

Navigating the Mise-en-Page: Interpretive Machine Learning Approaches to the Visual Layouts of Multi-Ethnic Periodicals

Benjamin Charles Germain Lee^{1,*}, Joshua Ortiz Baco^{2,*}, Sarah H. Salter^{3,*} and Jim Casey^{4,*}

¹University of Washington, USA

²Univeristy of Texas at Austin, USA

³Texas A&M University-Corpus Christi, USA

⁴Pennsylvania State University, USA

* = equal contribution

Abstract

This paper presents a computational method of analysis that draws from machine learning, library science, and literary studies to map the visual layouts of multi-ethnic newspapers from the late 19th and early 20th century United States. This work departs from prior approaches to newspapers that focus on individual pieces of textual and visual content. Our method combines *Chronicling America's* MARC data and the *Newspaper Navigator* machine learning dataset to identify the visual patterns of newspaper page layouts. By analyzing high-dimensional visual similarity, we aim to better understand how editors spoke and protested through the layout of their papers.

Keywords

machine learning, periodicals, newspapers, editors, document layout analysis, *Chronicling America*, *Newspaper Navigator*, MARC, multi-ethnic press

1. Introduction

This paper presents a set of developing methods for analyzing Black, Latinx, and other ethnic newspapers in the late nineteenth and early twentieth-century United States. We address these newspapers in this essay because multi-ethnic newspapers faced challenges above and beyond their era's respective forms of mainstream journalism. Contending with forces of racism and xenophobia often required editors to tailor their paper's visual grammar, communicating in carefully guarded ways that were still capable of circulating to distant readers and communities. Those traditions of complex editorial craft in the multi-ethnic presses represent an especially rich archive for analysis.

Editors in the historical multi-ethnic press often chose to proceed circumspectly by focusing on their newspaper's "mise-en-page." Rather than publish bold condemnations in headlines or

CHR 2021: Computational Humanities Research Conference, November 17–19, 2021, Amsterdam, The Netherlands

✉ bcgl@cs.washington.edu (B.C.G. Lee); joshuaortizbaco@utexas.edu (J. Ortiz Baco);

Sarah.Salter@tamucc.edu (S.H. Salter); jccasey@psu.edu (J. Casey)

🌐 <https://www.bcglee.com/> (B.C.G. Lee); <https://www.sarahhsalter.net/> (S.H. Salter);

<https://www.jim-casey.com/> (J. Casey)

📄 0000-0002-1677-6386 (B.C.G. Lee); 0000-0002-9723-4262 (J. Ortiz Baco); 0000-0001-7548-0932 (S.H. Salter); 0000-0001-5847-5242 (J. Casey)

© 2021 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

editorials, it was frequently safer to rely on other affordances of newspapers provided by the aggregate, often invisible techniques collapsed under the heading of editing. [22]. Editorial techniques include the labor required to gather, arrange, and issue a newspaper as well as the organization of those labors behind the scenes. They might vary greatly by paper, time, and place. Readers might recognize traces of editorial techniques in a paper’s use of headlines for emphasis, photographs to dramatize articles, or the separation of content into sections. At times, these techniques are performed by dedicated editors: news editors, copyeditors, managing editors, or photo editors. In this paper, we focus on the design of front pages, where many of the above-named editorial techniques intersect to create meaningful relationships between items on the page. This feature of multi-ethnic editorial communication makes it untenable to extract and analyze isolated textual or visual items from these newspapers.

Computational methods make it possible to analyze the visual patterns of multi-ethnic newspapers at scale. While prior research has approached these questions through textual analysis, a pivot to computational visual analysis compensates for the frequently noisy OCR text produced from historically neglected multi-ethnic newspapers.

Our working method uses machine learning to map the layouts of hundreds of thousands of newspaper pages, quantified by visual similarity, using the Newspaper Navigator dataset (as pictured in Figure 1) [29]. Visualizing and comparing layouts at scale provides a wayfinding function, helping to direct archival researchers to distinctive and meaningful places in large collections of digitized newspapers and cultural heritage collections writ large.

