A World of Fiction

Bode, Katherine

Published by University of Michigan Press

Bode, Katherine.

For additional information about this book
https://muse.jhu.edu/book/59018

For content related to this chapter
https://muse.jhu.edu/related_content?type=book&id=2119799
Introduction

1. I compare major mass-digitized historical newspaper collections in chapter 3.

2. Dickens’s “Little Dorritt” in Melbourne’s Leader in 1856 is the first title by an established international author listed in Johnson-Woods’s Index. Earlier works by Dickens in daily and weekly Australian newspapers include “Pickwick Papers” (Sydney Gazette, 1838), “Master Humphrey’s Clock” (Australasian Chronicle, 1840), “Barnaby Rudge” (Australasian Chronicle, 1841), “American Notes for General Circulation” (Southern Australian and Sydney Morning Herald, 1843), “Martin Chuzzlewit” (Launceston Examiner, 1845), and “Hard Times” (Argus, 1854). Record IDs for these works in the curated dataset are: #177; #856; #860; #498 and #499; #2259; and #460.

3. Johnson-Woods describes Braddon as “queen of the colonies” due to the scale of her fiction (“Mary”), but this project has discovered more titles by local author James Joseph Wright than Braddon (eighty-nine versus eighty-six).

4. As I discuss in chapter 1, Moretti proposed this concept in “Conjectures on World Literature” and developed it in two subsequent books: Graphs, Maps, Trees and Distant Reading.

5. For articles and blogs that use distant reading in this way, see Alexander et al.; Erlin and Tatlock; Liddle; Goldstone, “Distant”; and Underwood, “Dataset”.

6. See Dinsman; Lohr; Piepenbring; Rothman; Schultz; and Sunyer.

7. Book history often employs a systemic conception of print culture—
most famously in Darnton’s notion of the “communication circuit”—as well as quantitative methods. On parallels between distant reading and book history and the advantages of the latter framework over the former, see Bode, Reading 7–25.

8. Stories of more than sixty thousand words include Walter Besant’s “They Were Married!” in the Sydney Mail in 1887 (#13291) and Rolf Boldrewood’s “The Final Choice” in the Australasian in 1885 (#13337).

9. The database is accessible at http://cdhrdatasys.anu.edu.au/tobe-

10. For instance, the essays in a recent special issue of Australian Literary Studies explore “the ways in which Victorian literary texts and ideas were transformed by their arrival and reception in the Australasian colonies and then re-transmitted around the trade lines of Empire” (Martin and Mirmohamadi np).

Chapter 1

1. For example, the final chapter of Moretti’s Distant Reading (211–40) uses network analysis to analyze Shakespeare’s Hamlet.


3. On the limitations of bibliographical records for quantitative analysis, see Elliot.

4. Folsom’s essay on database as a new genre also associates digital technologies with comprehensive and direct access to the literary-historical record, even as the resource he refers to—The Walt Whitman Archive (whitmanarchive.org/)—enacts a carefully historicized approach to the documentary record. Responses to this essay, including by Freedman and by McGann (“Database”), represent early rejections of this paradigm of transparency in digital humanities.

5. Moretti is a coauthor on the pamphlet but, unlike the other authors, no specific role or insight is ascribed to him.

6. An exception to this lack of data publication occurs in a pamphlet that Moretti and Jockers authored collaboratively with others (Allison et al.).

7. In a copublished article, Jockers confirms the hints he gives in Macroanalysis: that his data are unpublished because they are derived from proprietary collections (Jockers and Mimno 752). But this does not explain why he cannot name the authors and titles studied or provide textual data at the level of word frequencies.
8. Jockers’s copublished article with Kirilloff includes a bibliography of the 3,329 works in their corpus, as well as a note about the various sources they are derived from.

9. For foundational work in this area, see McKenzie; and McGann, *Textual*.

10. The technique misclassifies 33 percent of works in terms of nationality, 14 percent of works in terms of gender, and an unspecified proportion of works in terms of chronology (Jockers, *Macroanalysis* 153).

11. Reinforcing Moretti’s disinterest in his underlying datasets, although he describes an increase to fifty novels per year for all national contexts, the graph shows British novels increasing to thirty titles, Japanese and Spanish novels to a little over forty, Italian novels to thirty-five, and Nigerian novels to only twenty-five titles per year (*Graphs*, 6).

12. Oft cited is Moretti’s claim that “between interpretation (that tends to make a close reading of a single text) and explanation (that works with abstract models on a large groups [sic] of texts) I see an antithesis. Not just difference, but an either/or choice” (“Moretti” 74).

**Chapter 2**

1. Shillingsburg offers an accessible account of the shift in scholarly editing from an archetypal to a historical and material approach (167–68) and of the distinctions between the implied, represented, and interpreted work (170–81).

2. Algee-Hewitt and McGurl model a literary system using ranked lists of books judged to be important by various organizations. Literary systems could also be constructed based on data employed in book history, such as number of print runs, sales records, or library holding and borrowing data.

3. Humanities scholars have long recognized archives as manifesting structures of authority rather than as neutral containers; see Manoff for a concise summary of this extensive debate.

