Castronovo and Susan Gillman have pushed for American Studies to examine issues through additional interdisciplinary frameworks and

¹ Smith, "Can 'American Studies' Develop a Method?," 206; Marx, "Thoughts on the Origin and Character of the American Studies Movement," 399; Kaplan, "Left Alone with America: The Absence of Empire in the Study of American Culture," 11.

² Wise, "'Paradigm Dramas' in American Studies," 332; Castronovo and Gillman, "The Study of American Problems," 2.

³ Washington, "Disturbing the Peace."

⁴ Radway, "What's in a Name?," 6.

fields. They argue that issues might be examined as “objects” to be traced through time and across multiple disciplines in the context of “American problems.”⁵

This discourse on interdisciplinary approaches to American Studies makes a recent appearance in *American Quarterly*'s 2018 issue that focused on digital projects. Along with the projects presented, scholars offer insight into how digital tools can inform American Studies, showcasing how the discipline of Digital Humanities offers new ways to critique and analyze issues. The articles also attempt to address concerns and hesitations of scholars incorporating digital tools in their research, whether from fear of neoliberal influences, lack of technical skills, or other misunderstandings. However, Lauren Tilton points to the strengths of using digital tools, especially to provide new views on what is now a largely digital culture. She even acknowledges that the tension Digital Humanities develops within the humanities should be used as an advantage and as inspiration to critically analyze the software and programs that influence daily life. Here, she calls on scholars in American Studies to ask, “What might computational analysis of our scholarship tell us about our very field? Let’s find out!”⁶

Tracking Keywords in American Studies is a project that answers this call. My project uses text analysis tools and data visualization techniques to open examination of the use of keywords in *American Quarterly*. Utilizing Bruce Burgett’s and Glenn Hendler’s *Keywords for American Cultural Studies* to guide word counts in *American Quarterly* and complementing this with algorithmically generated topics based on articles in *American Quarterly*, the project provides a new perspective on the development and current status of American Studies.

⁵ Castronovo and Gillman, “The Study of American Problems.”

⁶ Tilton, “American Studies + Computational Humanities,” 637.

The project also aims to demystify the process of computational text analysis to non-technical academics by providing transparency into the processes and methods used. I make the data the visualizations use available to provide additional transparency and accountability to potential biases that are inherent in text analysis algorithms.

Some selected findings brought out from the keyword charts and topic model charts reflect the histories provided by Smith, Marx, and Kaplan, and show that American Studies is indeed increasingly using interdisciplinary methods. The keyword counts alone offer new perspectives on Washington's argument, for example, by suggesting that the keyword White is at the center of discourses involving Race. Finally, the topic models offer options of at least two sets of topics that show potential as possible keywords not described in *Keywords for American Cultural Studies*, and give insight in discourses in American Studies that no longer are present.

Context on Field Formation

In brief, American Studies is an interdisciplinary academic field that examines American history, society, and culture. Gene Wise attributes one origin of the field to Vernon Louis Parrington, with Parrington's publication *Main Currents in American Thought* in 1927.⁷ Wise views the development of the field as a way to create an academic precedence to study and teach American materials in History and English departments, and at the time used by non-WASP (White Anglo-Saxon Protestant) groups such as Jewish, Catholic and other groups to gain footholds in academia.⁸ Washington later criticized these early scholars for setting a tradition of insularity, looking only at "America" in the context of the United States of America, and disregarding the hemispheric Americas and the transnational influences that create the American

⁷ Wise, "'Paradigm Dramas' in American Studies," 298.

⁸ Marx, "Thoughts on the Origin and Character of the American Studies Movement," 399.

identity.⁹ In the 1940s, many major universities began to develop American Studies programs (often titled American Civilization programs), officially institutionalizing the field and promoting the examination of Americanness by looking at the American “mind” and American “thought.”¹⁰ Leo Marx attributes the success of this “campaign to sanction the scholarly respectability of American subject matter” from the 1930’s to the 1950’s to the nationalism of World War II and early Cold War climate.¹¹ This institutionalization also led to the creation of the journal *American Quarterly* in 1949, at a time “when American Studies needed somewhere to gather” and “focus inward...its great centrifugal energy.”¹² Shortly following the creation of *American Quarterly*, in 1951 the American Studies Association (ASA) was formed to further unify the field, and adopted the journal as the ASA’s official publication.¹³

However, even early on arguments were made to include methods of inquiry from the social sciences into American Studies, with scholars like Henry Nash Smith claiming that cultural works cannot be examined without understanding the cultural context in which the works were created. To enable this, Smith and these scholars promoted a shift away from only examining “high” culture, to looking at popular culture.¹⁴ The shift only began to happen in the late 1960’s, when “the movement would widen its boundaries to include black studies, popular culture studies, folklore studies, women’s studies ... Third World studies, Native American studies, among others.”¹⁵ Along with widening of the field’s boundaries, was the more nuanced inclusion of each field’s approach to similar topics.

⁹ Washington, “Disturbing the Peace,” 2.

¹⁰ Wise, “‘Paradigm Dramas’ in American Studies,” 306.

¹¹ Marx, “Thoughts on the Origin and Character of the American Studies Movement,” 400.

¹² Nye, “Robert E. Spiller and the ASA.”

¹³ Bode, “The Start of the ASA.”

¹⁴ Smith, “Can ‘American Studies’ Develop a Method?”

¹⁵ Sklar, “American Studies and the Realities of America”; Wise, “‘Paradigm Dramas’ in American Studies.”

Despite this shift to include interdisciplinary approaches, the disciplines were not represented in the ASA conferences. In Mary Helen Washington's 1997 ASA Presidential Address, she brings to focus that even in 1985, the conference had only a limited representation of minorities in attendance and few sessions that dealt with the "newer" disciplines, mentioning African American Studies, Women and gender studies, popular culture, ethnic studies, Hispanic Studies, Native American Studies and Asian Studies, that were incorporated into the field in the previous years, especially compared to the conference held in 1997.¹⁶

More recently, leading scholars of American Studies examined issues as "objects." Russ Castronovo and Susan Gillman describe these objects as a way to examine an issue temporally and disciplinary, connecting it to a variety of different "American problems."¹⁷ They use Rodrigo Lazo's chapter on "the archive" in *States of Emergency: The Object of American Studies* as an example, focusing on how archives appear both as constructed official archives and organic unofficial archives, and then examining how archives grow over time and the issues of inclusion in the archive. Lazo then switches to the archive's "present" as a supposedly impartial mediator of history supporting the nation-state and issues of power. This recent approach truly brings to focus the interdisciplinary aspects of American Studies and America as the issue being examined, rather than looking for Americanness in the issue.

Keyword Background

In 2016, Robert Warrior celebrated the increased diversity of that year's ASA conference.¹⁸ In his address, Warrior points out that starting in 2006 the ASA began indexing its

¹⁶ Washington, "Disturbing the Peace," 4.

¹⁷ Castronovo and Gillman, "The Study of American Problems," 5.

¹⁸ Warrior, "Home / Not Home," 205.

conference programs using keywords. The index he highlights became a way for attendees to gain better insight into the program. While in this case scholars might use “keywords” much like a book’s index, scholars might also use keywords as definitions for concepts in the field. These definitions allow for scholars to contextualize different issues in American Studies and play an important role in how the field develops a common vocabulary between multiple disciplinary studies. Keywords are essential to this interaction between different disciplines and allows for scholars from multiple disciplines to converse together. With such a diversity of disciplines and issues addressed in American Studies, scholars have sought to define and summarize the keywords to create a common understanding of topics and issues in an American Studies context.