In this paper, we describe work in progress that brings machine learning, library science, and literary studies into conversation. This work includes: (1) a dataset of visual features from 16.3 million pages of historical newspapers; (2) remediation in 207 catalog records for 309 ethnic editors; and (3) modes for reading the practical languages of editing in specific historical contexts. Prior research efforts blend together in our exploratory machine learning prototypes for quantitatively and qualitatively measuring the visual similarities of the formats of a newspaper page. These visual similarities can help illuminate the ideas and activism that flowed through the editorial craft and conduct of historical multi-ethnic newspapers. Moreover, our work formulating a “similarity score” affirms the promise of multi-disciplinary collaborations for improving access and analysis of larger-scale digitized newspaper collections.

2. Related Prior Work on Developing Computational Approaches to Multi-Ethnic Periodicals

Multiple fields are eager to explore the new possibilities for analyzing digitized collections of historical newspapers. Complementary conversations in computational, archival, and literary studies have begun to pursue those possibilities in rich but largely disparate ways. Silos persist even as many scholars make parallel use of the Library of Congress’s *Chronicling America* and other national newspaper digital initiatives [15, 6, 36, 42, 26, 2, 3, 34, 35].

This paper builds on multi-ethnic press scholarship and archival research in *Chronicling America* and other repositories. The prevailing view of political activism in historical multi-ethnic newspapers has tended to focus on the textual and visual content that drove demands for social justice [31, 7, 43, 40]. A complementary approach has focused on social histories, tracing the networks of a paper’s contributors and readers [20, 21, 37, 39, 19]. What can be easy to miss, however, is that editors can use the format of a paper itself to voice critiques and protests. The formal nature of editors’ craft is difficult to grasp in the meticulous approaches



Figure 1: A visualization of the front page of the May 22, 1917, issue of *La Prensa* [8], annotated with bounding boxes of visual content derived from the Newspaper Navigator dataset [29]. These bounding boxes are predictions made by a machine learning model showing, in this case, the locations of headlines, photographs, and maps (each bounding box includes the predicted class, as well as a confidence score, in the top left corner). In this paper, we utilize these bounding boxes to compare newspaper titles according to visual similarity.

to archival research on a case-by-case basis. Our work aims to develop experimental machine learning applications to map the larger patterns of editorial practices and editing conventions that span languages, communities, and eras.

Conversations in library and archival studies are exploring how best to account for histories of race and ethnicity in metadata. Rather than being reduced to a theme or topic within a collection, specialists have rethought how race and ethnicity might work as organizing principles for collecting and describing historical materials. Race and ethnicity as categories of identity are historically contingent and socially constructed. That fluidity challenges the rigors of knowledge organization systems that impose stable categories and schemas. For some

archivists, however, categories of race and identity productively serve as a form of provenance to guide collection development and access policies [45, 17, 5]. Many of these changes at the collection level are being addressed in the metadata and, more recently, through linked open data. Because the development of controlled vocabularies, such as the Library of Congress Subject Headings, and Machine Readable Cataloging (MARC) largely predate sensitivity to these histories in libraries and archives, metadata has become a vital space for rediscovering the historical editors of multi-ethnic newspapers [41].

At the same time, an overabundance of work with more “computationally robust,” institutionally privileged white print collections has prompted preliminary conversations regarding the directions of computational humanities research. Indeed, from OCR engines that do not support indigenous languages or different dialects, to entire subfields of natural language processing and textual analysis that assume English as the default language of study, algorithmically-mediated erasure continues to be uncovered and detailed by scholars [18, 4, 25, 46, 24, 44, 23, 38, 32]. Many have advocated for the development of computational approaches for those collections that remain marginal. This paper concerns one such corrective for digitized periodicals: utilizing machine learning to study visual content and page layout rather than textual content, circumventing the use of OCR engines entirely.

