ARTICLE

What We Didn't Know a Recipe Could Be: Political Commentary, Machine Learning Models, and the Fluidity of Form in Nineteenth-Century Newspaper Recipes

Avery Blankenship[1]

[1] Northeastern University

In this article, I use document embedding models and a training set of nineteenth-century American recipes to build a pipeline classifier for identifying recipes in the broader nineteenth-century newspaper press. The model reveals a much more expansive understanding of the recipe form, which primarily centers around measurement words and prescriptive language rather than a heavily reliance upon the culinary. This fluidity of form allows nineteenth-century writers to harness the recipe form as a tool for political commentary all while no appearing to disrupt the careful divides between the public and domestic spheres. These recipe-adjacent texts, which are both recipe and not, offer a broader picture of short-form political commentary in the nineteenth century which can include genres and forms once thought unable to gestured beyond the confines of the kitchen.

Introduction

"Take three gallons of confidence; one peck of international agreement; one quart of gold standard; seven table spoonfuls of prosperity; two quarts, of sound money, and one pound of protection," instructs a recipe for "Confidence Pudding" from the March 13, 1897 edition of *The Dawn* submitted to the paper by a self-identified, unemployed mechanic, "Boil it with twelve months of no work for an hour and eat it while hot." Not quite a recipe, but not-*not* a recipe, this strange text triggers many questions, most pressing of which is: What is this doing in the newspaper? Simultaneously funny and commiserative, this not-quite-recipe takes a form that is instantly familiar—folkish, instructive, and vernacular in that way that recipes so often are, though food is hardly the focus. In the absence of actual culinary instruction, can we even call this recipe? If it isn't a recipe, then what else could it be? Whatever it may be, there is something useful to the writer about the recipe form. There is something about the recipe that the writer felt was a provocative and effective form through which to comment on the state of the nation and its particular impact on working-class citizens in 1897.[1] Recipes, this newspaper-writer implies, are uniquely useful political tools.

---

1 In 1897, the United States was in the midst of the Spanish-American war. During this time, the United States embraced the gold standard and increased involvement in international politics. The writer appears to be offering commentary on the nation's politics and also commenting on the impact these politics have on working-class people–namely unemployment.
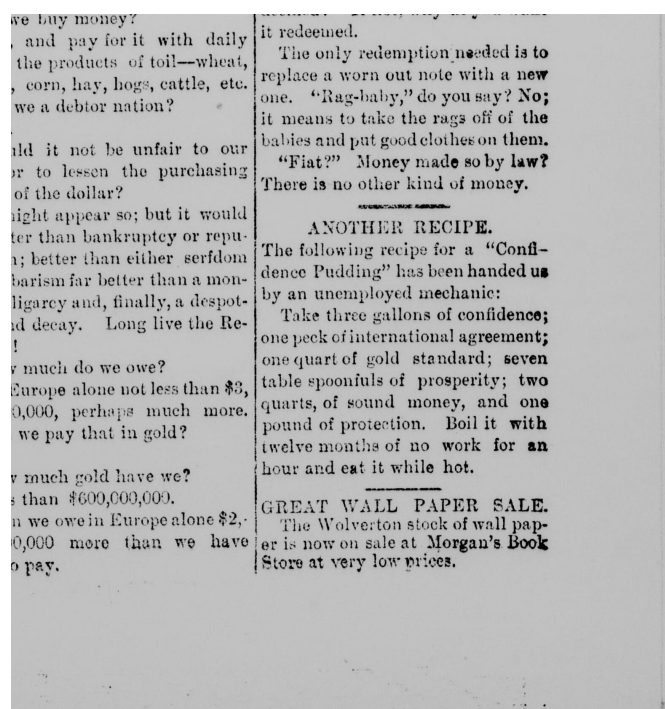
Recipe for Confidence Pudding from the March 13, 1897 edition of The Dawn

While the writer could have chosen to formulate this piece as an essay or even narrative, the use of the recipe form allows the writer to both joke with the reader and offer social commentary in a vernacular and commiserate way, all while adopting a form that perhaps a foreman or a business person might have written off and overlooked while reading the paper. The form allows the writer to speak to readers in the language of the everyday, the non-threatening, the *domestic*. Through this invocation of the recipe form, the mechanic's political viewpoint is presented to the reader in a way that is inoffensive—easy to chew and swallow down. For contemporary readers, the recipe for "Confidence Pudding" reveals two things. One, that nineteenth-century people considered the recipe form to be quite flexible and not confined to the purely culinary or domestic.[2] Second, that the dual tone that recipes can so easily take on allows writers to trouble the status-quo while not appearing to do so: to comment on politics and social issues from the safety of a form that some readers may have skimmed over altogether at first glance.

While the study of commercial cookbooks and manuscript recipes has increased dramatically in the last decade, nineteenth-century newspapers have always presented a problem to scholars of nineteenth-century domestic print

---

2 Other scholars have argued for the flexibility of the recipe form and its usefulness across generations of a family. For example, Elaine Leong writes in "Collecting Knowledge for the Family: Recipes, Gender and Practical Knowledge in the Early Modern English Household." *Centaurus*, vol. 55, no. 2, 2013, pp. 81–103. *Wiley Online Library*, https://doi.org/10.1111/1600-0498.12019 that recipes and recipe books in early modern europe reflect the needs and interests of particular families. Similarly, Madeline Shanahan writes in *Manuscript Recipe Books as Archaeological Objects: Text and Food in the Early Modern World*. Lexington Books, 2014 that manuscript recipes represent generational knowledge and familial practices in domesticity. My intervention here is to focus on the usefulness of the recipe form beyond the household—and in particular its usefulness to nineteenth-century people who otherwise might not have considered the recipe a particularly meaningful genre for political commentary if not for the flexibility of its form.

culture. Although we know that newspapers often published domestic writing, these pieces are often only a single paragraph or even just a few lines in length. When faced with the massive quantity of text that may appear on any given nineteenth-century newspaper page, locating a single paragraph or two of interest is akin to finding a needle in a haystack. In many ways, working with nineteenth-century newspapers requires that you already know what you are looking for in the paper when the articles often don't include titles, section headings, or other forms of delineation. Ultimately, the result of this foreknowledge is that there is a world of text—stories, narratives, and lives—that we pass over without even knowing.

My purpose in this article is twofold. First, I introduce a workflow designed to tackle that haystack of newspaper text by designing a needle-detecting pipeline of nineteenth-century newspapers. This workflow uses word embedding models and document vectors to automate the process of identifying recipes in newspaper data by turning the task into a classification problem. Using this model, I am able to identify texts with shared semantic relationships which can lead to grouping them with texts I already know are recipes. Second, because word and document vectors prioritize semantic relationships, the model allows me to explore the fluid space between recipes and not-recipes based on similarities in contextual use. I analyze how instances of these not-quite-recipes, what I call "recipe-adjacent" texts, in the nineteenth-century newspaper demonstrate both the flexibility of the recipe form and its ability to offer social commentary in ways that do not immediately appear to trouble the status-quo. Recipe-adjacent texts offer an invocation of the recipe form that is intimately intertwined with the context of the newspaper. Formally printed and published cookbooks showcase the author or editor's idea of what cooking is or should be, and thus any political commentary is to be understood within this context. The recipe-adjacent text, however, exists within the broader context of the newspaper, placing it in conversation with both the culinary and political texts that surround it, with the rapid fire speed with which the newspaper is published in the period, and most importantly with newspaper readers. This article thus argues for a reshaping of what recipes are and for what reasons they are written as well as a broader consideration of the political possibilities of the recipe form.

Domestic Genres

The question of genre and form is one that complicates the study of nineteenth-century domestic writing and recipes in particular. While recipes today may typically follow a more formulaic structure, recipes in the nineteenth century could take many forms (lists, prose, poetry, pedagogy, etc.) and appear just about anywhere: in diaries, pamphlets, magazines, newspapers, books, scrapbooks, and even stray bits of paper. Zboray and Zboray (2009) write that the fluidity of genre in the nineteenth-century domestic sphere offers "a glimpse into the structural relationship between writing as a practice and lived experience" (103). Nineteenth-century newspaper writers and readers
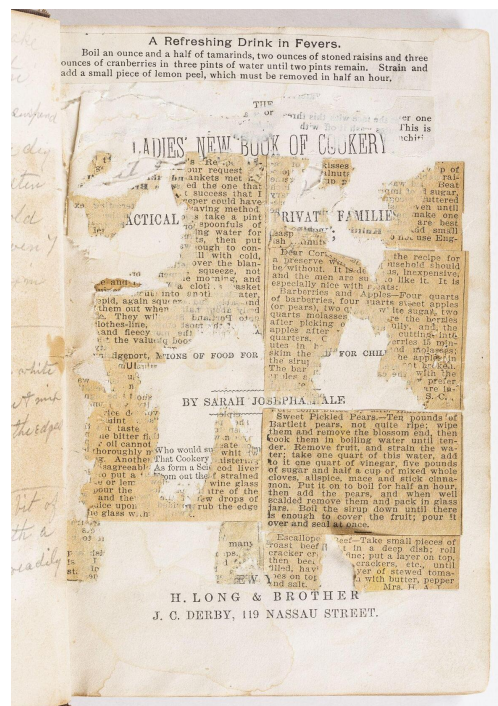
Figure 1. This copy of The Ladies' New Book of Cookery (1852) by Sarah Hale previously had newspaper recipes pasted over the title page, but these clippings were later ripped off of the page, possibly in an attempt to restore the original text. This copy is currently held by the American Antiquarian Society.

often push at the boundary lines between genres and forms, and in some ways, the obscurity of genre and form was part of the appeal of the newspaper.[3] The ways in which domestic writers in the nineteenth century transgress the boundaries between genre and form emphasize the particularities of nineteenth-century domestic writing, and recipes in particular, which often straddle the line between formalism and folklore.