Raymond Williams, the British cultural studies academic, first wrote about a need to understand common vocabulary, and analyzed how scholars use such terms in cultural studies.¹⁹ This developed vocabulary is akin to the common understanding scholars in the sciences have with common methods and proven theories, but for a much more subjective field. His book, *Keywords: A Vocabulary of Culture and Society*, published in 1976, was prompted by his experience of returning to academia after being in the military during WWII. After this return, Williams noticed how academic language had changed and became more aware of how it differed from nonacademic usage. His book created a list of important terms in the cultural studies field and defined them with essays that traced the history of the term and how cultural studies scholars use the terms. The book left a large impression in the humanities, making those that followed think more critically about the terms they used and the meanings behind those terms. It also spurred many others to create similar vocabularies.²⁰

¹⁹ Williams, *Keywords*.

²⁰ Patterson, “Keywords,” 68; Burgett and Hendler, *Keywords for American Cultural Studies*, 2007, 3.

Since *Keywords* was written and published with a British/Eurocentric mindset, Americanists sought to define and redefine terms to better fit the American Studies field. The first from the ASA was *Contemporary America*, a special issue of *American Quarterly* published in 1983. *Contemporary America* followed a format that had solicited authors to write essays on broad keywords like “Politics,” “Art,” and “Education.”²¹ The *Contemporary America* topics were chosen and aimed at an audience who were not familiar with American Studies, and served as a brief introduction to how issues were discussed in the field. While notable as an attempt to define conversations in American Studies, the special issue was overall lacking, focusing only on 12 broad themes rather than concrete keywords. Overall, *Contemporary America* showed the insular tradition of American Studies that Mary Hellen Washington criticizes by largely omitting diverse disciplines and outside points of view, including any references to Williams.

In 2007 *Keywords in American Cultural Studies*, a much more thorough publication than *Contemporary America*, created a new redefined list of keywords, providing the terms with definitions that aim to reflect current trends in American Studies. The essays in *Keywords for American Cultural Studies* differ from *Contemporary America* and *Keywords: A Vocabulary of Culture and Society* in that they not only seek to contextualize the different terms in the field currently and historically, but also encourage readers to think about the future evolution of the term. Also, the target audience of the publication were scholars and students new to the field, and geared especially for classroom use and discussion. This takes Williams’s idea of introducing terms in a field and links the keywords specifically to how such words might be used in American Studies.²²

²¹ Ribuffo, “Introduction.”

²² Burgett and Hendler, *Keywords for American Cultural Studies*, 2014, 3.

Bruce Burgett and Glenn Hendler, editors of *Keywords in American Cultural Studies*, also aimed to create a dynamic set of keywords. After the first edition in 2007, Burgett and Hendler created a digital collection of essays submitted by students in order to promote creative responses and keep the list of essays open ended. Their digital collection incorporates crowdsourcing, or an outsourcing of work over the internet to have people contribute to the project independently. It also is an online publication, often used in the Digital Humanities field, that allows the project to reach a wider audience. These digital editions were also unique in that students and scholars submitted the articles freely to the publication, rather than the editors requesting specific chapters. This allowed Burgett and Hendler to move away from the Williams model (one single author writing all the keywords) and the model followed by *Contemporary America* (invited scholars who the editors trusted as experts on the subject). Burgett and Hendler instead received submissions from outside their academic and social circles, diversifying the disciplines represented and the points of view they included.

A second edition, published as a print/digital hybrid in 2014, put 64 essays in print and made an additional 33 essays available online.²³ The decision to create a hybrid publication allowed Burgett and Hendler to publish a wider variety of keywords, incorporating more disciplines and viewpoints into the publication. A list of all the keywords from *Keywords for American Cultural Studies* had been put in Appendix A. For purposes of my project, I searched these keywords and counted both the literal words and their interpretations to create the keywords chart. The details of this are described in the Development - Data Collection & Transformation section. These keywords also serve as an authoritative reference point for

²³ Burgett and Hendler, *Keywords for American Cultural Studies*, 2014.

interpreting the topic model results, letting the user compare the algorithmically generated topics to keywords created by scholars.

While *Keywords for American Cultural Studies* captures the keywords in their current state, they also provide insight into how the words developed within American Studies. This hindsight opens conversation on when the words entered the field and shows how the terms have been impacted by other disciplines. By positioning the keywords this way, it allows for any quantitative/computational examinations of the keywords to include prior scholarship in the field. These results can then be compared to trends that have otherwise only been qualitatively described by scholars like Henry Nash Smith, Leo Marx, Mary Helen Washington, Janice Radway, and Amy Kaplan.

Starting in 2004, *American Quarterly* has been actively highlighting different disciplines and contexts of study by putting out annual “Special Issues” that focus on individual themes in the field. Often, these special issues also structure the conversation around a few specific keywords. While *American Quarterly* had occasional special issues in the past, this current series of special issues actively seeks to broaden the scope of the journal by dedicating a series of articles to look at a single discipline or theme, at times making pointed efforts to include scholarship that had been previously overlooked or excluded.

Although the keywords from *Keywords for American Cultural Studies* serve as excellent introductions for how issues might be addressed in American Studies, having a set of formalized words can be problematic. Burgett and Hendler acknowledge this in their publication by calling for other keyword projects to be collaborative and open to elaboration and amendment, while also maintaining a critical approach to the questions of the field.²⁴ Even before *Keywords in*

²⁴ Burgett and Hendler, 6.

American Cultural Studies was written, Janice Radway in her 1998 Presidential Address to the ASA raised the question “what do names do?” She did this to warn how the word “American” can lead to the “premature closure through an implicit, tacit search for distinctively American ‘common ground.’”²⁵ This warning shows how using the word “American” possibly closes off different concepts of “American” by other cultures, tying it too closely to the concept of the United States of America and the culture that is defined by US borders. This usage also excludes other viewpoints on “American” and effectively colonizes for the academic culture of the United States as developed by earlier American Studies research. Radway prompts the notion that scholars in American Studies should not look for Americanness in their objects of study, as Americanists were doing in the field’s earlier history. Instead, she says that scholars should evaluate the interactions of the objects within issues and how “America” impacts these interactions. Radway admits that her solution was not perfect, but she was optimistic that change was possible based on the increasingly interdisciplinary trends she saw in the field.²⁶

In the 2018 *American Quarterly* special issue highlighting digital projects, Lauren Tilton takes the argument about methods and self-analysis in American Studies into the context of Digital Humanities. She makes a case for using critical computational methods in American Studies, saying that “computation offers a methodology for studying culture.” To support this, Tilton lists projects that highlight key computational methods like text analysis, spatial analysis, and distant viewing. She argues that these methods and Digital Humanities as a discipline offer a critical examination at the “computational undergird,” composed of the software, algorithms, and user interfaces that make up everyday life.²⁷ She echoes Radway by arguing American Studies

²⁵ Radway, “What’s in a Name?,” 3.

²⁶ Radway, 24.

²⁷ Tilton, “American Studies + Computational Humanities.”

must expand its boundaries, stating that computational methods bring new critical approaches, not only by offering new takes on the field, but also by moving computational methods from the domains of companies and government agencies into the academic humanities.²⁸ She calls for explorations of how computational analysis can be used within American Studies but also to analyze the field itself.

Methods - Overview

In this project, I combine computational text analysis tools (using programs, code, and algorithms to examine text-based documents) with the method of exploratory data analysis (EDA) to expand on traditional text analysis. As a method, EDA allows for using statistical models to see what data can tell us before applying a formal thesis and approach to a data set.²⁹ My inclusion of EDA allows for using statistics and data visualization in the discourse of how American Studies has developed as a field over time. In this case EDA is used to explore the results algorithms that find similarity in texts presented as visualizations, which allows the researcher to decide what avenues they want to develop hypotheses towards. For example, if the visualizations bring up a strong presence of words involving gender, you might be prompted to analyze the text with an eye to how gender plays a role in it.

In my project, there are two data sets presented as EDA. The first set examines *American Quarterly* through the 96 keywords presented in *Keywords for American Cultural Studies*. The second set examines the “topics” that represent themes and discourses in *American Quarterly* I generated using a topic model algorithm. The two data sets act as companions to each other, with the keywords showing how American Studies is “defined” today. The topics give insight

²⁸ Tilton, 637.