Computational scholars have isolated visual content within periodicals in order to produce datasets of isolated content. Humanistic scholars have utilized these atomized items for interpretation. We are instead interested in the visual layouts themselves and the relations they imply on the page. By treating relations and patterns within layouts as forms of interpretive content – reading them as a “visual grammar” of photographs, advertisements, headlines, and beyond – we are in effect exploring a publication’s creative history, local contexts, and circulation networks.

3. Methods

The work described in this paper began in part with the Newspaper Navigator project, created as part of Library of Congress’s Innovator in Residence program in 2019 with the goal of excavating visual content in *Chronicling America*. Developed in partnership with LC Labs, the National Digital Newspaper Program, and IT Design & Development at the Library of Congress, Newspaper Navigator utilizes a machine learning model to extract seven visual features from 16 million pages in *Chronicling America*: advertisements, photographs, illustrations, maps, comics, editorial cartoons, and headlines [29]. This machine learning model consists of an object detection model that had been finetuned on thousands of bounding box annotations of visual content created by volunteers as part of the Beyond Words crowdsourcing initiative launched by LC Labs in 2017 [27]. The resulting extracted visual content was publicly released as the Newspaper Navigator dataset, available in the public domain for all to use [29].

Newspaper Navigator coincides with current trends in periodical studies. As Sean Latham and Robert Scholes argue in an influential essay, “We have often been too quick to see magazines [and other periodicals] as containers of discrete bits of information rather than autonomous objects of study” [28]. In practice, this means that periodical scholars tend to examine not only the individual pieces of content in a given publication, but also its overall construction, taking into account such architectural elements as a publication’s formats, columns, and visual layout. The typical methods to apply this approach require intensive reading of individual newspaper pages. Researchers gradually develop an intuition about the general style and flavor of a certain

publication. A publication’s implicit patterns can be relatively intangible because periodical formats are necessarily fixed—to preserve continuity—and flexible—to accommodate new and changing scenarios. A paper looks like itself almost always, but not always. Even beyond academic inquiry, dedicated readers of any periodical often develop this kind of intuitive mental map of a publication’s page designs and formats.

We saw the possibility of combining these two types of reading. On one hand, Newspaper Navigator makes it possible to chart visual features across the entirety of *Chronicling America*. On the other hand, the methods in periodical studies encourage attention to the juxtaposition of visual features on a page. Starting with these two methods, we developed a set of preliminary working questions. How do the visual features on the page of a newspaper interlock and mutually inform each other? If we look at all of those features holistically, what kinds of styles and ideas can we detect over the lifespan of a single newspaper? More broadly, what added perspective can we gain by using the Newspaper Navigator to trace the sum of newspaper layouts over hundreds of publications? Finally, can we establish a “baseline” for newspaper production in the later 19th and early 20th century United States that would allow us to see the critical departures made by multi-ethnic newspapers in the service of social justice during moments of elevated racism or xenophobia?

3.1. Methods I: Adapting MARC 21 Metadata for Research

Adapting the Newspaper Navigator dataset to answer these questions required us to collect additional data about the subsets of multi-ethnic newspapers within *Chronicling America*. This information was partially extracted from the bibliographic data of newspaper titles in the machine-readable cataloging (MARC) format. All records for newspapers in *Chronicling America* follow the cataloging standards of the Cooperative Online SERIALS Program (CONSER) [1], providing a somewhat normalized set of features for sorting newspapers. Additionally, *Chronicling America*’s partnering institutions provide a “title essay” that describes and contextualizes each newspaper. These unstructured data has recently been the focus of efforts for metadata enhancement in remediation projects geared to identifying ethnic editors [33].