However, while generic and formal fluidity is a common feature of nineteenth-century newspaper recipes, these very same writerly transgressions can make locating them in libraries and archives complicated. Printed books were often repurposed as repositories for recipes clipped from newspapers, which were pasted over the original text.[4] When these texts then appear in institutional collections, the likelihood that they will be categorized as "recipe books" is understandably low. In some cases, the newspaper clippings may have even been removed at some point to reveal the text underneath (Figure 1). As a result, while commercial, printed cookbooks are relatively straightforward to locate—if they have been preserved—other forms of texts that were utilized as cookbooks or were modified by their users remain much more difficult to pinpoint.

---

3 See Roggenkamp, Karen. *Narrating the News: New Journalism and Literary Genre in Late Nineteenth-Century American Newspapers and Fiction.* The Kent State University Press, 2005. Roggenkamp argues that the lines between factual reporting and fictional storytelling were often obscured in the newspaper. Distinguishing between was was real and imaginary was part of the appeal of penny papers, in particular (3-5).

4 See *Writing with Scissors: American Scrapbooks from the Civil War to the Harlem Renaissance* (2012) by Ellen Gruber Garvey

In many ways, the slipperiness of domestic genres defies scholarly taxonomies. When domestic writing appears embedded within the hybrid framework of the nineteenth-century newspaper, even simple categorization becomes a perplexing task. Diedre Lynch argues that studying handmade domestic texts such as albums or scrapbooks shifts our attention away from what these texts are and towards how they work or how they connect readers and writers to one another (91-93). However, categorization is crucial to the study of historical texts. Categorizing texts allows them to be searchable, retrievable, and thus studyable. However, as I mentioned, part of the difficulty of working with ephemeral recipes in newspapers, is how scattered they are across both the newspaper as well as in other places such as diaries or scrapbooks. "There is no way of ever getting access to the past," Geoffrey Bowker and Susan Star write in *Sorting Things Out* (2000), "except through classification systems of one sort of another—formal or informal, hierarchical or not" (41). Shifting the focus of the study of newspaper recipes to how they work is important, but in order to do so, we must first locate them at all and expand the methods that we use to do so.

In *Data Feminism* (2020), Catherine D'Ignazio and Lauren Klein emphasize that "what gets counted counts." The decision to include or exclude an item from a data set is in itself a form of categorization—one that asserts that that item does or does not belong. Similarly, Digital Curator for the Smithsonian National Museum of African American History and Culture, Dorothy Berry, writes that what gets included in archives and libraries is what becomes the truth.[5] Although the newspaper recipes are *there* in the sense that they are intermixed with the millions of other pages of newspapers in collections such as *Chronicling America*, it is precisely their intermixedness which renders them as a brief hiccup of a presence to scholars.

Studying recipes in the nineteenth-century newspaper is also complicated by the fact that, despite how commonly they appear in various issues across the country, their position and style on the pages of newspapers is rarely fixed. Nineteenth-century newspapers differ significantly from how newspapers in the twenty-first century look and how they are read. Without defined column headings or, in some cases, delineation between articles, searching for a specific genre of writing within any particular issue of a newspaper often requires reading through each individual page. It is also not guaranteed that a recipe will include the word "recipe" in the text, meaning that many relevant clippings will slip through keyword searches. Broersma and Harbers argue that keyword searching is not even particularly useful to researchers working with historical newspapers unless the researcher is interested in specific dates or events (1151). With the Library of Congress' *Chronicling America* collection, in particular, the best match for a keyword search is not guaranteed to be listed as the

---

5 Berry, Dorothy. "The House Archives Built." up//root. Accessed July 14, 2022. https://www.uproot.space/features/the-house-archives-built.

first result of a search, for instance, when a keyword has multiple uses and meanings or when the selected keyword is too vague. As Deacon (2007) notes, "key word searching is best suited for identifying tangible 'things' (i.e. people, places, events and policies) rather than 'themes' (i.e. more abstract, subtler and multifaceted concepts)" (8). With recipes, this becomes somewhat tricky. For instance, the word "receipt" used to mean "recipe," but its usage and definition has changed over time—though not all at once. As a result of these limitations, locating recipes by hand in a nineteenth-century newspaper until the researcher is too exhausted with text to continue sifting has thus far been the primary mode of studying them.

Word and document embedding models are equipped to tackle such problems by focusing on the classification of a type of document rather than strictly its content. Johanna Drucker describes distant reading methods as "ways of viewing the cultural record that are as useful as seeing the earth from space" (630).[6] The classification of genre, in particular, has been significantly addressed through the application of these methodologies.[7] Statistical modeling has shown to be a useful tool for classifying genre in a body of text. In *Distant Horizons* (2019), Ted Underwood advocates for the use of statistical modeling for genre detection, writing that "statistical models can embrace the contextual and perspectival character of historical interpretation, while giving scholars more control over the contexts and perspectives they provisionally adopt" (37). Underwood emphasizes, however, that genre can be a slippery concept—it means different things to different people, readers and scholars alike. While this indicates that no statistical model will ever be able to fully capture the entirety of a particular genre, their ability to emphasize context and semantic relationships makes the application of statistical modeling and machine learning methodologies to the identification of fluid genres like the recipe particularly useful.

Laura Nelson argues that machine learning, in particular, provides an "inductive framework" through which to analyze complex categories in an intersectional way. Machine learning's potential is realized when combined with inductive research—when the researcher and the model work together to inform one another. For example, D'Ignazio and Klein argue that while categorization is an inevitable part of working with data, in instances where data complicates or pushes at the boundaries of systems of classification it is important to apply feminist frameworks in order to question the very systems of classification in the first place. This process of questioning requires that

---

6 See, for example, "A Matter of Scale" by Matthew Jockers and Julia Flanders, "The Literary, the Humanistic, the Digital" by Julia Flanders, and "Text: A Massively Addressable Object" by Michael Witmore.

7 See, for example, Sarah Allison, et al. "Quantitative Formalism: An Experiment." *Stanford Literary Lab*, 2011. and Schöch, Christof. "Topic Modeling Genre: An Exploration of French Classical and Enlightenment Drama." *Digital Humanities Quarterly*, vol. 011, no. 2, May 2017.

the researcher reflect not just on the model's accuracy, but how that model's categorization decisions reflect a broader framework of classification that may at times not be particularly useful.

Rather than view instances where the model is not behaving in an expected way as failures, through the application of ideas that Nelson, D'Ignazio and Klein present, these supposed "failures" can present useful opportunities for the researcher to reflect on their own systems of categorization and whether or not these systems are particularly suited to a corpus. For example, for this article, I trained a machine learning model using recipes from professionally published nineteenth-century cookbooks. When that model then makes a mistake in classifying a recipe—such as passing over what a human would consider to be a recipe or including texts that we would not typically consider to be recipes—those mistakes don't necessarily reflect a flaw in the model. Rather, those mistakes offer a lens through which to reevaluate how we understand the genre of recipe, which is often mistakenly understood as a siloed genre, or that a book-recipe and a newspaper-recipe are the same thing. Long and So (2016) write that "there is more to be gained by explicating the singular aspects of a text, or by describing how it deviates from an assumed normative model, rather than by trying to define the model itself" (241). Similarly, I take misclassifications—or what I would have once perceived as misclassifications—as opportunities to enrich my work rather than deter it. Long and So also write that the act of categorizing literary texts in itself "also fundamentally determines the outcome of analysis and requires that a set of internally diverse texts be ostensibly pigeonholed into a limited number of categories" (252). They go on to argue, however, that this practice can open up the analysis to consider the boundaries of the categories assigned to a text, and whether or not those boundaries are usefully recognized by a machine. Conversations about digital humanities methods tend to unequally privilege technical disciplines such as statistics and computer science as granting legitimacy to literary or humanistic scholars—robustness of method, more "correct" numbers, proper implementation.[8] This view of digital humanities work limits the possibilities that technical disciplines may have to be shaped by work in literary studies, instead. Namely, that robustness of method is not always the goal as much as trying to make sense of a complicated, and often contradictory, literary landscape which often defies rules and struggles against categorization.

Recently, digital humanist Gabi Kirilloff has argued that rather than focusing on computation and methods as a deliverable, that digital humanists should instead view computation as a type of context building like archival work.[9]

---

8 See for example Da, Nan Z. "The Computational Case against Computational Literary Studies." *Critical Inquiry*, vol. 45, no. 3, Mar. 2019, pp. 601–39. *www-journals-uchicago-edu.ezproxy.neu.edu (Atypon)*, https://doi.org/10.1086/702594.

9 See Gabi, Kirilloff. "Computation as Context: New Approaches to the Close/Distant Reading Debate." *College Literature*, vol. 49, no. 1, 2022, pp. 1–25. *Project MUSE*, https://doi.org/10.1353/lit.2022.0000.