²⁹ Arnold and Tilton, “New Data? The Role of Statistics in DH.”

into past trends and trends that may not be addressed by the keyword. My project presents the data as interactive temporal charts, which will allow users to select the data points they are interested in, and to make their own charts to show trends in the data they thought were important. This flexibility allows the user to come up with their own interpretations of the data, analyzing the texts through the issues that interest them. Frank Moretti's work in literary history is an excellent example of examining data trends to find new insights. He analyzed the trends in the number of novels published, revealing the impact of historical events on the industry and the 25-year lifecycle of popular genres in 19th century British novels that was otherwise unquantified.³⁰

The added benefit of EDA is that it offers a less biased analysis and raises overlooked trends, letting a researcher see the “bigger picture” of the data. It is important to note that since I am the person creating the project, the initial data sets are based on my interpretation of keywords and what topics make sense to me. However, to combat biases I may have, I have made the data the visualizations are based on available and open for critique. I also present the project as EDA to enforce the idea that these results are prompts for further investigation and discussion, and not a final quantitative analysis on American Studies. In the next few paragraphs, I will highlight examples of algorithm-based projects and discuss the issues with algorithms as a tool.

In theory, an algorithm does not approach an analysis with preconceived notions of what should or should not be a topic, theme, or grouping, so it can possibly bring up things that have gone unnoticed. The Viral Texts project is an excellent example of this, as it creates visualizations that clearly depict how different 19th century periodicals from the United States,

³⁰ Moretti, *Graphs, Maps, Trees*, 18.

the United Kingdom, and Australia shared articles. The results show how some news publications had large influence over others and how texts flowed across the periodical landscape.³¹ Viral Texts used algorithms to identify similar blocks of texts between periodicals, and then traces how texts moved through American periodicals, depicting the flow of texts from periodical to periodical through network graph visualizations. The algorithms brought up some interesting findings in the texts. One finding titled “Fugitive Verses” brought up anonymous poems that were widely shared and read, but which have been generally omitted from the literary canon. Other findings also showed widespread misattribution of poems, and allowed the user to see how attribution changed from newspaper to newspaper. These results also give excellent insight into what periodicals were central to the 19th century in disseminating original content through network graphs, a type of visualization. *Tracking Keywords* utilizes visualizations similarly, making comprehensible what are otherwise large sets of numbers. This makes it simpler for users to clearly see how the keywords and topics track through time and how they correlate to each other.

However, while Viral Texts had a straightforward approach to analyzing its data, algorithms are only as unbiased as their data sets and the people in control of the algorithm training process. Excessively “cleaning” the data, a process to remove errors or create a more uniform data set, can lead to erasures of content. This happens because decisions about what to remove is left to the researcher and their own biases.³² The same goes for training the algorithms, where the researcher makes changes to the algorithm until it creates results that are meaningful to the researcher. In topic modeling, explained later in this paper under “Methods for Data Analysis,” the algorithm creates a different outcome each time it is run, so the decision of which

³¹ Cordell, “The Viral Texts Project.”

³² Rawson and Muñoz, “Against Cleaning.”

set of results is “best” is mine and based on my judgement of what is most relevant to American Studies. For this project, topic modeling will be used as a companion to the keyword counts. To account for my own bias of what may be “good” results, the topics provided are contextualized as themes or discourses, rather than alternative topics to keywords. This allows for open interpretation of which issues these results can be applied to. This contextualization also signifies this is not the only possible set of results, and that there are others yet to be created.

The algorithms themselves can also have implicit bias due to the social context of who created them or have bias built into the texts themselves through stereotyping language. Recent studies have shown how gender and racial stereotypes in texts can skew machine learning algorithms to repeat and reinforce those stereotypes.³³ Passing along these stereotypes becomes even more problematic when the algorithms are tasked with embedding data into images and websites, though algorithms can be trained to interpret texts otherwise if supervised with a critical approach.³⁴ This issue can be used to the advantage of the project by bringing up problematic language of past issues in *American Quarterly* and seeing when, or even if, the language changed to mark changes in cultural trends.

In addition to negating bias in algorithms, it is important to provide insight into the project methods in order to demystify and make accessible the technical process of computational text analysis. Following the direction of Digital Humanities, my project takes steps to make as much of the project as accessible to nontechnical people as possible by describing the process in plain language. For example, instead of saying the data was “transformed” to fit a certain model, the actual steps of the transformation will be described. This

³³ Caliskan, Bryson, and Narayanan, “Semantics Derived Automatically from Language Corpora Contain Human-like Biases,” 183.

³⁴ Bolukbasi et al., “Man Is to Computer Programmer as Woman Is to Homemaker? Debiasing Word Embeddings,” 8.

type of access provides transparency into decisions made at different points in the process and to promote a general understanding of computation text analysis. By developing this general understanding of the methods, it allows the user to critique the interaction of keywords in American Studies more effectively. This general understanding also develops an academic infrastructure for the users to evaluate other projects.³⁵

Tracking Keywords in American Studies combines the methods and techniques of EDA, algorithms, and transparency to show not only how digital tools can be used in American Studies, but also provides a framework for others to work within and possibly replicate the results. This framework can as be applied to other academic journals or other fields. The concept of EDA will guide the usage of algorithms and text analysis to create visualizations that prompt deeper analysis into how American Studies has developed over time, and into the role keywords and specific topics have played in the field. The data transparency will also hold the project accountable for any biases and oversights in analyzing the texts.

Methods - Data Analysis

The project focuses on two methods of text analysis: word counts and topic modeling. These two methods offer two different variations of the results, which can be compared or can supplement one another.

Word counts are based on the keywords presented in *Keywords in American Cultural Studies*. Word counts let the researchers see how often a word or set of words appear in a document, and then this data is graphed to show how much the word was used within a time frame. This method provides a very literal way of looking at the keywords, letting the researcher

³⁵ Liu, "Drafts for Against the Cultural Singularity (Book in Progress)."

assume that if a writer is talking about a topic, they have to use a word that is associated with the topic. While this method is a strong way to evaluate terms like Islam, it is weaker when applied to abstract topics like Interiority (describing inner character and/or nature), which encompass a collection of feelings or emotions. Word counts also do not account for a keyword whose definition has changed over time (i.e. Queer), nor ones that may have several connotations (i.e. Sex).

To supplement the word counts, the articles are also analyzed using topic modeling. Topic modeling is a statistical model used to cluster words that frequently occur together using algorithms, creating different “topics.” This type of clustering allows for a uniform analysis across large bodies of text, bringing up topics that might be missed by a human reader. Each topic is assigned a statistical percentage, here called a “weight,” that shows how strongly a topic is present in any given document. By using visualizations to present these statistics, it is possible to show how different articles share the same topics and how prevalent topics were at different points in time. To develop the topics, the researcher needs to tell the algorithm to search for a specific number of topics that each contain a set number of words. For example, if I set the algorithm to search for 100 topics each containing 5 words, the algorithm will look for 5 words that statistically correlate to each other and group them as a topic, while looking for 99 other combinations of 5 words that are statically relevant.

Each topic is then evaluated to see if the results are meaningful to the analyzed texts, and then repeated if better results are needed. For example, a hypothetical five-word topic containing “cows dairy cheese market pasture” is fairly clear and understandable to a person. On the other hand, a five-word topic of “toast american cat digital yes” is unlikely to be meaningful, and the algorithm’s topic settings would need to be adjusted. This can be done either by changing the

number of topics to look for, or by changing the number of words that make up that topic. By changing these numbers, it is also possible to decide how broad or granular a topic is created. An example of granularity in topics is provided in context of *American Quarterly* later in the paper. In addition to this, the topic model algorithm evaluates the texts slightly differently each time, meaning the chances of getting exactly the same result every time are very small. Due to this issue, I use topic modeling to represent themes and discourses within the texts. The topics should not be considered as tagging of articles with words.