The last step in collecting and organizing our data was combining MARC fields and structured data generated from the text of title essays to find relevant publications. Metadata from the title essays allowed us to select only ethnic newspapers with identifiable editors. From this subset, we selected MARC fields used by cataloguers to record title changes, race, ethnicity, and languages associated with each publication. We initially hoped to find an indicator of visual change within the metadata for newspapers. We considered that serials with more than one language edition, language notes, or different titles over their publication runs would feature changes to visual and layout elements of newspapers to mark these relationships or changes.¹ Title changes, for example, are a cataloguing practice for linking preceding and succeeding titles of a publication that has had a substantial name change, which we assumed would also produce modifications to the graphic elements on the page. We were able to parse these metadata fields where the change inherent in the seriality of newspapers is recorded, which in turn provided parameters for capturing difference, similarity, and linking in an otherwise static knowledge organization system.

¹The 780 and 785 fields indicate all preceding and succeeding titles for a publication, respectively. The 546 field describes the language of a resource while the 765 indicates the original language of a translated publication and the 775 captures other language editions of a publication.

Heat Maps of Advertisements on the Front Pages of *The Opelousas Courier*, 1870-1909

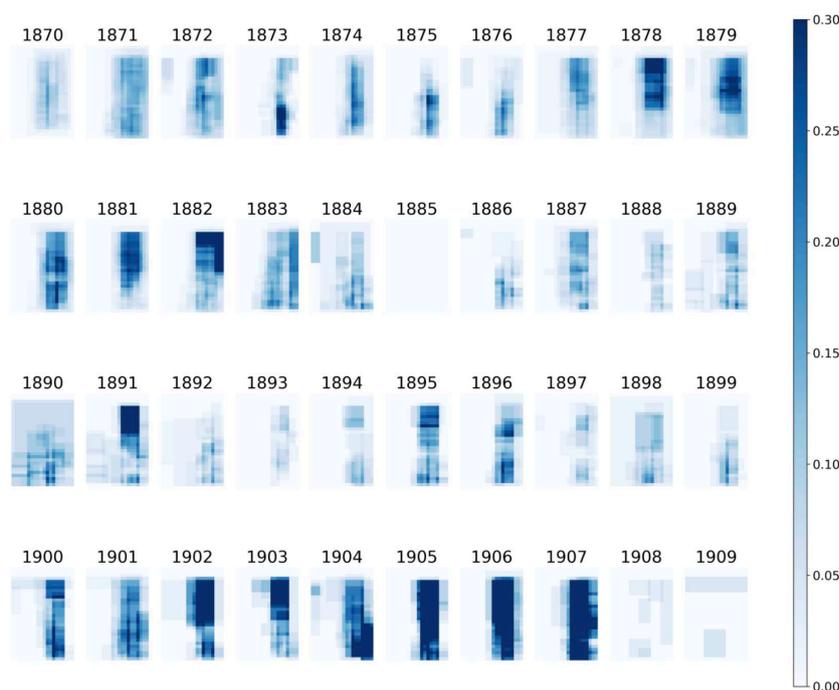


Figure 2: A heatmap of advertisements appearing on the front page of the Opelousas Courier, 1870-1909. Darker regions on the heatmaps correspond to a higher concentration of advertisement pages appearing in that region in aggregate over a given year.

3.2. Methods 2: Quantifying and Visualizing the Mise-en-Page

The first visual output we experimented with were heat maps, created with Newspaper Navigator data that encodes the positions of visual features on a newspaper page. Inspired by PageOneX [16], these heat maps concisely summarize visual content patterns in an interpretable fashion. Figure 2 shows the locations and relationships of advertisements across four decades of *The Opelousas Courier*.² As an initial experiment in visual layout representation, the heat maps enabled us to trace the development of discrete visual features, otherwise impossible to quantify by examining front pages by eye.

This approach facilitates several comparative opportunities. We rendered multiple visual categories into heat maps to understand the interplay between visual components in a single publication. We also utilized the heat map graphics to compare different newspapers using the same visual categories. The heat maps offered a useful preliminary system for studying newspaper layout as historical information.