The field's preoccupation with methodological validity and innovation can at times stand in the way of innovative interpretation. Kirilloff argues that even incorrect data or flawed methods can lead to insightful interpretation. I would add that the pressure to be an inventor in digital humanities—to both invent methods and to invent new readings—distances researchers from one another and places us in the very academic silos that Nelson argues underutilize the full potential of computational methods. Thus, the purpose of the methodological and analytical decisions I outline in this article is less to invent a new, or even entirely valid, way of classifying recipes and instead to think about why classifying recipes is so difficult to begin with and how these difficulties can surface unique features of recipes that may otherwise go overlooked. In this article, I consider my machine learning model to be less of a tool which can itself be classified as either effective or ineffective and more a type of perspective. When the model misclassifies recipes, those misclassifications can be used to, instead of criticizing the model, turn that criticism inward in order to ask myself why I believe those misclassifications were made in error. In this way, I use misclassifications not to tinker with the model until it reads recipes the same way I do, but to instead reconfigure my own understanding of what a recipe is.

Methodology
The Digitized Newspaper

Before scholars can conduct a computational analysis of a newspaper, the newspaper must undergo various transformations to shift it from analog to digital form. Because I approach my work from an ethics of transparency, I believe that it is essential to not only unpack the effect that this transformation has on the data, but also its impact on our understanding of the nineteenth-century newspaper. As Lauren Klein powerfully writes:

> Among the greatest contributions of the digital humanities is its ability to illuminate the position of the critic with respect to his or her archive of study, and to call attention to the ethical and affective as well as epistemological implications of his or her methodological choices. (672)

As an essential first step to outlining my methodology, I will acknowledge some of the perspectival limitations inherent in working with nineteenth-century newspapers as digital objects.

Algee-Hewitt et al. describe a corpus as "smaller than the archive, which is smaller than the published: like three Russian dolls, fitting neatly into one another" (2). Nineteenth-century newspapers are key examples of this analogy: published newspaper articles are like an ocean and digital collections such as *Chronicling America* constitute only a cove within that ocean. Decisions regarding which newspapers should be preserved and whether they will be digitized are impacted by many factors, including funding, equipment

availability, staff, and the quality of the paper itself. As Lyneise Williams discusses, not only should texts be understood as mediators, but the technology used to preserve and access those texts should also be approached as a representational tool.[10] For newspapers, this means that both the newspaper and its digitized form are social mediators that prevent us from accessing nineteenth-century society directly.

Digital archives, Katherine Bode argues, "are interpretive constructs," (83) and "neither the analog record nor the digital one offers an unmediated and comprehensive view of the literary-historical record; both are partial, and not necessarily in complementary ways" (96). An understanding of these varying degrees of scale and mediation are essential to the study of nineteenth-century newspapers, allowing researchers to shift analysis away from totalizing statements about nineteenth-century society and towards analytic precision. Any conclusions one can draw about a body of newspapers are in actuality conclusions about a particular body of newspapers preserved by an institution and represented digitally. Ryan Cordell argues that the digital instantiation of an analog textual object constitutes an entirely new object and should be treated as such.[11] Digital representations of historical texts must be approached as archival objects in their own right, which means that they come with the same limitations and impacts of decision-making that physical archival objects do. While digital scans may increase access to archival material, that access is not even or without its own set of politics.

Benjamin Fagan points out that the absence of many Black newspapers in open-access databases, as well as *Chronicling America*'s limited representation of the Black press, means that conclusions drawn from the study of *Chronicling America*'s collection of digitized newspapers apply very specifically to white newspapers and only those that have been preserved.[12] Because my newspaper data is sourced from the *Viral Texts* project database,[13] which sources many of its newspapers from *Chronicling America* and *Gale Historical Newspapers*, any conclusions I can draw about recipes in newspapers are similarly impacted. However, while the glimpse of domestic writing offered by *Chronicling America* may reflect a very white, American perspective, the collection does include many regional papers, which makes studying middle- and working-class domestic writing much more fruitful. While larger papers in cities included recipes sparingly, regional newspapers are much more likely to include recipes and other forms of domestic writing with regular publication.

---

10 See Williams, Lyneise. "What Computational Archival Science Can Learn from Art History and Material Culture Studies." 2019 *IEEE International Conference on Big Data (Big Data)*, 2019, pp. 3153–55. IEEE Xplore, https://doi.org/10.1109/BigData47090.2019.9006527.

11 See Cordell, Ryan. "'Q i-Jtb the Raven': Taking Dirty OCR Seriously." *Book History*, vol. 20, no. 1, 2017, pp. 188–225. Project MUSE, https://doi.org/10.1353/bh.2017.0006.

12 Fagan, Benjamin. "Chronicling White America." *American Periodicals: A Journal of History & Criticism*, vol. 26, no. 1, 2016, pp. 10–13.

13 See Smith, David A., et al. "Computational Methods for Uncovering Reprinted Texts in Antebellum Newspapers." *American Literary History*, vol. 27, no. 3, Sept. 2015, pp. E1–15. Silverchair, https://doi.org/10.1093/alh/ajv029. And Ryan Cordell and David Smith, *Viral Texts: Mapping Networks of Reprinting in 19th-Century Newspapers and Magazines* (2022), http://viraltexts.org.

The collection shifts the attention away from being centered on urban life and towards the day-to-day which is best captured in rural papers. This perspective is especially important to the study of recipes given their often vernacular or folkish nature.

Acknowledging the many ways in which the data I work with has already been mediated before I have encountered it is an essential move in feminist data practice as D'Ignazio and Klein outline.[14] This mediation shifts the conclusions that I can draw from my analysis towards more precision. For instance, while I may be able to analyze the recipes present in the newspapers available through *Chronicling America*, there are millions of newspaper pages that are not part of that collection. The absences of these data are particularly pertinent to the study of domestic writing, which has historically itself been limited by gaps in available records.

Messy OCR and its Impact on Classification

Optical Character Recognition (OCR) is a technology designed to automatically recognize text and transcribe digitized textual objects to make them searchable. Though OCR is commonly used in newspaper databases to improve discoverability, newspapers present a number of challenges to automatically detecting text. As Smith et al. point out, OCR makes frequent mistakes when applied to historical newspapers due to the difficulty of navigating around damaged or worn pages.[15] Strange et al. further note that small-scale, regional papers tend to produce poorer quality scans as a result of a combination of variable historical and contemporary resources such as low quality ink or type and a lack of access to preservation equipment. As a result, metropolitan papers tend to be cleaner and more receptive to OCR, while rural papers are likely to contain many more mistakes.[16] These mistakes and variations in OCR can be useful, as Cordell has argued, in examining circulation in the nineteenth-century periodical press. However, the quality of OCR can also negatively impact classification tasks.[17] Strange et al. explain that the degree to which OCR errors can impact text mining depends on the task at hand, as "tasks of sentence boundary detection, tokenization, and part-of-speech tagging on text are all compromised by OCR errors" (para. 17).

---

14 See D'Ignazio, Catherine, and Lauren F. Klein. *Data Feminism*. MIT Press, 2020.

15 See Smith, David A., et al. "Computational Methods for Uncovering Reprinted Texts in Antebellum Newspapers." *American Literary History,* vol. 27, no. 3, Sept. 2015, pp. E1–15. Silverchair, https://doi.org/10.1093/alh/ajv029.

16 See Strange, Carolyn, et al. "Mining for the Meanings of a Murder: The Impact of OCR Quality on the Use of Digitized Historical Newspapers." *Digital Humanities Quarterly*, vol. 008, no. 1, Apr. 2014.

17 See Cordell, Ryan. "'Q i-Jtb the Raven': Taking Dirty OCR Seriously." *Book History*, vol. 20, no. 1, 2017, pp. 188–225. *Project MUSE,* https://doi.org/10.1353/bh.2017.0006.

Furthermore, OCR errors can impact language modeling tasks, such as topic modeling, which has been popularly used for downstream classification tasks.[18] As Mutuvi et al. note, the impact of OCR "noise" depends on the task being performed. For example, while OCR errors have minimal impact on certain NLP tasks, the same level of "noise" produced by OCR errors can adversely affect methods such as topic modeling (5). This, in turn, impacts the quality of the topics derived from the task. Since topic modeling, a statistical method for determining the most frequent topics in a text, focuses on addressing content rather than semantic relationships, the method was not ideal for the purposes of determining such a slippery genre as recipes. As the recipe for "Confidence Pudding" demonstrates, recipes in the nineteenth century span across topics, meaning that a topic classification focused on food as a mainstay of the genre may overlook a broad swath of recipes. The reverse is also true: not all writing in newspapers on the topic of food or food preparation are recipes. As a result, classification of topical genres in the newspaper would not necessarily capture the recipe as a genre or form. Instead, I needed a method that would successfully capture the semantics and context of newspaper writing with as little negative impact from OCR noise as possible.

Word vectors have steadily grown more popular in the digital humanities. With tutorials,[19] theoretical breakdowns,[20] and entire institutes dedicated to word vectors and big data,[21] there is no shortage of resources for getting started. While word vectors, like topic models, are adversely impacted by OCR errors, it is possible to obtain robust results with their usage.[22] In a study on the impact of OCR noise on language models, Todorov and Colavizza found that although the impact of OCR errors on end results is significant, this performance is not consistent across methods. For example, transformer-based algorithms such as BERT tend to be outperformed by 'simpler' models, including those trained using Word2Vec. Todorov and Colavizza further explain that Word2Vec is more resilient to OCR noise in the presence of smaller datasets (7). Since even a large dataset of recipes is guaranteed to have fewer words than, say, a dataset of novels, this made Word2Vec suitable for

---

18 See for example, Walker et al., Daniel, et al. "Evaluating Models of Latent Document Semantics in the Presence of OCR Errors." *Faculty Publications*, Jan. 2010, https://scholarsarchive.byu.edu/facpub/1648., Todorov, Konstantin, and Giovanni Colavizza. *An Assessment of the Impact of OCR Noise on Language Models*. arXiv:2202.00470, arXiv, 26 Jan. 2022. *arXiv.org*, http://arxiv.org/abs/2202.00470., and Dutta, Haimonti, Mutuvi, Stephen, et al. "Evaluating the Impact of OCR Errors on Topic Modeling." *Maturity and Innovation in Digital Libraries*, edited by Milena Dobreva et al., Springer International Publishing, 2018, pp. 3–14. *Springer Link*, https://doi.org/10.1007/978-3-030-04257-8_1., and Aayushee Gupta. "PNRank: Unsupervised Ranking of Person Name Entities from Noisy OCR Text." *Decision Support Systems*, vol. 152, Jan. 2022, p. 113662. *ScienceDirect*, https://doi.org/10.1016/j.dss.2021.113662.