Topic modeling has been used extensively in the *Quantifying Kissinger* project, for example.³⁶ There, scholars used topic modeling to “read” memos from Henry Kissinger between 1969 and 1976, showing how different terms group closely together. Visualization of these results show interesting trends, such as how different topics were talked about in Secret and Top-Secret documents. By design, *Quantifying Kissinger* does not analyze the papers with a specific thesis or hypothesis in mind, but first processes and presents the data to create an opening for others to develop their own interpretations. My project follows the same model by presenting the visualizations as another way to interact with *American Quarterly*.

Development - Data Collection & Transformation

The data collections and transformation process contained multiple steps. I first collected the articles, then converted them into text files, analyzed them via word counts and topic modeling, and then had that raw data transformed into a format that can be visualized. Each stage of the processes is outlined below in Figure 1.

³⁶ Kaufman, ““Everything on Paper Will Be Used Against Me.”

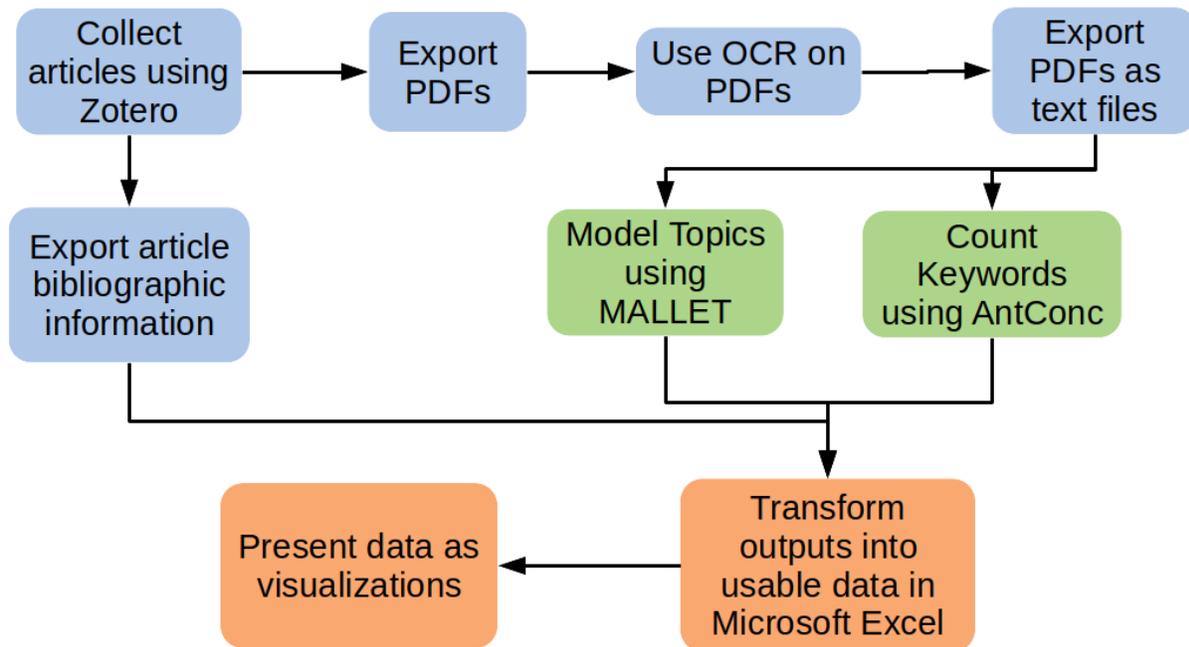


Figure 1. Flowchart describing data collection, processing and transformation.

The first step in the project was gathering relevant articles. I decided to only use articles printed within *American Quarterly*, excluding book and event reviews, notes, updates, and lists of recent publications. This was because the articles are the main purpose of the journal and a standard type of text that appeared throughout its history, while reviews and other types of articles were not as consistently published. Access to *American Quarterly* was spread between two different academic research databases, JSTOR and ProjectMuse, each with different metadata and PDF file names. To help create uniform PDF file names, Zotero harvested all the articles and then exported them into a single folder for further processing. Zotero is an open source software program that is normally used as a citation management tool. The file names contained the last name of the author, the year of publication, and article title. Other data, such as database provided subject terms, were not included when collecting the articles due to copyright

concerns (database providers generally hold copyright on those terms), and due to lack of consistency within and between the databases.

As another avenue of collecting articles, I investigated JSTOR's Data for Research service. This service allows a researcher to request batches of text files for text analysis, along with some word usage analysis statistics. However, the service only covered issues of *American Quarterly* that JSTOR held (1949-2012 at time of project), and there was no way to easily separate out unneeded selections like book reviews, or front and end matter. The naming conventions of the files were also not as useful for visualization. They had the file names encoded with numerical IDs but did not have a key file or list to connect IDs to any useful information. However, this service showed potential for examining everything within the journal between the start in 1949 until the last seven years of publication (likely ProjectMuse's embargo date on the publication).

For the files to be processed by the different programs, I converted the PDFs into text files. This presented some problems due to the slight differences in PDFs. Some have pages that were stored as images and not text, making them unable to convert straight into plain text files. Using Adobe Acrobat DC, I removed the first page of each article, which only contained the cover page with general citation information. After this, the PDFs were run through Adobe's OCR (Optical Character Recognition) function to make sure all the text in each document was stored as machine accessible text. Following the OCR, the documents were exported as text files, a standard format that both AntConc and MALLET could use.

The first stage of analysis was to get counts on how often each of the 96 keywords were used within a text by using the program AntConc. AntConc is a corpus analysis toolkit for concordancing and text analysis, created and made available for free by Laurence Anthony of

Waseda University, Japan. Anthony is an academic focused on education technology and linguistics who has developed many software tools to support the corpus linguistics field.

The *Keywords for American Cultural Studies* website presents 96 keywords between their Web Essays, Print Essays, and Post Publication Essays. I searched each keyword and its derivatives based on the literal word, or if distinct clarification/inclusion of other applicable terms were listed in the essay. An example of this distinct inclusion was for the keyword “Islam,” which also stated that Arab and Muslim have historically been used as synonyms for Islam. Since the article in *Keywords for American Cultural Studies* explicitly mentioned Arab and Muslim as being part of the “Islam” keyword, I included them. While this word usage is problematic, either in not representing groups accurately or being outdated language, it was included into the search to make sure the keyword was tracked as accurately as possible. An example of this problematic language was where to include negro. It fit both in the “Black” and “African” keywords, but I ultimately decided to include it with “Black” to represent the “American” nature of the word, where the “African” article was more focused on the transcontinental African influence in America. However, not all keywords can be accurately represented this way. With “Religion” as an example, would it have been best to list every religion and variants of spiritualism in the search? Or look for just where the word religion was included? This is the main weakness of the word count method, and one of the reasons why I included the method of topic modeling. Topic modeling is explained later in this section.

To explain how the keyword derivatives were designated in AntConc, truncation needs to be explained. Truncation is a search technique that broadens your search to include various spellings and endings. It allows a user to add an asterisk (*) into a word to denote that the rest of the term is open ended, and for the program to include variations of that word. Taking the

keyword “West” as an example, using truncation on “west” would turn the search term into west* in order to capture the usage of west, western, westerner, westerners, westernization, etc. Here, you can see that by truncating words in AntConc, the user can include more words that fall within the scope of the term, without having to list them all and possibly miss some. A full list of terms used to search each keyword is available in Appendix B, and can be compared to the keywords list in Appendix A.

AntConc outputs its data as large text files, with each line in the file listing the occurrence of the word and the filename of where it was found. To combine these files into a single spot that can then be used with a visualization, each file was imported into Microsoft Excel as an individual sheet by copying and pasting the data in. A “master” sheet, which combines the article with keyword counts, was created using a Zotero CSV export (CSV is a “comma separated value” file commonly used for saving data). This was modified to only leave in the filename of the article, publication year, author’s name, title of the article, and volume/issue numbers of the journal the article was published in.