Heat maps added greater nuance to our historical questions about newspaper layout. For example, what are the relationships between visual content, editorial practices, and print technology? What are the roles of regional and local context in predicting layout? How do we account for journalistic trends, such as the rise of syndicated content or nationally-circulating advertisements?

²*The Opelousas Courier* is described in *Chronicling America* as a Louisiana multilingual newspaper, with content in French and English. <https://chroniclingamerica.loc.gov/lccn/sn83026389/marc/>.

The Appeal



March 14, 1891

The Washington Bee



February 29, 1896

The Colored American



April 7, 1900



May 9, 1891



August 29, 1896



September 8, 1900

Figure 4: Sample front pages from *The Appeal*, *The Washington Bee*, and *The Colored American* from 1891, 1896, and 1900, respectively – all of which are title-year pairs appearing in the same cluster in Figure 3 [9, 10, 14, 13, 11, 12]. All six front pages feature visual content – namely, illustrations and photographs of individuals – prominently in the center of the page.

from years included within the cluster. All six pictured front pages feature visual content on the center of the page. Notably, all six feature photographs and illustrations of individuals, a trend that is even more apparent when browsing these front pages across different issues. These images demonstrate that the similarity clusters, generated using Newspaper Navigator's rubric of seven content classes, accurately reflect common visual patterns of layout in distinct newspapers across the Black newspapers held in *Chronicling America*. The shared qualities across each of the three newspapers pictured in Figure 4 are suggestive. They are roughly contemporaneous African American newspapers that locate specific types of images near the center of the front page. These results indicate the capacity of the method to produce a multifaceted research question: what can we learn from the use of visual culture and portraiture in the material pages of late nineteenth-century African American newspapers?

These constellations of similarity provide a means to uncover small networks to focus on and interpret at a much more granular level. This is not a fully quantitative representation, but rather an exploratory framework for identifying and interpreting networks of related

publications and the editorial practices that shape them. Given our overarching interest in the historical trajectories and meanings of newspaper formats, these visualizations show great promise. The constellation map is a model for how to convert raw bounding box information into interpretable provocations about historical similarity. Similarity clusters introduce visual layouts as clues to a shared context, inviting new research into the different historical traditions, technological moments, or political causes that comprise the histories of editorship.

4. Findings & Future Work

Our ongoing collaboration offers three central findings to date. First, this work finds that full level CONSER MARC records provide useful faceted metadata for accurately selecting multi-ethnic newspapers in machine-learning applications. Specifically, the most value comes from fields that register change over time, such as preceding and succeeding titles, and describe relationships between different editions of the same publication. Further metadata enhancement focused on identifying ethnic editors will significantly improve the interpretative value of visualizations and pattern recognition from large-scale collections like *Chronicling America*.

Second, this collaboration reveals how computational notions of layout similarity can be leveraged for humanistic inquiry. Using the machine learning-constructed Newspaper Navigator dataset of visual content, we have generated heat maps and constellation maps for exploring trends of visual layout within and across newspaper titles, respectively. These visualizations provide us with new visual grammars and affordances for excavating editorial practices.

Third, this work required us to find a collaborative working process for creating multidisciplinary methods of analysis that engage with machine learning, library science, and the humanities. Rather than trying to blend or transplant our respective research methods, we focused on creating a shared vocabulary that could speak to current research agendas in each of the relevant fields. This set of theoretical exchanges required iterative conversations, frequently pedagogical in nature, to find a method that is not merely extractive and static but generative and dynamic. This process will help guide the future growth of conversations in computational periodicals research.

Our roadmap for future work is guided by the provocations offered throughout this paper. First, we plan to further utilize MARC data and the constellation maps in concert with one another in order to inform our understanding of both. The MARC data has the capacity to guide our navigation and understanding of the constellation maps; conversely, the constellation maps provide mechanisms for record enhancement by allowing scholars, catalogers, and librarians alike to sift through incomplete legacy records of ethnic publications. We offer this direction of future work in pursuit of uncovering editorship in ethnic presses. We plan to refine our similarity metrics and visualizations from a computational perspective in pursuit of this goal as well. The results of this research will drive new inquiry into the longer and varied histories of radical dissent and veiled protest in the multi-ethnic press.