19 See Ben Schmidt's *Vector Space Models for the Digital Humanities*. http://bookworm.benschmidt.org/posts/2015-10-25-Word-Embeddings.html. Accessed 23 Apr. 2022.

20 See Ryan Heuser's http://ryanheuser.org/word-vectors-2/. Accessed 23 Apr. 2022.

21 See the Women Writers Project's https://www.wwp.northeastern.edu/outreach/seminars/neh_wem.html

22 See Dutta, Haimonti, and Aayushee Gupta. "PNRank: Unsupervised Ranking of Person Name Entities from Noisy OCR Text." *Decision Support Systems*, vol. 152, Jan. 2022, p. 113662. *ScienceDirect*, https://doi.org/10.1016/j.dss.2021.113662.

developing my model. The handling of OCR errors as well as the stability of the results, makes Word2Vec and word embedding models a valuable method for classifying OCR'ed newspapers.

Word2Vec and Doc2Vec

My classifier was built by first training a Doc2Vec model on a dataset of recipes that I individually selected from published, nineteenth-century cookbooks.[23] In 2013, Mikolov et al. et al. introduced Word2Vec, a model for representing words as vectors using a neural network.[24] Word-embedding models represent the relationships between words in a corpus with each word occupying a point in space some distance from related words, and this location is represented by what is called a "vector," or a numerical representation of that word. As a result, word embedding models can model the semantic relationships between words based on their closeness or distance in a shared semantic space.

In 2015, Mikolov et al. et al. proposed a modification of Word2Vec titled Doc2Vec, which models semantic relationships at the document level rather than the word level.[25] For example, rather than focus on individual words as their own entities, with Doc2Vec words could be linked together in sentences, paragraphs, or several paragraphs. The model assigns vectors to individual words in a paragraph, and then assigns the entire paragraph a vector which allows the model to understand paragraphs as joint entities which are made up of individual words.[26] With training, the paragraph vectors can be used for classification tasks using pre-determined labels that the researcher provides the model. In this way, the model comes to learn what the vector of a particular document is, its relationship to other documents in the corpus, and its particular relationship to documents in categories with shared labels. This method is particularly useful for the exploratory research in large corpora of text since it maintains the unsupervised way that word embedding models uncover the relationships between documents while allowing for the semi-supervision of relating those relationships to categories of text that the researcher has already identified.

---

23 I sourced the initial training data from cookbooks available through *Project Gutenberg*. I used only cookbooks published in the nineteenth-century and in America in order to more closely match the newspaper data. The initial training set included one thousand recipes. The data is available on GitHub. https://github.com/ViralTexts/nineteenth-century-recipes

24 See Mikolov, Tomas, et al. "Distributed Representations of Words and Phrases and Their Compositionality." *Advances in Neural Information Processing Systems*, vol. 26, Curran Associates, Inc., 2013. *Neural Information Processing Systems*, https://proceedings.neurips.cc/paper/2013/hash/9aa42b31882ec039965f3c4923ce901b-Abstract.html.

25 See Le, Quoc V., and Tomas Mikolov. Distributed Representations of Sentences and Documents. arXiv:1405.4053, arXiv, 22 May 2014. arXiv.org, http://arxiv.org/abs/1405.4053.

26 The model understands and links paragraphs by learning the vectors for each individual word in a paragraph as well as the vector for each paragraph. The model then averages those two vectors in order to predict what the next word in context might be using the skip-gram algorithm.

Corpus Description and Doc2Vec Implementation

In the absence of a robust set of nineteenth-century newspaper recipes to use as training data, I sourced my training data from commercial cookbooks published in the nineteenth century. While the machine learning community has increasingly become interested in analyzing recipes and developing recipe datasets, these studies tend to be focused on modern recipe analysis.[27] Though recipes tend to be passed down from generation to generation, using modern recipes, or anything published in the twentieth century, to develop a machine learning task for determining nineteenth-century recipes poses several challenges. For example, the end of the nineteenth century marks a shift in technologies of the kitchen. Early nineteenth-century recipes often include instructions for both cookstoves and fireplaces, with wood still accounting for ninety percent of the nation's fuel in the 1850s. By the end of the century, however, many more kitchens made use of coal-burning cookstoves.[28] Similar innovations such as the Dover Egg Beater and chemical leavening and baking soda, as well as the widespread availability of ingredients after the invention of the refrigerator, make recipes at the end of the century and beyond significantly different from recipes in the mid- or early-period. There, of course, also tends to be significant delays in technological uptake in rural kitchens or households with limited access to the supplies abundant in urban environments.

The recipe form and genre conventions also shifted significantly at the end of the period. Due to the rise of domestic science and scientific cookery at the end of the nineteenth century, recipes became much more standardized and cooking itself professionalized. Recipes from the mid to early nineteenth century are written in a more descriptive style, with instructions and calls for ingredients intermixed with prose. Recipes at the end of the period would look more familiar to cooks today, as their ingredients are placed at the top of the recipe with standardized measurements and the recipes themselves are longer and more detailed and often lack the more narrative or folklorish style more prevalent prior to the rise of domestic science.[29]

The final point of departure from contemporary recipes is the way nineteenth-century recipes cling to a manuscript and oral culture. As rhetorician Sarah Walden notes, Southern cookbooks are uniquely important to American

---

27 See for example, Jiang, Yiwei, et al. "Recipe Instruction Semantics Corpus (RISeC): Resolving Semantic Structure and Zero Anaphora in Recipes." *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, Association for Computational Linguistics, 2020, pp. 821–26. *ACLWeb*, https://aclanthology.org/2020.aacl-main.82., *Recipe1M+: A Dataset for Learning Cross-Modal Embeddings for Cooking Recipes and Food Images - MIT*. http://pic2recipe.csail.mit.edu/. Accessed 13 June 2022., and Popovski, Gorjan, et al. "FoodBase Corpus: A New Resource of Annotated Food Entities." *Database: The Journal of Biological Databases and Curation*, vol. 2019, Nov. 2019, p. baz121. *PubMed Central*, https://doi.org/10.1093/database/baz121.

28 For more detail on the transition from fireplace to cookstove, see Brewer, Priscilla J. *From Fireplace to Cookstove: Technology and the Domestic Ideal in America*. Syracuse University Press, 2000.

29 For more on the rise of the domestic science movement, see Neuhaus, Jessamyn. *Manly Meals and Mom's Home Cooking: Cookbooks and Gender in Modern America*. JHU Press, 2012., and Matthews, Glenna. *"Just a Housewife": The Rise and Fall of Domesticity in America*. Oxford University Press, 1989.
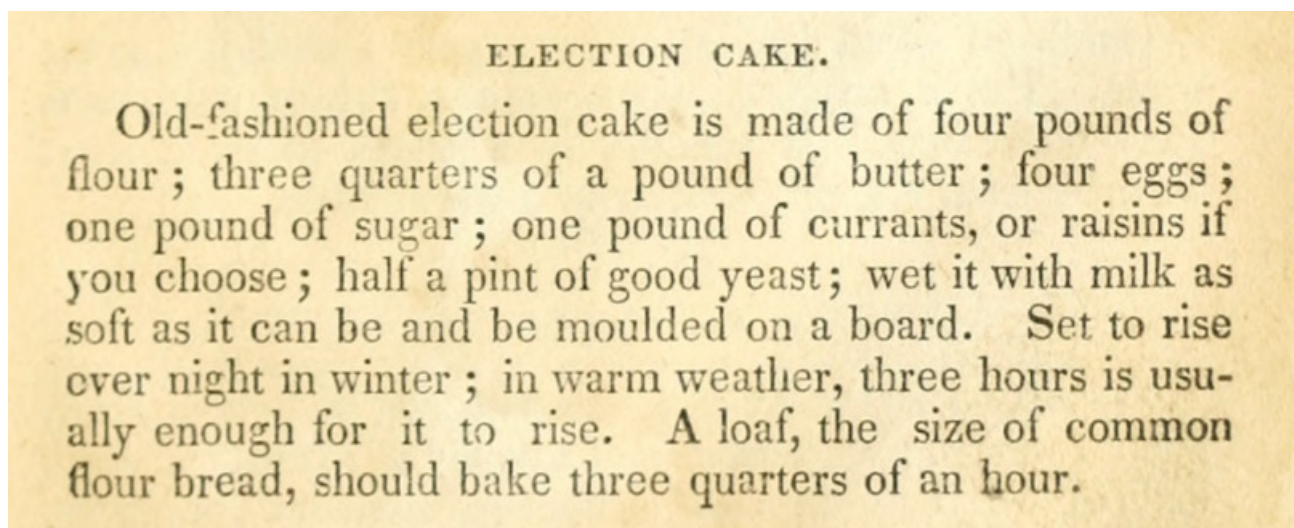
Figure 2. Recipe for Election Cake from Lydia Maria Child's The American Frugal Housewife (1838) courtesy of HathiTrust.

domestic print culture. Due to the South's rural and agrarian culture, cookbooks primarily remained in manuscript form until well into the nineteenth century. Few commercially published cookbooks were produced in the South and those that did were typically printed locally with few editions in circulation (83). Many nineteenth-century recipes reflect this tie to oral or manuscript culture through vague instructions to "sweeten to taste" or "bake in the usual way," which assume prior knowledge or universal training on the behalf of the reader. Figures 2 and 3 show two recipes for the same cake, one published at mid-period and one at the end of the nineteenth century. The recipes, when read together, demonstrate how the form and generic conventions evolved rather quickly in the nineteenth century, shifting from a more narrative style to a formulaic style.