4. Because typically created from existing analog collections, digital ones can grow at the faster rate of digitization than manual collection. Database and interface features can be renovated more quickly than physical infrastructure; even if development is lengthy, when changes are implemented, for the end user they take effect instantaneously. Crowd-sourced corrections—a celebrated feature of *Trove*’s newspaper collection—mean that searchable text can change even when the number of digitized documents remains constant.

5. Research that discusses this relationship includes Mussell, *Nineteenth*; Solberg.

6. For these projects see McGann, “Complete”; Brown, Clements, and Grundy; McGregor et al.

7. For instance, humanitiesdata.com seeks to identify publicly available
data of interest to humanities research so as to support verification and collaboration; the *Journal of Cultural Analytics*, launched in 2016, has a platform for data publication and reviews, and its first “debate” explored the requirements of data reuse (Allison; Goldstone, “From”).

8. More recently, McGann describes the scholarly edition as “a model, a theoretical instantiation, of the vast and distributed . . . network in which we have come to embody our knowledge” (*New 26*).

9. Underwood and Sellers make this point in the online working paper (“How” 31) that accompanies their article (“*Longue*”) and published datasets and code (“*Code*”). The number of publications required for their argument emphasizes the necessity of a dedicated format to support data-rich literary history.

10. All of the fields I created are available for export—wholesale or selective—from the database. But due to spatial limitations not all fields are available for searching and browsing through its interface. As with the creation of future formats for data publication that combine sustainability and accessibility, an interface that enables users to select displayed fields would be preferable to the format I provide.

11. Full publication details for any of the authors, titles, or newspapers cited in this book can be accessed by searching the database. Where I refer to a specific publication event—including in chapter 6, for which specific text files were subjected to topic modeling—I provide the title ID (and, where quoting the text, the chapter number).

**Chapter 3**

1. Changes in printing technologies include the inventions of the cylinder and rotary presses, both of which significantly increased printing rates. Scholarship on political and cultural developments in nineteenth-century British and American literary culture is extensive and includes Brantlinger on literacy; McGill (“Copyright”) on copyright; Hewitt on taxation; Easley (*Literary*) on celebrity authorship.


3. While noting an “absence of local evidence of specific prices charged by booksellers,” Johanson contends that there was essentially no market in the Australian colonies for expensive multivolume novels and that readers were “not interested in buying . . . 6s editions” either (231, 213). Askew and Hubber comment that colonial lending libraries “probably reached only about 3%” of the population (122).

4. Editor, “*Australian Serial Literature,*” *South Australian Register*, July 23, 1867, 2, nla.gov.au/nla.news-article39194214
5. Editor, “Marjory’s Mistake”: Another Original Story for the ‘Miner,’” *Barrier Miner*, April 22, 1895, 3, nla.gov.au/nla.news-article44167084

6. For instance, “Another Admiring” writes, “SIR,—The story of ‘Andrew Fairfax’ was novel, interesting, and powerful, and the same may be said of ‘Ishmael,’ even though in many points he is very like ‘Andrew.’” “To the Editor,” *Euroa Advertiser*, February 23, 1894, 2, nla.gov.au/nla.news-article65534559


11. Editor, “A Novel without a Name,” *Newcastle Chronicle*, June 17, 1876, 3, nla.gov.au/nla.news-article110990578

12. Some of these projects have been published (Crittenden; Johnson-Woods, *Index*; Morrison, “Contribution”; Webby, *Early*) while others are available only on-site, as card indexes, or are unavailable.


14. For more information on these projects and recent steps by *AustLit* to index international newspaper fiction, see Bode and Hetherington (nn 2, 4, 5).

15. For those unfamiliar with the Australian context, Victoria was one of six colonies, along with New South Wales, Queensland, South Australia, Tasmania, and Western Australia.

16. Analysis of *Trove* uncovered five titles—including Dickens’s “A Tale of Two Cities” (#13814)—published in the *Age* prior to 1872.


19. NLA, “Australian Newspaper Plan.” Funding cuts to *Trove*, discussed in the conclusion, will likely change processes for selecting, and responsibilities for digitizing, historical newspapers.


21. For instance, to the (supplied) question “What does the collection cover?” the *British Newspaper Archive* answers, “The British Library’s newspa-
per collections are among the finest in the world, containing most of the runs of newspapers published in the UK since 1800” (findmypast, “About the British Newspaper Archive,” britishnewspaperarchive.co.uk/help/about). Conflating the cultural institution’s analog holdings with the commercial organization’s digital ones, this reply obscures the nature and scope of the digital collection while denying the existence of gaps between historical and digital records.

22. NLA, “Australian Newspaper Digitization Program Selection Policy.”

23. NLA, “What Is a Newspaper?,” nla.gov.au/content/what-is-a-newspaper


27. NLA, “Australian Newspaper Digitization Program Selection Policy.”


29. I have found no basis or explanation for this estimate, which occurs a number of times in Trove’s documentation and associated publications (see Berthon and Wan Wong; NLA, “Australian Newspaper Digitization Program Selection Policy”; NLA, “Trove Help Centre: New Titles: Why Isn’t My Newspaper Title Digitised?” help.nla.gov.au/trove/using-trove/digitised-newspapers/new-titles).