In addition to the bibliographic information from the Zotero CSV export, each keyword followed in its own column. The COUNTIF() spreadsheet function in Excel was used to look at the individual AntConc output sheets. This function creates a tally of how many times a target cell (in this case the article filename) appears in a data set (the AntConc output sheets). The COUNTIF() functions are combined in a single spreadsheet, with each article listed and how many times a keyword appears in a single article (see fig. 2). While this fine-grained data is informative on its own, showing which keywords are relevant to which article, it did little to show what was happening on a yearly basis. To summarize the data in a more meaningful way, another sheet was created in the Excel document to combine the keyword counts by year. This

utilized the SUMIF() function, which adds together target cells, showing the total number of times a keyword appears in a single year (see fig. 3).

	A	B	C	D	E	F	G	H	I
1	Filename	Author	Title	Issue	Volume	Publication Year	Aboliton	Aesthetics	Affect
2	Den_Hollander-194	Den Hollander, A. N. J.	A Preliminar	4	1	1949	0	0	0
3	Cantwell-1949-Haw	Cantwell, Robert	Hawthorne a	4	1	1949	0	0	0
4	Lens-1949-Labor_B	Lens, Sid	Labor Betwix	4	1	1949	0	0	0
5	Morgan-1949-Amer	Morgan, Edwin	American Art	4	1	1949	0	0	0

Figure 2. A snapshot of the keyword count spreadsheet that collects how many times each keyword is present in a single document. The COUNTIF() function is shown in the cell.

	A	F	G	H	I	J	K	L	M
1	Year	American	Asian	Black	Body	Border	Capitalism	Citizenshi	City
2	1949	0	0	3	0	0	0	0	0
3	1950	0	0	6	0	0	0	0	0
4	1951	0	0	1	0	0	0	0	0
5	1952	0	0	17	0	0	0	0	0
6	1953	0	0	3	0	0	0	0	0
7	1954	0	0	1	0	0	0	0	0

Figure 3. A snapshot of the keyword count spreadsheets that shows how many times a keyword is present in a given year. The SUMIF() function is shown in the cell.

Following the keyword counts, I analyzed the texts using topic modeling. For this I used the package MALLET, an open source software toolkit made for text analysis created by Andrew Kachites McCallum at University of Massachusetts Amherst. After downloading and setting up the program, I ran the topic model algorithm a few times to test the quality of results. I changed the number of topics to look for and how many words define a topic until results were created that made sense to me. In order to keep the topics suitable for a visualization, I decided to have the program output 100 topics that consist of three words each. This decision limited the

granularity and nuance of the topics, but it made for a visualization that was less overwhelming and easier to comprehend and interact with, due to either too many words or too many lines.

For instance, to explain the differences in the number of words for each topic, here is a set of outputs that all revolve around the 2008 financial crisis:

- 1-word topic - “crisis”
- 3-word topic - “financial crisis debt”
- 5-word topic - “crisis financial debt neoliberal economic”
- 10-word topic - “crisis financial economic neoliberal debt percent economy global neoliberalism capital”

One can see how by having more words associated with a topic, the more nuanced the topic becomes. Starting with just a concept of crisis, a topic can be associated with the keywords Finance, Market, and Economy. The words in the topic are also ordered from most relevant to least relevant to help guide the person interpreting them. By deciding to limit the topics to only three words, MALLET’s output allowed for enough nuance to see how topics can relate to keywords, but kept the visualization from turning into word soup. I present some examples of topics that relate to keywords in the Findings section of the paper. I similarly decided to only create 100 topics, limiting the granularity of the results compared to what 150 or 200 topics might be. If I created more topics, there would have been a wider variety of granular topics to interact with. However, this would have been an overwhelming number of topics for the user to scroll through and examine. Figure 4 is a snapshot of MALLET’s keys file which offers insight into what the output files being evaluated look like.

	Topic #	Overall Weigh	Topic
0	Topic 1	0.0191	muslim arab muslims
1	Topic 2	0.21984	history historical past
2	Topic 3	0.22833	literary literature book
3	Topic 4	1.25395	time fact point
4	Topic 5	0.02018	civilization civilized tommo
5	Topic 6	0.06307	narrative narratives story
6	Topic 7	0.03429	latin puerto cuba
7	Topic 8	0.10722	james boston henry
8	Topic 9	0.02123	cherokee missionary missionaries
9	Topic 10	0.02524	digital humanities data
10	Topic 11	0.06146	indian native indians
11	Topic 12	0.03331	twain mark twain's
12	Topic 13	0.01393	baldwin wright shock
13	Topic 14	0.04333	melville thoreau nature
14	Topic 15	0.06885	england english massachusetts

Figure 4. An example of a MALLET keys file. First column is the key number, second is the topic number added in myself, third is the weight of each topic in all of *American Quarterly*, and the fourth are the words that comprise each topic.

In addition to the keys file that lists all the topics, MALLET outputs a composition file. The composition file is a table that lists each document within the corpus and assigns a percentage of how relevant each one of the 100 topics is within that single article. The key file shows each topic's weight throughout the entire corpus, while the composition file shows the weight of each topic within a single document. See Figure 5 for an example of a MALLET composition file.

Filename	Publication Year	Topic 1	Topic 2	Topic 3
Ayres - 1949 - Land of the Free Ride.txt	1949	2.04E-05	0.002366	2.43E-04
Ayres - 1949 - The Industrial Way of Life.txt	1949	1.40E-05	0.003826	0.005298
Beloff - 1949 - The Projection of America Abroad.txt	1949	1.66E-05	1.92E-04	0.010657
Brown - 1949 - The Ash Can School.txt	1949	1.60E-05	0.001023	1.91E-04
Cantwell - 1949 - Hawthorne and Delia Bacon.txt	1949	7.67E-06	0.010131	0.008527
Couch - 1949 - A Case History in Book Publishing.txt	1949	1.43E-05	1.64E-04	0.174386
Curti - 1949 - The Reputation of America Overseas (1776-1860).txt	1949	3.88E-06	0.002481	0.02847

Figure 5. Example of a MALLET composition file. The second column is the filename and the following ones are the weight of each topic within the file and correspond with the keys file. Headers are my own.

The composition file was handled similarly to the keyword counts. The key file was first opened in Microsoft Excel as a single sheet and a column added labeling each row as Topic 1, Topic 2, etc. Next the composition file was opened and added into the Excel file as a second sheet. A column there was added showing the publication date of each article, copied from the Zotero CSV export file. The data from these two files were then combined and averaged according to year and topic using the AVERAGEIF() function. The average of the topic in a year was used since the numbers assigned to the topic are percentages and cannot be added. This function was set to combine all the weights of a single topic in a given year, and outputs an average sum.

This left me with a final step of converting the Excel documents into something the visualizations could reference as data. To do this, I used a Python script that reads an Excel sheet, “melts” the data using the pandas.melt() function, and exports a new Excel sheet that has. The melt function transposes the cells in the spreadsheets so that each row in the new document is a single data point, rather than multiple data points as it was displayed before. After the document is created, it was saved as a CSV, making a file that is compatible with the visualization code. Only the topic model file was modified further, adding in a column that listed

the words that make up the topic, as can be seen in Figure 6. Now that the files are in the correct format, they can be displayed clearly in a visualization.

		date	weight
1	composition		
2	muslim arab muslims	1949	0.0000142
3	history historical past	1949	0.01568058
4	literary literature book	1949	0.030391965
5	time fact point	1949	0.111288538
6	civilization civilized tommo	1949	0.001345718

Figure 6. A snapshot of how the topic model CSV is organized.