My initial training set consisted of one thousand published recipes that I sourced from books available through Project Gutenberg, and only those published in the United States.[30] This training data is available, open-source, on Github.[31] Notably, the model could not successfully identify newspaper recipes using the dataset of commercially published recipes alone. However, as I added newspaper recipes to the training data and retrained the model, its accuracy increased significantly.

Using Document Vectors in a Classification Problem

In this section, I will describe the general technical workflow I used for training my Doc2Vec model as well as the logistic regression model for classification. While I will use correct technical terms, so that readers familiar with such

---

30 https://www.gutenberg.org/ebooks/bookshelf/419

31 Blankenship, Avery. "A Dataset of Nineteenth-Century American Recipes," Viral Texts: Mapping Networks of Reprinting in 19th-Century Newspapers and Magazines. 2021. https://github.com/ViralTexts/nineteenth-century-recipes/

Election Cake.

½ cup butter.
1 cup bread dough.
1 egg.
1 cup brown sugar.
½ cup sour milk.
⅔ cup raisins seeded and
    cut in pieces.

8 finely chopped figs.
1¼ cups flour.
½ teaspoon soda.
1 teaspoon cinnamon.
¼ teaspoon clove.
¼ teaspoon mace.
¼ teaspoon nutmeg.

1 teaspoon salt.

Work butter into dough, using the hand. Add egg well beaten, sugar, milk, fruit dredged with two table-spoons flour, and flour mixed and sifted with remaining ingredients. Put into a well-buttered bread pan, cover, and let rise one and one-fourth hours. Bake one hour in a slow oven. Cover with Boiled Milk Frosting.

Figure 3. Recipe for Election Cake from Fannie Farmer's The Boston Cooking School Cook Book (1901) courtesy of HathiTrust.

methods can assess my work, I will explain each step for readers less familiar with these methods. After compiling my initial set of training data, I trained a Doc2Vec model using Python's Gensim library which includes a fast and robust instantiation of Doc2Vec. I began by importing my recipe training data into a dataframe, which is similar to a spreadsheet in Python, and adding a column to tag each of the recipes as "recipe." Adding the "recipe" column ensures that when the model is being trained, that it has access to both the text of each recipe and the label which will later identify the text as a recipe to the logistic regression model. I then pulled in hand-tagged generic training data from the *Viral Texts* project, which is available on Github which includes examples of news, advertisements, poetry, and literature all pulled from newspapers.[32] This hand-tagged training data was pulled into a single dataframe and tagged "not recipe." With this done, the text in my corpus was either text that belonged to the "recipe" category or text that belonged to the "not recipe" category. Once all of the data was assigned a category label, I "cleaned," or standardized it, by lowercasing all of the words and removing punctuation, which prevents words with varying punctuation and capitalization from being mistaken for distinct words. I did not correct any OCR errors given the research described above.

After cleaning the data, I trained a Doc2Vec model. When a Doc2Vec model is trained using binary tags—with only two labels—document vectors are only generated for each label rather than for each individual document, resulting in

---

32 https://github.com/ViralTexts/viral-texts-classification/tree/master/hand-tagged

a much smaller model. Ultimately, this means that my model was not interested in learning the distinct features of each individual document in a particular category, but more so understanding that features the documents in that category share. This method is particularly efficient for downstream classification problems because the Doc2Vec model has the ability to infer a vector for new, or unseen, documents and assign it a label based on how similar that document is to other documents within one of the two categories. This means that Doc2Vec, since it is trained to understand what features documents in a particular category share, is able to take new documents, analyze their features, or determine their vectors, and then place these new documents in a location in vector space that is close to other documents in the same category which is a workflow that Word2Vec, since it considers individual words, is not able to do. With the binary model trained, I then used the model to pipeline into training a logistic regression model. By pipeline, I mean that I took the Doc2Vec model and its vectors which represents the contextual or semantic relationships between documents in two classes, and used that data as training data for a logistic regression model. The logistic regression model is what would then allow me to introduce newspaper texts that the model has not seen before in order to determine which category the text belongs to.

After developing the classification pipeline, I sampled one thousand random clusters, or groupings of newspaper articles, from the *Viral Texts* newspaper corpus so that I could introduce newspaper recipes to my model and retrain it as I increased the training data to include more of these recipes. In order to ensure that at least one item in that random sampling would include a recipe, I sampled one thousand random clusters that had at least one occurrence of the word "recipe." These random samples were then given to the logistic regression model to classify as either "recipe" or "not recipe." However, although I required the appearance of the word "recipe" in this initial sampling in order to ensure that there would be some recipe representation that I could then use to introduce my model to more newspaper recipes, the model was also successful at categorizing recipes which did not include the word "recipe." The initial model, trained on printed and commercialized recipes, only achieved an accuracy score of approximately fifty percent, meaning it was only accurately classifying recipes given a random sample of newspaper data fifty percent of the time. Following the classification, I worked through the clusters myself and corrected any major misclassifications. This new newspaper data was introduced to the Doc2Vec model and used to retrain to increase accuracy. As I added more and more newspaper recipes to the training data, the logistic regression model became significantly more adept at classifying recipes. I was eventually able to achieve an accuracy score of seventy-one percent with a final training set of 2,110 recipes—over half of which were newspaper recipes.

Challenging Genre with Recipes

That the model was not particularly successful at identifying recipes from a training set of newspaper recipes alone, but improved dramatically once newspaper recipes were incorporated into the training data, suggests two things. The first is simply that the model became better at identifying recipes because the newspaper recipes added to the training set included similar OCR errors to those found in the random clusters it was then asked to identify. Second, this indicates that there is some generic or formal difference between recipes that appear in the newspaper and recipes that appear in published books. My initial goal in developing the classification pipeline for the *Viral Texts* newspaper corpus was simply to locate recipes. The final model, which has a seventy percent accuracy rate, does serve as an efficient way to pull the somewhat elusive texts out of mountains of newspaper pages. However, an unexpected result of this classification task was the sheer number of cases in which the model identified a text that may not be considered a recipe at first glance but certainly utilizes the recipe form. These recipes, which I refer to as "recipe-adjacent" texts, demonstrate both the flexibility of the recipe form as well as its usefulness beyond the domestic sphere. These recipes also exemplify the usefulness of computational classification for identifying generic adaptation and fluidity.

Sari Edelstein writes that women writers of the nineteenth century used the newspaper to "telegraph their engagement with the sociopolitical world" (6). Many of these writers, as I will demonstrate, also recognized the power of domestic forms such as the recipe. Rather than using the masculine-associated essay form, women writers used the recipe to disseminate their political opinions, allowing them to challenge social norms from the safety of a genre associated with domesticity and the home. Pulling double duty, recipe-adjacent texts operate as both socially acceptable jests and political prescription. For example, Figure 5 depicts a recipe titled "A Recipe for Cooking Husbands" that is attributed to "a Baltimore lady." This recipe, as identified by the model, adopts a teasing tone to describe how a wife should manage her husband, suggesting that a husband should be stewed, roasted, or pickled. However, beneath the teasing tone is an air of commiseration. The writer acknowledges the limited power that women have over their husbands in the home, warning women to not select their husband based solely on material wealth. By presenting this text as a recipe rather than a more traditional narrative or essay, the "Baltimore lady" is able to subversively communicate to other women the difficulty of gendered expectations in the home without appearing to transgress social expectations.
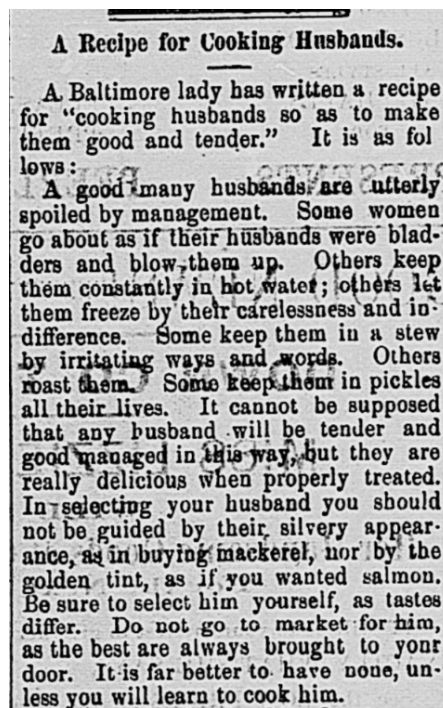
A Recipe for Cooking Husbands.

A Baltimore lady has written a recipe for "cooking husbands so as to make them good and tender." It is as follows:

A good many husbands are utterly spoiled by management. Some women go about as if their husbands were bladders and blow them up. Others keep them constantly in hot water; others let them freeze by their carelessness and indifference. Some keep them in a stew by irritating ways and words. Others roast them. Some keep them in pickles all their lives. It cannot be supposed that any husband will be tender and good managed in this way, but they are really delicious when properly treated. In selecting your husband you should not be guided by their silvery appearance, as in buying mackerel, nor by the golden tint, as if you wanted salmon. Be sure to select him yourself, as tastes differ. Do not go to market for him, as the best are always brought to your door. It is far better to have none, unless you will learn to cook him.