Acknowledgments

The authors would like to thank and acknowledge Dr. Molly Hardy, Senior Program Officer at the National Endowment for the Humanities, along with Georgia Higley and Vanessa Mitchell from the Serial and Government Publications Division at the Library of Congress, for sharing their expertise. This work would not be possible without Dr. Hardy's input and advice.

References

- [1] *About CONSER - Program for Cooperative Cataloging (Library of Congress)*. URL: <https://www.loc.gov/aba/pcc/conser/about/>.
- [2] *About Delpher*. URL: <https://www.delpher.nl/nl/platform/pages/helpitems?title=wat+is+delpher>.
- [3] *About The British Newspaper Archive*. URL: <https://www.britishnewspaperarchive.co.uk/help/about>.
- [4] H. Alpert-Abrams. “Machine Reading the Primeros Libros”. In: *Digital Humanities Quarterly* 010.4 (2016).
- [5] J. A. Bastian. “Documenting Communities Through the Lens of Collective Memory”. In: *Identity Palimpsests: Archiving Ethnicity in the U.S. and Canada*. Ed. by D. Daniel and A. S. Levi. Archives, archivists and society book 5. Sacramento, CA: Litwin Books, 2014, pp. 15–34.
- [6] S. Cassidy. “Publishing the Trove Newspaper Corpus”. In: *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC’16)*. Portorož, Slovenia: European Language Resources Association (ELRA), 2016, pp. 4520–4525. URL: <https://www.aclweb.org/anthology/L16-1715>.
- [7] J. L. Cole. *How the Other Half Laughs: The Comic Sensibility in American Culture, 1895-1920*. Jackson: U Press of Mississippi, 2020.
- [8] L. of Congress. *La Prensa, 22 May 1917. Chronicling America: Historic American Newspapers*. 1917. URL: <https://chroniclingamerica.loc.gov/lccn/sn83045395/1917-05-22/ed-1/seq-1/>.
- [9] L. of Congress. *The Appeal, 14 March 1891. Chronicling America: Historic American Newspapers*. 1891. URL: <https://chroniclingamerica.loc.gov/lccn/sn83016810/1891-03-14/ed-1/seq-1/>.
- [10] L. of Congress. *The Appeal, 9 May 1891. Chronicling America: Historic American Newspapers*. 1891. URL: <https://chroniclingamerica.loc.gov/lccn/sn83016810/1891-05-09/ed-1/seq-1/>.
- [11] L. of Congress. *The Colored American, 7 April 1900. Chronicling America: Historic American Newspapers*. 1900. URL: <https://chroniclingamerica.loc.gov/lccn/sn83027091/1900-04-07/ed-1/seq-1/>.
- [12] L. of Congress. *The Colored American, 8 September 1896. Chronicling America: Historic American Newspapers*. 1900. URL: <https://chroniclingamerica.loc.gov/lccn/sn83027091/1900-09-08/ed-1/seq-1/>.
- [13] L. of Congress. *The Washington Bee, 29 August 1896. Chronicling America: Historic American Newspapers*. 1896. URL: <https://chroniclingamerica.loc.gov/lccn/sn84025891/1896-08-29/ed-1/seq-1/>.
- [14] L. of Congress. *The Washington Bee, 29 February 1896. Chronicling America: Historic American Newspapers*. 1896. URL: <https://chroniclingamerica.loc.gov/lccn/sn84025891/1896-02-29/ed-1/seq-1/>.
- [15] L. of Congress & National Endowment for the Humanities. *Chronicling America*. Text. 2021. URL: <https://chroniclingamerica.loc.gov/>.