Development - Data Visualization

The goal with the project is to create two visualizations, one for keywords and one for topics created by topic modeling, that are clear and meaningful, but also interactive to allow for people to analyze the different slices of the data. The visualizations needed to show both the prevalence of the keywords and topics, but also how they changed over time. Being able to see how these keywords and topics change over time gives scholars the opportunity to analyze the trends of what American Studies has addressed and what events have impacted the field.

As a concept both the keyword counts and topics were fairly straight forward to work with. Line graphs are traditionally used to represent this type of data and made the most sense to represent keyword usage. The issue with using a line graph was how to clearly represent 96 different keywords or 100 topics and allow for interaction. Portraying 96 lines on a single graph creates a mess that hides more than it shows. On the other hand, having 96 different graphs or some combination of grouping charts by themes also seemed impracticable and subjective. Further, the interactive visualization needed to allow users to export their own findings from the tool if they were to use it in their own research. A final consideration was that the tool be free and open source to allow others to replicate findings.

The decision to use line graphs was also influenced by Castronovo and Gillman, who argue that objects in American Studies should be approached with a temporal dimension in mind. This approach traces the object of study through time and its shifting context through different disciplines to show how issues do not just exist in a single time period or affect a single aspect of society. Line graphs are important to show the temporal context of the keywords and topics, especially if these terms are not to just be examined as discourses or issues in American Studies, but as objects in American Studies. Viewed as objects, each line can be traced as it interacts with other lines, intertwining with other objects in the field. This is particularly relevant to the keywords chart. By presenting the data on a line graph, the keywords can be traced temporally, and the user can see how each of the keywords relate to each other at any given point on the chart. Now the charts are not just a historical analysis but are recontextualized as taking the object of keywords and examining them within the issue of “defining American Studies.”

There were a few code packages that seemed to meet these needs, with D3 and Plotly standing out as the most likely choices. I ultimately decided on Plotly, an open source JavaScript code library based on D3. Though Plotly itself is a private company, the code library is made publicly available online as an open source library under the MIT license. The code library also is well documented and was simpler to use than D3 based on my current coding experience. The downside of using Plotly over D3 was that it did not allow for as much flexibility in the type of charts displayed or as many animation options.

Findings

Tracking Keywords in American Studies serves as a web platform for hosting charts that users can interact with to examine trends, and from these results they can develop questions of their own. The website will also act as a central place to access the raw data sets I created for the keyword counts and topic modeling, the list terms used to create the keyword counts, and a condensed version of the text analysis process. The project can thus be used by faculty and students in two main ways. One, they can look at the results of the keywords counts and topic modeling and attempt to explain trends they see in the visualization. Second, they could look critically at the process, the terms used, and the data, and then argue the pros and cons of how the methods have been used and how it can be improved.

From my own analysis, the keywords chart brings up some interesting insights into American Studies. The chart supports the analysis of the evolution of American Studies made by Smith, Marx, Kaplan, and Washington. The chart clearly shows that based on the current usage of keywords, there has been a significant increase in the variety of topics addressed by the field, showing the increasing inclusion of interdisciplinary approaches and issues addressed. The 1960's see the start of this rise, with some peaks of activity in the mid 1980's and 1990's, with a steady representation of many keywords starting in the 2000's, such as "Empire," "Globalization," "Immigration," and "State." Even taking into consideration that more recent issues of *American Quarterly* have more pages than earlier issues, which means there are more words to count, there is a distinct increase in usage of keywords the closer scholarship is to the present. Predictably, the word "America" dominates the field; the same is true with "Society" and "Culture." They're all relevant to the field from the beginning. Keywords like "War" and

“Finance” show peaks relating to recent events in American society, like post 9/11 military activities and the 2008 financial crisis, respectively.

One of the more important correlations (a statistical relationship between two or more variables) within the keywords I found was the interplay between “Race” and the race/cultural group related keywords. As shown in Figure 7, “Race” and “White” correlate fairly strongly together. When scholars write about “Race,” they will most likely mention “White.” However, when other keywords like “African,” “Asian,” “Black,” “Latino,” and “Mestizo” get added in, it becomes apparent that while these words also correlate with “Race”, they peak and dip around the constant trend of “Race” and “White.” This means that some years “African” and “Black” may be talked about more than “Latino” and “Mestizo,” and other years “Asian” might be the focus. For example, in 2005 and 2015, “Black” dips in representation while “Asian” peaks. However, since “White” does not fluctuate the same way, it shows that scholars in American Studies often write about these other keywords in relation to “White.” On seeing this correlation, it is difficult not to remember Mary Helen’s Washington’s ASA address “Disturbing the Peace.” She argues that as much as the ASA claims that it has accepted racial/ethnic/social diversity and includes interdisciplinary methods, in practice that inclusion is not present and the field is still centered around traditional mainstream white academics. While scholars in American Studies today are more diverse than what Washington criticized in 1997, the center of the conversation has not moved. By looking at the keywords it is clear that “White” is still the center of conversation of race in American Studies, and that scholars in the field may not be doing enough to recontextualize the discussion about race away from comparing groups to “White,” or going beyond the context of “White”.

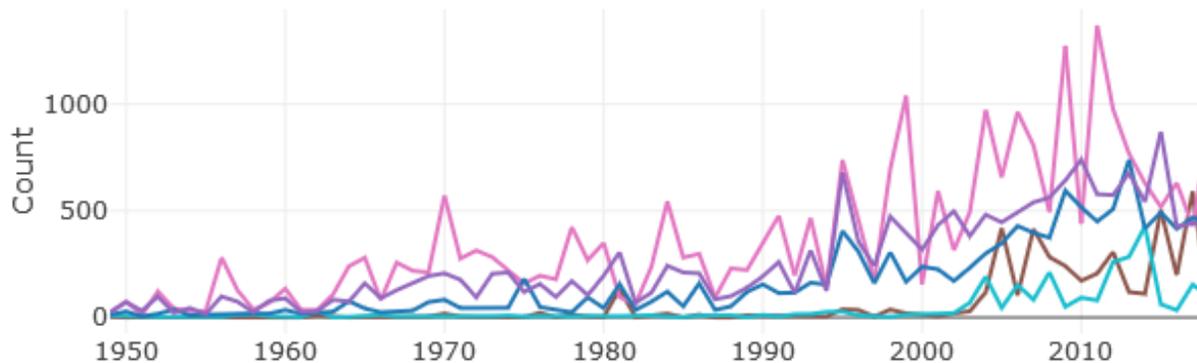


Figure 7. Chart showing keyword counts for Asian (brown), Black (pink), Latino (cyan), Race (blue), and White (purple). Notice the dip in Black as Asian and Latino rise after 2010.

After looking at the keyword chart, I looked at the topic model chart to see how its results compare to the keyword results. The 100 topics listed are a mix that relate to the keywords, and others that give insight into the field’s past and shifts in academic discourse. For example, the topic “law state legal” relates to keyword “Law,” and even follows roughly the same trend of usage. Another example is “muslim arab muslims,” relating to “Islam.” Other topics can be interpreted as parts of a keyword, such as “japanese japan tourism” relating to either the “Asian” or “Orientalism” keywords.

When the topic model chart is examined as a tool that brings out topics that are not named in *Keywords for American Cultural Studies*, I see strong cases for the inclusion of “Art” and “Education” as potential keywords. For “Art,” the “art artists modernism” and “art painting paintings” topics show how the topic has been talked about historically (see Figure 8), but considering that in recent history *American Quarterly* publishes event reviews that cover museum and gallery exhibitions, performances, media events, festivals and films and more, which were not included in the analysis, the topic has a strong case for inclusion as a keyword. Remember, the topic modeling algorithm had no instructions on what topics to look for, just how many topics to create and how granular the topics should be. The algorithm has now read every

article in *American Quarterly*, and after reading every article hundreds of times, it decided that there should be two separate topics for that focus on art out of 100 possible topics. Statistically, that is a result that is difficult to ignore and means that it is something to be investigated further.