Figure 5. A recipe for cooking husbands. December 31, 1885 edition of the *South Carolina Anderson Intelligencer*

Existing scholarship on the topic of non-culinary recipes tends to focus on either their narrative style or their teasing tone.[33] I wish to bridge these conversations to address how the recipe form allows writers to both tease and prescribe actions to the reader through narrative. The appearance of these recipes in the nineteenth-century newspaper moves the recipe form out of the domestic sphere and into public discourse. Jennifer Rachel Dutch defines "joke recipes" as "vignettes of frustrated domesticity" and as an "identifiable form of folk humor" (251). Certainly, this recipe for cooking husbands fits Dutch's description. However, while Dutch primarily focuses on joke recipes that circulate amongst women readers through community or church cookbooks, removing a joke recipe from the context of a cookbook and circulating it to a broader newspaper audience introduces new levels of political engagement—a type of engagement that pushes domestic writing forms into public discourse and demonstrates the significance of the recipe form to those outside the kitchen. Juri Joensuu describes "literary recipes" as recipes that work "against our expectations by using the instantly recognizable form for nonculinary purposes" (579). He continues that infeasibility, "is an important factor that distinguishes literary recipes from real ones, even if the difference is always open to interpretation" (582). However, the infeasibility of literary recipes and the folk humor of joke recipes does not fully account for the ways in which the recipe form was taken up and adapted by writers in the newspaper. Accounting for the diverse applications of the recipe form in periodicals allows

---

33 See Joensuu, Juri. "Fictitious Meals, Culinary Constraints: The Recipe Form in Four Oulipian Texts." *Poetics Today*, vol. 42, no. 4, Dec. 2021, pp. 575–95. *DOI.org (Crossref)*, https://doi.org/10.1215/03335372-9356868 and Dutch, Jennifer Rachel. "Not Just for Laughs: Parody Recipes in Four Community Cookbooks." *Western Folklore*, vol. 77, no. 3/4, 2018, pp. 249–76.

us to understand the political potential of recipes and the ways recipes were used to both commiserate with readers and prescribe social change, particularly for middle- and working-class readers.

With the advent of the penny press in the 1830s, mass-produced newspapers made broader reading accessible to middle- and working-class readers at minimal cost. While newspapers became rapidly more affordable, printed books still remained costly. Printed and commercialized cookbooks, in particular, were written for an affluent audience. The cookbook reader was often a literate, middle-class housewife who read the recipes more than she implemented them. Megan Elias writes that "when she read cookbooks, then, the middle-class housewife did so as a reader, rather than as a worker" (15). She might step into the kitchen to prepare a dessert or two—cookbooks in the nineteenth century are heavily focused on indulgences such as desserts—but rarely are cookbooks written with servants or working-class people in mind. Instead, they are written in such a way that domesticity, or at least the domestic ideal, is not accessible to the working class. Domesticity, then, is positioned as stable, impenetrable, and unreachable. Because published cookbooks would not have been as accessible to the working class or the lower-middle class, the domestic ideal as configured in their recipes is mediated to them by the middle-to-upper-middle class mistress. While innovations in printing technology made cookbooks more widely available, much of the writing is still aimed at emulating a more affluent class.[34] Jessamyn Neuhaus writes that the trouble with studying prescriptive literature like cookbooks is that they can only account for the advice provided by authority figures, not what people actually did (3). Thus, the commercial cookbook represents the author or editor's idea of what American cooking *should* be as well as how a house *should* be managed. This aspiration may or may not have actually been put into practice.

Newspaper recipes, by contrast, defy the generic expectations of the cookbook recipe. While cookbook recipes are more focused on making an argument as to what the American domestic ideal looks and tastes like, the public platform of the nineteenth-century newspaper allows writers to position themselves as knowledgeable instructors of topics other than food. The recipe form as it is presented in the newspaper becomes fluid and flexible, allowing a particular type of expression that is not quite achievable through the use of other forms. These recipes also tend to be focused more on mutual aid: on helping the reader, often to save money, without emphasizing class differences or appearances. For example, <u>Figure 6</u> is a recipe for a cheap ice-chest which instructs the reader to construct an ice-chest using two dry-goods boxes, explaining that "for family use, this has proved quite as serviceable and as economical as more costly 'refrigerators'." The recipe, identified as such within the text and by the model, is focused not so much on instructing the reader as

---

34 See Neuhaus, Jessamyn. *Manly Meals and Mom's Home Cooking: Cookbooks and Gender in Modern America*. JHU Press, 2012.
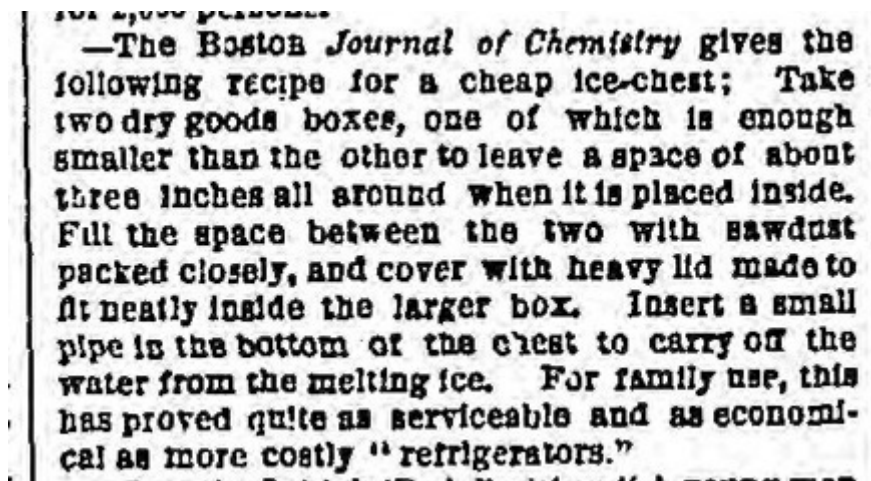
Figure 6. A recipe for a cheap ice-chest. July 4th, 1871 edition of the Chicago Tribune

to the most effective way to make a cheap ice-chest but more so on presenting the reader with an alternative to what the recipe calls "more costly 'refrigerators.'"

Just as the recipe for the ice-chest is focused less on positioning the writer as an absolute authority, the recipe in Figure 7, a recipe for making pencil writing indelible, provides mutual aid in the form of a simple, economical solution to a common problem, and particularly those writing to soldiers. This distinction in authority from the recipe for cooking husbands, which positions the writer as a knowledgeable instructor, is that the writers in Figure 6 and Figure 7 are positioned more as friends offering a suggestion for something that they have found useful. This distinction may largely be based on class and social rank. While the recipe for cooking husbands is clearly speaking to a more affluent group of women, women who are courted and who will manage a household, the recipe for the ice-box is more focused on saving money and the recipe for pencil writing on efficiency and waste. The writer describes the purpose of the recipe as providing a "process [that] will make lead pencil writing or drawing indelible as if done with ink." The recipe, which the writer describes as both an old and good one, is intended to assist those sending letters to soldiers, to ensure that their letters will not be damaged as letters often are in transit. The casual, conversational tone of the recipe suggests that the writer is less interested in commanding or monitoring the reader's behavior and more in offering friendly advice. This focus on casualty differs largely from published cookbooks in the period which are much more invested in training women to adhere to a domestic ideal. For example, Lydia Maria Child's *The American Frugal Housewife* (1829) and Mary Randolph's *The Virginia Housewife* (1824) which promote and help to establish New England cuisine and Southern cuisine respectively. These cookbooks are designed to be less advice manuals and more instructions for what it means to be a New England or Southern housewife. To defy their carefully laid out instructions is to defy domesticity.

How to Make Pencil Writing Indelible.

A correspondent of an agricultural paper gives the following information, which may be of service to some of our readers:

A great many valuable letters and other writings are written in pencil. This is particularly the case with the letters our brave soldiers send home from the army. The following simple process will make lead pencil writing or drawing as indelible as if done with ink: Lay the writing in a shallow dish and pour skimmed milk upon it. Any spots not wet at first may have the milk placed upon them lightly with a feather. When the paper is wet all over with the milk, take it up and let the milk drain off, and whip off with a feather the drops which collect on the lower edge. Dry it carefully, and it will be found to be perfectly indelible. It cannot be removed even with India rubber. It is an old recipe and a good one.
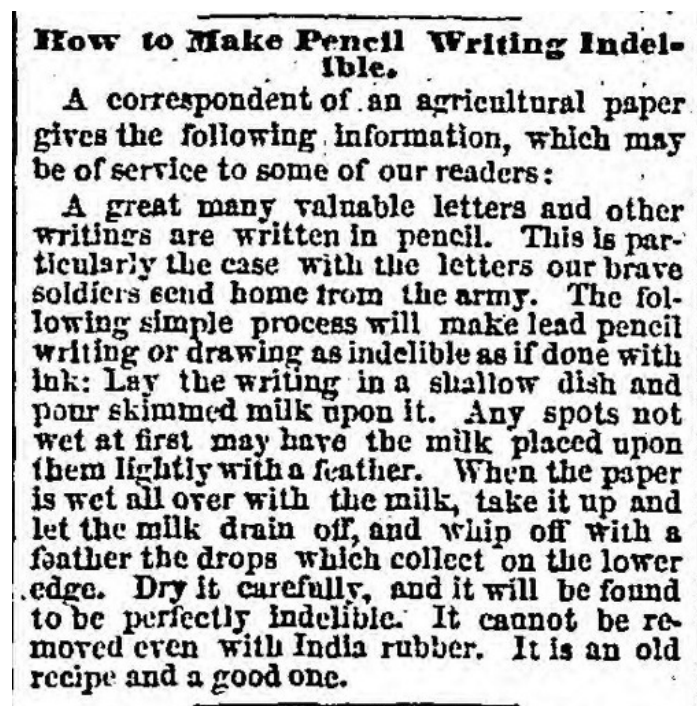
Figure 7. A recipe to make pencil writing indelible. December 24th, 1864 edition of the Chicago Tribune