- [16] S. Costanza-Chock and P. Rey-Mazón. “PageOneX: New Approaches to Newspaper Front Page Analysis”. In: *International Journal of Communication* 10 (2016), p. 28.
- [17] D. Daniel. “Archival Representations of Immigration and Ethnicity in North American History: From the Ethnicization of Archives to the Archivization of Ethnicity”. In: *Archival Science* 14.2 (2014), pp. 169–203. DOI: 10.1007/s10502-013-9209-6. URL: <http://link.springer.com/10.1007/s10502-013-9209-6>.
- [18] B. Fagan. “Chronicling White America.” In: *American Periodicals* 26.1 (2016), pp. 10–13. URL: <http://search.proquest.com/docview/2152559277/F95B44D158E64163PQ/13>.
- [19] B. Fagan. *The Black Newspaper and the Chosen Nation*. Athens, GA: University of Georgia Press, 2016.
- [20] F. S. Foster. “A Narrative of the Interesting Origins and (Somewhat) Surprising Developments of African-American Print Culture”. In: *American Literary History* 17.4 (2005), pp. 714–740. URL: <http://muse.jhu.edu/journals/american%5C%5Fliterary%5C%5Fhistory/v017/17.4foster.html>.
- [21] E. Gardner. *Black Print Unbound: The Christian Recorder, African American Literature, and Periodical Culture*. Oxford University Press, 2015.
- [22] P. Giddings. *Ida: A Sword Among Lions: Ida B. Wells and the Campaign against Lynching*. New York, N.Y.: HarperCollins, 2009.
- [23] K. L. Gustafson. “Ethnic Newspaper Producers Face Archiving Challenges”. In: *Newspaper Research Journal* 36.3 (2015), pp. 314–327. DOI: 10.1177/0739532915600744. URL: <https://doi.org/10.1177/0739532915600744>.
- [24] M. Hardy and L. DiCuirici. *Critical Cataloging and the Serials Archive: The Digital Making of “Mill Girls in Nineteenth-Century Print” - Archive Journal*. 2019. URL: <http://www.archivejournal.net/?p=8073>.
- [25] P. Joshi, S. Santy, A. Budhiraja, K. Bali, and M. Choudhury. “The State and Fate of Linguistic Diversity and Inclusion in the NLP World”. In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Online: Association for Computational Linguistics, 2020, pp. 6282–6293. DOI: 10.18653/v1/2020.acl-main.560. URL: <https://www.aclweb.org/anthology/2020.acl-main.560>.
- [26] K. Kreitz. “American Alternatives: Participatory Futures of Print from New York City’s Nineteenth-Century Spanish-Language Press”. In: *American Literary History* 30.4, 1 (2018), pp. 677–702. DOI: <https://doi.org/10.1093/alh/ajy032>.
- [27] L. Labs. *Beyond Words (“Mark”)*. 2017. URL: <http://beyondwords.labs.loc.gov/%5C#/mark>.
- [28] S. Latham and R. Scholes. “The Rise of Periodical Studies”. In: *PMLA/Publications of the Modern Language Association of America* 121.2 (2006), pp. 517–531. DOI: 10.1632/003081206x129693.