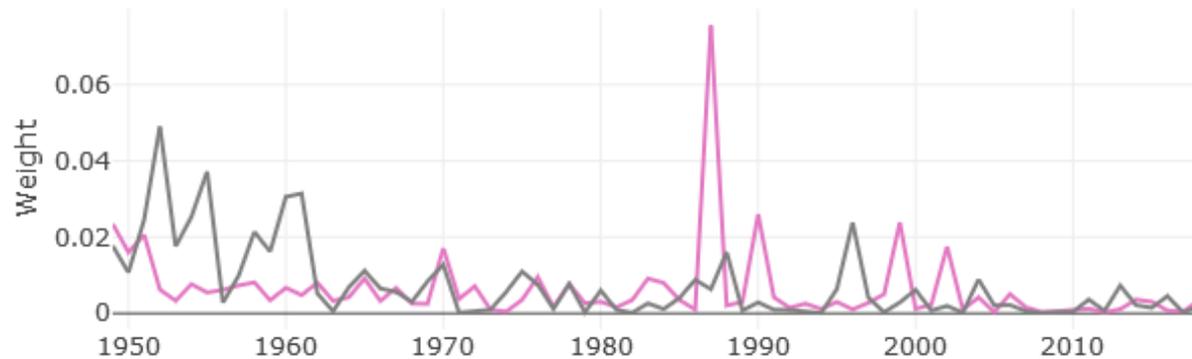


Figure 8. Topic model chart showing “art artists modernism”(pink) and “art painting paintings” (grey)

“Education” as a proposed keyword is a bit more of a nebulous concept, but the topics “school students education” and “university press state” point to a theme about higher education, and if “american studies scholars” is included, “Education” can be interpreted with a strong focus towards academia. Figure 9 shows how the topics track over time. If “Education” is taken as a keyword, questions can be asked of where are the schools with prominent American Studies programs, what types of students do these schools serve, what is the composition of the faculty like, and what impact does this all have on the field. Historically, American Studies exists purely in an academic setting, so the overall absence of keywords in *Keywords for American Cultural*

Studies is something that stands out, especially when it is statistically prominent as a topic.

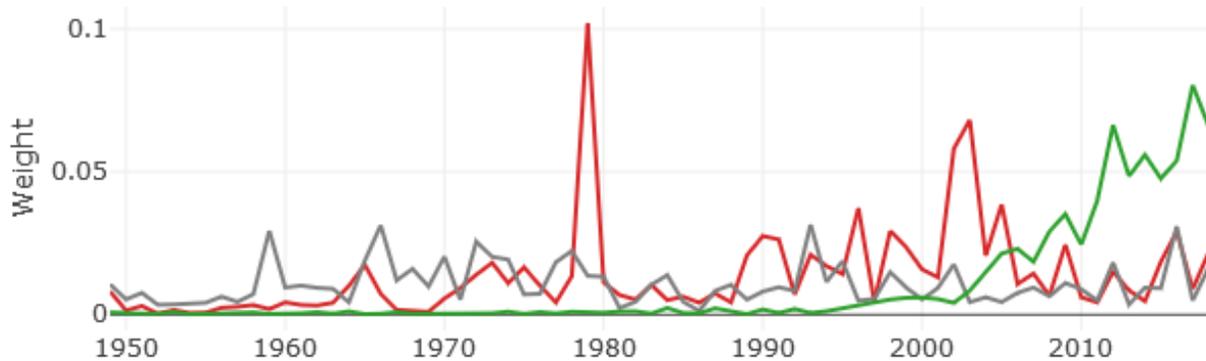


Figure 9. Topic modeling showing “school students education” (grey), “university press state” (red), “american studies scholars” (green).

Two topics give insight into how the discourse and tone of *American Quarterly* has changed over time. As seen in Figure 10, “time fact point” starts off as a very dominant topic in *American Quarterly*, steadily declining towards the present. I interpret this combination of words as academics using objective language, with scholars making strong “points” about “facts.” The decline of this topic reflects the influence of postmodernism on scholarship, and the focus on reinterpreting perceived truths through different points of view and experiences. On the other side of this topic, there is a rise of “cultural american discourse.” This topic is nonexistent in earlier literature, but begins to rise in the 1980s, showing how the *American Quarterly* starts to look at how culture is communicated and represented in an American context. Other topics of note are “folklore anthropology folk,” which starts in 1971 but ends in 2003, as well as all of the topics that focus on canon writers like Mark Twain (who had his own special issue in 1964), Herman Melville, Henry David Thoreau, and Stephen Crane, all declining towards the present. This decline is strongly representative of how American Studies has moved away from canonical authors and has incorporated a more diverse body over work along racial, gender and sexual orientation lines. Together, these two topics point to the shifting nature of American Studies, and

the absence of these keywords that represent these topics shows how *Keywords in American Cultural Studies* is a publication of the “now,” capturing how American Studies exists in the present, which is how Bruce Burgett and Glenn Hendler intended their project.

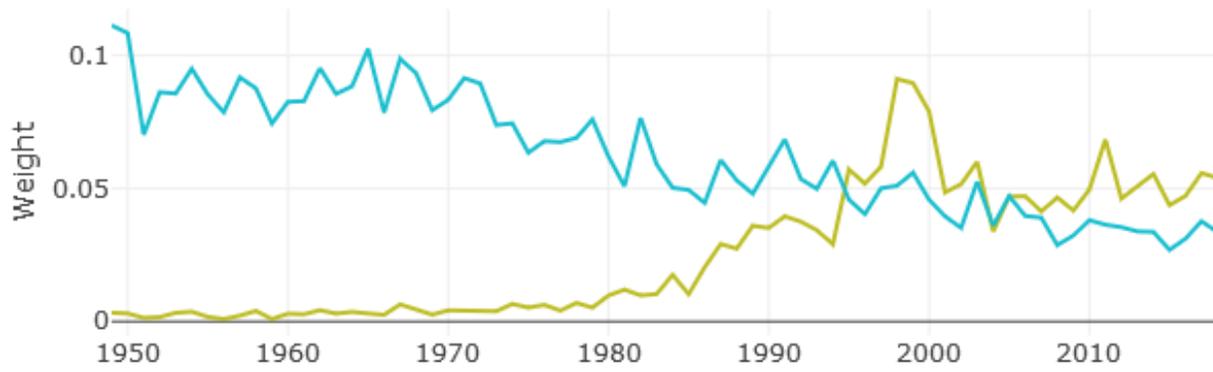


Figure 10. Topic model chart showing “time fact point” (cyan) and “cultural american discourse” (gold).

Future of the Project

Tracking Keywords in American Studies is an ongoing project. My next step will be to present the project to faculty of the CUNY Graduate Center American Studies program as a resource that their students can utilize in exploring *Keywords for American Cultural Studies* in the classroom. Following feedback on the project, I plan to present the project at relevant conferences, and condense it for a scholarly journal publication. I am self-hosting the website using a WordPress site via Reclaim Hosting, allowing me to have control over the intellectual content of the site and to make changes and expand the project in the future.

In the project’s current state, the two charts only track keywords and topics up to 2018. I intend to regularly update the keywords chart to reflect current articles of *American Quarterly*. I also intend to regularly update the topic model chart. However rather than replacing the existing chart, I will expand the chart to allow users to choose past data sets and explore how the MALLET algorithm provides different results. Another change I intend to make is to move my

data transformation process away from Microsoft Excel and explore and experiment with alternatives like Python scripting and relational databases. This will allow the project to be less reliant on proprietary software, and mitigate long term digital preservation issues, such as when application developers go out of business or are bought by other companies. When this happens, programs are often no longer updated and become unsupported, leading to files that were created in those applications becoming unusable.