These recipe-adjacent texts take up the recipe form to share friendly advice, advice that is removed from the vision of domesticity presented in published cookbooks though still presented within the context of a socially approved domestic writing form. The recipe form in these cases is both literary and common, allowing writers to establish rapport with readers and to provide context for the recipe. The ubiquitous nature of prescriptive writing, by contrast, positions the writer as an expert who is uniquely suited to the prescribed tasks. In cookbooks, this specialized knowledge is often wielded to prescribe the author's idea of ideal domesticity, while the newspaper recipe is much less invested in reifying social stratifications. This may be because the intended audience of a cookbook, a housewife who is likely responsible for a small fleet of servants, differs greatly from the audience of a newspaper, which includes working-class people seeking alternatives to new and costly technologies. The newspaper recipe is intended to be practiced by the reader, whereas the cookbook recipe is more likely to be read by a mistress and then dictated to a servant or enslaved person. These distinct readerships demonstrate the different kinds of knowledge that members of differing social classes have access to. The published cookbook emphasizes respectability and control over one's household, the author is positioned as an authoritative expert and the recipes are typically promised to be tried and tested to ensure their quality. The newspaper recipe, and particularly the recipe-adjacent text, opens up this authoritative form to members of other classes who otherwise would have found the cookbook inaccessible for many reasons. With access to the form through the newspaper, the form is opened up for intervention and its boundaries are pushed to include a more colloquial tone and sometimes even a distancing from domesticity.

> **To KILL A TOWN.**—The following is a good recipe for killing a town:
>
> "To kill a town, underrate every present and prospective public enterprise, speak ill of the churches and schools, tell everybody the hotels are 'bad,' enlarge the vices of the people, especially the young people; withhold the patronage from your merchants and tradesmen, and buy your goods and groceries at some other place; and by all means go to the city for your millinery and such like; never subscribe for the local paper, and if you are in business, refuse to advertise, and above all, if you have job printing to be done send it to a city office."

Figure 8. Recipe to kill a town. April 30th, 1873 edition of The Newberry Herald

Where the two forms intersect is in their ability to support social commentary. While the published cookbook is much more likely to offer commentary on what is respectable and appropriate in the home, with some exceptions such as *The Woman Suffrage Cook Book* (1886). Even the names of dishes—Election Cake, Lincoln Cake, etc.—point to a patriotic undertone. Recipe-adjacent texts, however, push social commentary outside of the home and towards larger, national politics. In [Figure 8](#), the recipe "To Kill a Town" adopts a tone that is both teasing and authoritative. The recipe describes how towns are destroyed, including steps such as driving business away to big cities and indulging the vices of young people. The instructions provided include: "underrate every present and prospective public enterprise," "withhold the patronage from your merchants and tradesmen," and "if you have job printing to be done, send it to a city office." Although the recipe is playfully written, the specificity of the instructions indicates real concern over the death of small towns, a phenomenon particularly pertinent to middle- and working-class people at this moment in the period. The commentary in this recipe could even be connected to broader concerns among the public regarding mass-industrialization. In any case, the clear purpose of the recipe to kill a town is to express dissent and civil unrest rather than represent the stability and control that recipes in published books often do.

By invoking the recipe form, the writer is able to subvert domestic authority but in a way that almost feels like an inside joke. While cookbooks are often written to act as a regulatory force of domesticity, the recipe-adjacent text flips that role on its head to reveal how well equipped the language and genres of domesticity are to conveying civil unrest and disobedience to a broader

readership. To the untrained eye, the self identification as a recipe may have even prompted a reader to skip over the article without reading it. To those in the know, however, rebellion and political strife are conveyed through the guise of the very symbol of domestic ideology and contentedness. The recipe genre and form works so well, is so covert in its conveyance of political opinion and class transgression, that it has even managed to go undetected by many who study newspapers or recipes. Without the model identifying these recipes as such, I likely wouldn't have paused over them for very long, myself, and my understanding of what recipes are and how they worked in the nineteenth-century would have still been limited to culinary purposes. Of course on one hand, that would mean that the recipe-adjacent texts are doing the work that they are intended to do. On the other hand, the absence of these recipes from broader conversations about newspaper writing and public platforms in the nineteenth century leaves out many writers for whom the recipe was the most useful or even safe form through which to convey such radical political ideas. The model, however, was able to see things that I couldn't see and in doing so, helped me expand and reflect on my own understanding of what a recipe is.

Formally, the pipeline reveals structures of the recipe which are counterintuitive. Namely, that the structure of recipes leans less of food words—which can vary significantly from region to region and from culinary style to culinary style—and more so on instructional and measurement words. According to this view of recipes which the model makes evident, the recipe, as a formal structure, can broadly be understood as short-form, instructional writing where actions, thoughts, or advice is prescribed according to measurements given at varying degrees of precision. This definition of the recipe form has very little to do with cooking and yet, the broader view of what recipes are and what kinds of messages they are useful for aren't outside the purview of the culinary, either. This expansive definition of the form poses the question: can the political agenda of chocolate cake coexist with commentary on society? Are the two messages implied by either equally useful to their readers, though for different reasons? And more importantly: Are the politics of the average citizen usefully couched in the forms and rhetoric of the domestic sphere which offers covert pathways for political critique without the risk of public scrutiny?

Conclusion

When researchers only attend to recipes as they appear in published cookbooks, they overlook an entire subgenre of recipes that circulated in nineteenth-century newspapers. Incorporating these texts into scholarship on recipes is essential because they offer a different perspective on the use and uptake of the recipe form in public spaces. The recipe form is far more flexible and adaptable than previous scholarship might suggest. From recipes for ice-boxes to cooking husbands, the recipe form as it was taken up in the

nineteenth-century newspaper demonstrates the unexpected ways domestic writing forms permeated public discourse, including ways that are not necessarily tied to the home or to kitchens.

That being said, these texts cannot be studied if researchers cannot locate them. The chaos of the newspaper page makes identifying recipes at a glance incredibly challenging, and their idiosyncratic form cannot easily be translated into keywords for database searches. Using computational methods allows readers to not only identify recipes in newspapers, but also recipe-adjacent texts that can expand our understanding of the form. Recipe-adjacent texts like those discussed above broaden scholarship on domestic writing forms and their uptake in public platforms, allowing us to understand the unique affordances the recipe form offers nineteenth-century writers. These affordances allow the writer to position themselves as both knowledgeable expert and friendly neighbor. The colloquial tone of newspaper recipes and their divergence from the authoritative book recipe, in turn, highlights the usefulness of the recipe form to those who would have found the published cookbook inaccessible, that same cookbook which describes servants and other members of the working-class as lazy or stubborn. Bringing the recipe into a public platform like the newspaper extends access to its usefulness to writers who otherwise would have been locked out of the readership of book-recipes. In this way, the recipe form is opened up for intervention and its boundaries shifted beyond the scope of the middle-class household and made to be reflective of lower-middle-class and working-class needs. Once these boundaries shift, the form does too, defying the often middle and upper-class understanding of recipes to include the language of the everyday—a language that does not neatly conform to any singular form.

Scholars who study the nineteenth century have much to gain from focusing on the ways forms shift and are stretched by different readers and writers. The affordances of digital humanities methods, which allow for the sorting and classification of numerous pages of text, can bring these forms to the forefront of attention where they might have otherwise been lost to the human eye. While scholars have made progress in leaps and bounds when it comes to the study of recipes, these studies still out of necessity privilege the published cookbook, which reflects primarily the interests of a affluent, middle-class reader and a writer attempting to prescribe their own brand of domesticity to their readers, one that often mocks servants or deems them a necessary evil. Expanding these studies to include the other places that recipes are printed–for instance the newspaper–not only broadens the intended readership of the piece but also expands our understanding of what a recipe is, what a recipe gets to be, and how importantly a recipe will be regarded—all of which are decisions we get to make, fluid as they may be.

WORKS CITED

Betti, Arianna, and Hein van den Berg. "Modelling the History of Ideas." *British Journal for the History of Philosophy*, vol. 22, no. 4, July 2014, pp. 812–35. *Taylor and Francis+NEJM*, https://doi.org/10.1080/09608788.2014.949217.

Bode, Katherine. "The Equivalence of 'Close' and 'Distant' Reading; or, Toward a New Object for Data-Rich Literary History." *Modern Language Quarterly*, vol. 78, no. 1, Mar. 2017, pp. 77–106. *Silverchair*, https://doi.org/10.1215/00267929-3699787.

Bowker, Geoffrey C., and Susan Leigh Star. *Sorting Things Out: Classification and Its Consequences*. MIT Press, 2000, https://doi.org/10.7551/mitpress/6352.001.0001.

Brewer, Priscilla J. *From Fireplace to Cookstove: Technology and the Domestic Ideal in America*. Syracuse University Press, 2000.