- [29] B. C. G. Lee, J. Mears, E. Jakeway, M. Ferriter, C. Adams, N. Yarasavage, D. Thomas, K. Zwaard, and D. S. Weld. “The Newspaper Navigator Dataset: Extracting Headlines and Visual Content from 16 Million Historic Newspaper Pages in Chronicling America”. In: *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. New York, NY, USA: Association for Computing Machinery, 2020, pp. 3055–3062. URL: <https://doi-org.offcampus.lib.washington.edu/10.1145/3340531.3412767>.
- [30] L. van der Maaten and G. Hinton. “Visualizing Data using t-SNE”. In: *Journal of Machine Learning Research* 9.86 (2008), pp. 2579–2605. URL: <http://jmlr.org/papers/v9/vandermaaten08a.html>.
- [31] J. Nerone and K. G. Barnhurst. “US Newspaper Types, the Newsroom, and the Division of Labor, 1750–2000”. In: *Journalism Studies* 4.4 (2003), pp. 435–449. DOI: 10.1080/1461670032000136541. URL: <https://doi.org/10.1080/1461670032000136541>.
- [32] S. U. Noble. *Algorithms of Oppression: How Search Engines Reinforce Racism*. NYU Press, 2018. URL: <http://www.jstor.org/stable/j.ctt1pwt9w5>.
- [33] J. Ortiz Baco. *Title Essays, Linked Data, and the Ethnic Press in Chronicling America*. 2019. URL: <https://www.neh.gov/blog/title-essays-linked-data-and-ethnic-press-chronicling-america>.
- [34] *OurDigitalWorld: Digital Newspapers*. URL: <https://ourdigitalworld.net/what-we-do/digital-newspapers/>.
- [35] *Papers Past*. <https://natlib.govt.nz/collections/a-z/papers-past>.
- [36] A. Pekárek and M. Willems. “The Europeana Newspapers – A Gateway to European Newspapers Online”. In: *Progress in Cultural Heritage Preservation*. Ed. by M. Ioannides, D. Fritsch, J. Leissner, R. Davies, F. Remondino, and R. Caffo. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 654–659.
- [37] J. Rhodes. *Mary Ann Shadd Cary: The Black Press and Protest in the Nineteenth Century*. Bloomington, IN: Indiana University Press, 1998.
- [38] R. Risam. *New Digital Worlds: Postcolonial Digital Humanities in Theory, Praxis, and Pedagogy*. Evanston, Illinois: Northwestern University Press, 2018.
- [39] D. R. Spires. *The Practice of Citizenship: Black Politics and Print Culture in the Early United States*. Philadelphia: University of Pennsylvania Press, 2019.
- [40] S. A. Stein. “Making Jews Modern: The Yiddish and Ladino Press in the Russian and Ottoman Empires, The Modern Jewish Experience”. In: *International Journal of Middle East Studies* 41.2 (2007). DOI: 10.1017/s0020743809090874.
- [41] G. B. Ventura, L. Gauthereau, and C. Villarroel. “Recovering the US Hispanic Literary Heritage: A Case Study on US Latina/o Archives and Digital Humanities”. In: *Preservation, Digital Technology & Culture (PDT&C)* 48.1 (2019), pp. 17–27. DOI: 10.1515/pdte-2018-0031. URL: <https://www.degruyter.com/document/doi/10.1515/pdte-2018-0031/html>.
- [42] M. Willems and R. Atanassova. “Europeana Newspapers: Searching Digitized Historical Newspapers from 23 European Countries”. In: *Insights* 28.1 (2015), pp. 51–56. DOI: 10.1629/uksg.218. URL: <http://insights.uksg.org/articles/10.1629/uksg.218/>.

- [43] A. N. Williams. “Cultivating Black Visuality: The Controversy over Cartoons in the Indianapolis Freeman”. In: *American Periodicals: A Journal of History & Criticism* 25.2 (2015), pp. 124–138. DOI: 10.1353/amp.2015.0032. URL: <http://muse.jhu.edu/article/601461>.
- [44] L. Williams. “What Computational Archival Science Can Learn from Art History and Material Culture Studies”. In: *2019 IEEE International Conference on Big Data (Big Data)*. 2019, pp. 3153–3155. DOI: 10.1109/BigData47090.2019.9006527.
- [45] J. Wurl. “Ethnicity as Provenance: In Search of Values and Principles for Documenting the Immigrant Experience”. In: *Archival Issues* 29.1 (2005), pp. 65–76. URL: <http://www.jstor.org/stable/41102095>.
- [46] E. Yakel. “Archival representation”. In: *Archival Science* 3.1 (2003), pp. 1–25. DOI: 10.1007/bf02438926. URL: <https://doi.org/10.1007/BF02438926>.