Although I do not currently have the technical skills to create my own topic modeling algorithm, I would also like to create my own that can be applied to the project. Within the context of the project, this algorithm can also be applied to other academic fields, particularly ones represented in the Keyword series by NYU Press. This work will also further my overall academic career as a librarian and archivist, providing a space to experiment with tools that I can then apply in an archival setting. Similar to Lauren Tilton's call to use computational approaches in American Studies, there have been calls to use more computational approaches in the archives and library science field, but to examine how these tools can provide better or alternative access to materials, providing services that are otherwise solely provided by large private companies, especially database providers.

I believe the *Tracking Keywords in American Studies* is successful in providing insight into how keywords and algorithmically generated topics can be used to create critical arguments in American Studies, and to show how American Studies has developed and continues to develop as an interdisciplinary field. As technology continues to transform academia, projects like this allow for an opportunity to use it for self-reflection and analysis, and also to introduce people to the computational text analysis process, and to think critically about how computational approaches are used and presented in their respective fields. The project should

see its greater success in the classroom, serving as a companion to the current usage of *Keywords for American Cultural Studies*, and providing another source to inspire debate on topics in American Studies.

Works Cited

- Arnold, Taylor, and Lauren Tilton. "New Data? The Role of Statistics in DH." *Debates in the Digital Humanities*, September 2019. <https://dhdebates.gc.cuny.edu/read/4805e692-0823-4073-b431-5a684250a82d/section/a2a6a192-f04a-4082-afaa-97c76a75b21c#ch24>.
- Bode, Carl. "The Start of the ASA." *American Quarterly* 31, no. 3 (1979): 345–54. <https://doi.org/10.2307/2712381>.
- Bolukbasi, Tolga, Kai-Wei Chang, James Y Zou, Venkatesh Saligrama, and Adam T Kalai. "Man Is to Computer Programmer as Woman Is to Homemaker? Debiasing Word Embeddings," 2016, 9.
- Buolamwini, Joy, and Timnit Gebru. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification," 2018, 15.
- Burgett, Bruce, and Glenn Hendler, eds. *Keywords for American Cultural Studies*. Second edition. New York ; London: New York University Press, 2014.
- Caliskan, Aylin, Joanna J. Bryson, and Arvind Narayanan. "Semantics Derived Automatically from Language Corpora Contain Human-like Biases." *Science* 356, no. 6334 (April 14, 2017): 183–86. <https://doi.org/10.1126/science.aal4230>.
- Castronovo, Russ, and Susa Gillman. "The Study of American Problems." In *States of Emergency: The Object of American Studies*. Chapel Hill: The University of North Carolina Press, 2009.
- Cordell, Ryan. "The Viral Texts Project." Accessed September 30, 2019. <http://viraltxts.github.io/>.
- Kaplan, Amy. "Left Alone with America: The Absence of Empire in the Study of American Culture." In *Cultures of United States Imperialism*, 3–21. Durham and London: Duke

University Press, 1993.

Kaufman, Micki. “‘Everything on Paper Will Be Used Against Me:’ Quantifying Kissinger | Text Analysis, Visualization and Historical Interpretation of the National Security Archive’s Kissinger Correspondence.” Accessed September 10, 2019.
<https://blog.quantifyingkissinger.com/>.

Marx, Leo. “Thoughts on the Origin and Character of the American Studies Movement.” *American Quarterly* 31, no. 3 (1979): 398. <https://doi.org/10.2307/2712392>.

Moretti, Franco. *Graphs, Maps, Trees: Abstract Models for a Literary History*. London ; New York: Verso, 2005.

Nye, Russel B. “Robert E. Spiller and the ASA.” *American Quarterly* 19, no. 2 (1967): 291–92.
<https://doi.org/10.2307/2711001>.

Radway, Janice A. “What’s in a Name? Presidential Address to the American Studies Association, 20 November 1998.” *American Quarterly* 51, no. 1 (1999): 1–32.
<https://doi.org/10.1353/aq.1999.0001>.

Rawson, Katie, and Trevor Muñoz. “Against Cleaning.” *Debates in the Digital Humanities*, September 2019. <https://dhdebates.gc.cuny.edu/read/4805e692-0823-4073-b431-5a684250a82d/section/07154de9-4903-428e-9c61-7a92a6f22e51#ch23>.

Ribuffo, Leo P. “Introduction: The Burdens of Contemporary History.” *American Quarterly* 35, no. 1/2 (1983): 3–18. <https://doi.org/10.2307/2712709>.

Sklar, Robert. “American Studies and the Realities of America.” *American Quarterly* 22, no. 2 (1970): 597–605. <https://doi.org/10.2307/2712227>.

Smith, Henry Nash. “Can ‘American Studies’ Develop a Method?” *American Quarterly* 9, no. 2 (1957): 197. <https://doi.org/10.2307/2710743>.

Washington, Mary Helen. “‘Disturbing the Peace: What Happens to American Studies If You Put African American Studies at the Center?’: Presidential Address to the American Studies Association, October 29, 1997.” *American Quarterly* 50, no. 1 (1998): 1–23.
<https://doi.org/10.1353/aq.1998.0005>.

Wise, Gene. “‘Paradigm Dramas’ in American Studies: A Cultural and Institutional History of the Movement.” *American Quarterly* 31, no. 3 (1979): 293.
<https://doi.org/10.2307/2712378>.

APPENDIX

Appendix A: Keywords in Keywords for American Cultural Studies

Abolition	Exceptionalism	Performance
Aesthetics	Family	Politics
Affect	Fashion	Prison
African	Finance	Property
America	Freedom	Public
Asian	Gender	Queer
Black	Gesture	Race
Body	Globalization	Racialization
Border	Government	Reform
Capitalism	Identity	Region
Citizenship	Immigration	Religion
City	Indian	Rural
Civilization	Indigenous	Science
Class	Interiority	Secularism
Colonial	Internment	Sentiment
Community	Islam	Sex
Contract	Labor	Skill
Coolie	Latino, Latina, Latin@	Slavery
Copyright	Law	Society
Corporation	Liberalism	Sound
Culture	Literature	South
Democracy	Market	Space
Dialect	Marriage	State
Diaspora	Media	Subject
Digital	Mestizo/a	Technology
Disability	Migration	Terror
Diversity	Modern	Time
Domestic	Nation	Visual
Economy	Naturalization	War
Empire	Neoliberalism	West
Environment	Normal	White
Ethnicity	Orientalism	Youth

Appendix B: Search terms used to develop Keyword counts in AntConc

abolition*
aesthetic*
affect*|feeling*|emot*|mood|atmosphere
africa*
america*
asia*
black|blacks|negro*
body|bodies|bodily
border*
capital*
citizen*
city|cities|urban*
civiliz*|civilit*
class|classes|caste|status
colony|coloni*|settler*
communit*
contract|contracts|contractual
coolie*
copyright*|intellectual property
corporat*
cultur*
democrac*
dialect*
diaspor*|homeland
digital*
disabl*
divers*
domest*
econom*
empire*|imperial*
environment*|ecolog*|wilder*|wildness
ethni*
exception*
famil*
fashion*
finan*
freedom*|free|freeing
gender*
gestur*
global*
govern*
ident*|personalit*
immigra*
Indian|indians
indigen*
interiori*
internment*
islam*|muslim*|arab*
labor|labors|laboring|laborer*|worker*|labour*
latin*
law|laws|legal*
liberal*
liter*
market*
marriag*|marry*
media*
mestiz*|mulatt*|chican*
migra*
modern*
nation*
naturaliz*
neoliberal*
normal*|abnormal*|norm*
oriental*|the orient
perform*
politic
prison*|jail*|penitentiary*|incarcera*|penal|carceral
propert*|own*
public*|privat*
queer*
race*|racist*|racism
racial*
reform*
region*
relig*
rural*|small town|agrarian*
scien*
secul*
sentiment*
sex*|heterosex*|homosex*
skill*
slave*
societ*|social|socially
sound*
south|southern*
space*|spatial*
state|states
subject*
techno*|tech

terror*
time|temporal*
visual*
war|wars|wartim*|warring|battl*
west*
white*
youth*|adolesc*|teen*