Brodhead, Richard H. *Cultures of Letters: Scenes of Reading and Writing in Nineteenth-Century America*. University of Chicago Press, 1993.

Broersma, Marcel, and Frank Harbers. "Exploring Machine Learning to Study the Long-Term Transformation of News." *Digital Journalism*, vol. 6, no. 9, Oct. 2018, pp. 1150–64. *Taylor and Francis+NEJM*, https://doi.org/10.1080/21670811.2018.1513337.

Cordell, Ryan. "'Q i-Jtb the Raven': Taking Dirty OCR Seriously." *Book History*, vol. 20, no. 1, 2017, pp. 188–225. *Project MUSE*, https://doi.org/10.1353/bh.2017.0006.

D'Ignazio, Catherine, and Lauren F. Klein. *Data Feminism*. MIT Press, 2020, https://doi.org/10.7551/mitpress/11805.001.0001.

Dobson, James. "Interpretable Outputs: Criteria for Machine Learning in the Humanities." *Digital Humanities Quarterly*, vol. 15, no. 2, June 2020.

Drucker, Johanna. "Why Distant Reading Isn't." *PMLA*, vol. 132, no. 3, May 2017, pp. 628–35. *Cambridge University Press*, https://doi.org/10.1632/pmla.2017.132.3.628.

Dutch, Jennifer Rachel. "Not Just for Laughs: Parody Recipes in Four Community Cookbooks." *Western Folklore*, vol. 77, no. 3/4, 2018, pp. 249–76.

Edelstein, Sari. *Between the Novel and the News: The Emergence of American Women's Writing*. University of Virginia Press, 2014. *ProQuest Ebook Central*, http://ebookcentral.proquest.com/lib/northeastern-ebooks/detail.action?docID=3444159.

Elias, Megan J. *Food on the Page: Cookbooks and American Culture*. University of Pennsylvania Press, 2017, https://doi.org/10.9783/9780812294033.

Fagan, Benjamin. "Chronicling White America." *American Periodicals: A Journal of History & Criticism*, vol. 26, no. 1, 2016, pp. 10–13.

Gabi, Kirilloff. "Computation as Context: New Approaches to the Close/Distant Reading Debate." *College Literature*, vol. 49, no. 1, 2022, pp. 1–25. *Project MUSE*, https://doi.org/10.1353/lit.2022.0000.

Gagliano, Andrea, et al. "Intersecting Word Vectors to Take Figurative Language to New Heights." *CLfL@NAACL-HLT*, 2016. *Semantic Scholar*, https://doi.org/10.18653/v1/w16-0203.

Garvey, Ellen Gruber. *Writing with Scissors: American Scrapbooks from the Civil War to the Harlem Renaissance*. OUP USA, 2012.

Guarneri, Julia. *Newsprint Metropolis: City Papers and the Making of Modern Americans*. University of Chicago Press, 2017, https://doi.org/10.7208/chicago/9780226341477.001.0001.

Jockers, Matthew, and Julia Flanders. *A Matter of Scale*. Faculty Publications -- Department of English, Mar. 2013, https://digitalcommons.unl.edu/englishfacpubs/106.

Joensuu, Juri. "Fictitious Meals, Culinary Constraints: The Recipe Form in Four Oulipian Texts." *Poetics Today*, vol. 42, no. 4, Dec. 2021, pp. 575–95. *DOI.org (Crossref)*, https://doi.org/10.1215/03335372-9356868.

King, Rachael Scarborough. "The Scale of Genre." *New Literary History*, vol. 52, no. 2, 2021, pp. 261–84. *Project MUSE*, https://doi.org/10.1353/nlh.2021.0012.

Klein, Lauren F. "Dimensions of Scale: Invisible Labor, Editorial Work, and the Future of Quantitative Literary Studies." *PMLA*, vol. 135, no. 1, Jan. 2020, pp. 23–39. *Cambridge University Press*, https://doi.org/10.1632/pmla.2020.135.1.23.

---. "The Image of Absence: Archival Silence, Data Visualization, and James Hemings." *American Literature*, vol. 85, no. 4, Dec. 2013, pp. 661–88. *read-dukeupress-edu.ezproxy.neu.edu*, https://doi.org/10.1215/00029831-2367310.

Kusner, Matt J., et al. *From Word Embeddings To Document Distances*. p. 10.

Latham, Rob. *Science Fiction Criticism: An Anthology of Essential Writings*. Bloomsbury Academic, 2017. *DOI.org (Crossref)*, https://doi.org/10.5040/9781474248655.

Le, Quoc V., and Tomas Mikolov. *Distributed Representations of Sentences and Documents*. arXiv:1405.4053, arXiv, 22 May 2014. *arXiv.org*, http://arxiv.org/abs/1405.4053.

Lee, Benjamin. "Compounded Mediation: A Data Archaeology of the Newspaper Navigator Dataset." *Digital Humanities Quarterly*, vol. 015, no. 4.

Long, Hoyt, and Richard Jean So. "Literary Pattern Recognition: Modernism between Close Reading and Machine Learning." *Critical Inquiry*, vol. 42, no. 2, 2016, pp. 235–67, https://doi.org/10.1086/684353.

Lynch, Deidre. "Paper Slips: Album, Archiving, Accident." *Studies in Romanticism*, vol. 57, no. 1, 2018, pp. 87–119. *Project MUSE*, https://doi.org/10.1353/srm.2018.0004.

Matthews, Glenna. *"Just a Housewife": The Rise and Fall of Domesticity in America*. Oxford University Press, 1989.

Mikolov, Tomas, et al. "Distributed Representations of Words and Phrases and Their Compositionality." *Advances in Neural Information Processing Systems*, vol. 26, Curran Associates, Inc., 2013. *Neural Information Processing Systems*, https://proceedings.neurips.cc/paper/2013/hash/9aa42b31882ec039965f3c4923ce901b-Abstract.html.

Mutuvi, Stephen, et al. "Evaluating the Impact of OCR Errors on Topic Modeling." *Maturity and Innovation in Digital Libraries*, edited by Milena Dobreva et al., Springer International Publishing, 2018, pp. 3–14. *Springer Link*, https://doi.org/10.1007/978-3-030-04257-8_1.

Nelson, Laura K. "Leveraging the Alignment between Machine Learning and Intersectionality: Using Word Embeddings to Measure Intersectional Experiences of the Nineteenth Century U.S. South." *Poetics*, vol. 88, Oct. 2021, p. 101539. *ScienceDirect*, https://doi.org/10.1016/j.poetic.2021.101539.

Neuhaus, Jessamyn. *Manly Meals and Mom's Home Cooking: Cookbooks and Gender in Modern America*. JHU Press, 2012, https://doi.org/10.56021/9780801871252.

Roggenkamp, Karen. *Narrating the News: New Journalism and Literary Genre in Late Nineteenth-Century American Newspapers and Fiction*. The Kent State University Press, 2005. *ProQuest Ebook Central*, http://ebookcentral.proquest.com/lib/northeastern-ebooks/detail.action?docID=3118710.

Schöch, Christof. "Topic Modeling Genre: An Exploration of French Classical and Enlightenment Drama." *Digital Humanities Quarterly*, vol. 011, no. 2, May 2017.

Schudson, Michael. *Discovering The News: A Social History Of American Newspapers*. Basic Books, 1981.

Smith, David A., et al. "Computational Methods for Uncovering Reprinted Texts in Antebellum Newspapers." *American Literary History*, vol. 27, no. 3, Sept. 2015, pp. E1–15. *Silverchair*, https://doi.org/10.1093/alh/ajv029.

Stewart, David M. *Reading and Disorder in Antebellum America*. The Ohio State University Press, 2011.

Theophano, Janet. *Eat My Words: Reading Women's Lives Through the Cookbooks They Wrote*. St. Martin's Publishing Group, 2016.

Todorov, Konstantin, and Giovanni Colavizza. *An Assessment of the Impact of OCR Noise on Language Models*. arXiv:2202.00470, arXiv, 26 Jan. 2022. *arXiv.org*, http://arxiv.org/abs/2202.00470.

Underwood, Ted. *Distant Horizons: Digital Evidence and Literary Change*. University of Chicago Press, 2019, https://doi.org/10.7208/chicago/9780226612973.001.0001.

---. "Genre Theory and Historicism." *Journal of Cultural Analytics*, vol. 2, no. 2, Oct. 2016, p. 11063. *culturalanalytics.org*, https://doi.org/10.22148/16.008.

Walker, Daniel, et al. "Evaluating Models of Latent Document Semantics in the Presence of OCR Errors." *Faculty Publications*, Jan. 2010, https://scholarsarchive.byu.edu/facpub/1648.

Warhol-Down, Robyn and American Council of Learned Societies. *The Work of Genre Selected Essays from the English Institute*. English Institute in collaboration with the American Council of Learned Societies, 2011.

Williams, Lyneise. "What Computational Archival Science Can Learn from Art History and Material Culture Studies." *2019 IEEE International Conference on Big Data (Big Data)*, 2019, pp. 3153–55. *IEEE Xplore*, https://doi.org/10.1109/bigdata47090.2019.9006527.

Zboray, Ronald J., and Mary Saracino Zboray. "Is It a Diary, Commonplace Book, Scrapbook, or Whatchamacallit?: Six Years of Exploration in New England's Manuscript Archives." *Libraries & the Cultural Record*, vol. 44, no. 1, 2009, pp. 101–23. *Project MUSE*, https://doi.org/10.1353/lac.0.0055.