30. These Gordon & Gotch figures, and hence my assessment of the proportion of nineteenth-century Australian newspapers digitized in the relevant years, differ from those presented in an article in Victorian Periodicals Review (and corrected in the following issue of that journal). Working from digitized copies of the directories, I did not realize that some tables extended over opposing pages while others were on a single page, and my counting was therefore inaccurate. I appreciate—in the sense of perceiving, not so much of enjoying—the irony of discussing errors arising from digitization in a project that seeks to articulate a humanities approach to such documents.

31. For provincial newspapers in all three colonies, rates of digitization for those operating in 1890 based on Kirkpatrick’s figures (Country 47) are very similar to those operating in 1892 based on that Gordon & Gotch directory.

32. While some limit paratext to textual documentary features, McGann
(Textual) and McKenzie extend the concept to other material and social elements, ranging from cover art, font, binding, and illustrations to markings on individual copies.

33. Supplements are often perceived as peripheral or irrelevant to newspapers’ “real” contents (see Brake, “Lost”).

34. Such stories’ length makes them easier to overlook when zoning pages, and a zoning error is more likely to affect their discovery: short fiction is often composed of one article, meaning a single zoning error renders that title effectively invisible to my paratextual method. In contrast, extended fiction typically involves multiple installments with the same or similar paratext.

35. Extensive crowd-sourced manual correction of Trove’s OCR-rendered newspaper text has significantly improved its quality but is mostly directed at genealogical information and has not greatly improved the textual data collected in this project.

36. NLA, “API Technical Guide,” help.nla.gov.au/trove/building-with-trove/api-technical-guide. Chronicling America and Europeana Newspapers also provide APIs, but these are less useful for targeted searching and exporting of content due to the lack (in the case of Chronicling America) or only partial implementation (in the case of Europeana Newspapers) of article segmentation.

37. All appendices to this book are published alongside the digital version on the University of Michigan Press website.

Chapter 4

1. Other useful studies of literary anonymity include Brake, Print; Easley, First-Person; Griffin; Mullan.


4. In most cases neither gender nor nationality has been determined, but in a small number either a gender or a nationality has been identified. For instance, the widespread use of “Bertha M. Clay” as a pseudonym for fiction in American newspapers presents a convincing argument that these titles were American in origin. But the gender of the individual or individuals behind the pseudonym is unknown.

5. Story appears 1,068 times in the titles of fiction in the curated dataset, followed by tale, 457 times, and Australia(n), 432 times; romance and life, with 371 and 343 appearances, respectively, are the next two most common terms.
6. This title appeared in multiple provincial newspapers in 1892; it had been published in the colonies fourteen years earlier in the *Australian Journal* as “Found Guilty; or, The Hidden Crime” and was featured in the *New York Ledger* the year before that, in 1877, under the title “The Lord of Strathmore; or, The Hidden Crime.”

7. Determining if the opening paragraphs of hundreds of works indicate a national origin was clearly time consuming. Stanford’s Named Entity Recognizer (Finkel, Grenager, and Manning), employed to develop the stopwords list for chapter 6, might have offered an automated solution to this problem but would not have helped—and would, in fact, have created inaccurate inscriptions—for stories that cite locations but where contextual features indicate the narrative perspective is foreign to that place. Deciding which references should be considered indicative of a story’s origins was also complicated. For instance, should stories prominently set in a manor house, or referencing lords and ladies, be inscribed as British? In these and other instances I decided such references were not unambiguously indicative of national origin.

8. For an interesting discussion of original and changing textual meanings, see Frow, “Reading.”

9. See, for example, Boyd 5–6; Brake, “Writing” 64; Casey n8; Hughes and Lund, “Textual/Sexual” 144; Hughes and Lund, *Victorian* 103; Lund 26.

10. Such studies include Brake, “Writing” 61; Brantlinger 32; Mays 178; Lovell 9–10; Pearson 196.

11. Feminist literary historians have extensively debated whether men or women were more likely to use opposite sex pseudonyms. Some argue that women were more likely to use male pen names to avoid either the shame associated with middle- and upper-class women earning a living by writing fiction and/or the nineteenth-century perception of “women novelists . . . as inferior to male writers” (Casey n6; see also Sutherland 156, 159–60). Others propose that men were more likely to use female pseudonyms due to the nineteenth-century view of fiction authorship as feminized (Judd 82; Tuchman with Fortin 53). Among the colonial newspaper fiction identified in this project, titles by known male authors were inscribed as “female” only seven times, compared with eighty-one cases where the opposite occurs.

12. Interestingly, these non-British titles—“The Life and Adventures of Toby Frundle,” by Australian author Timothy Short, published in 1839 in the *Southern Australian*; and the earliest title discovered by this project, “Travels and Adventures in Southern Africa,” by South African author George Thompson (Esq.), published in 1828 in the *Sydney Monitor*—are the only ones attributed to a named author prior to 1843, when Dickens’s “American Notes for General Circulation” was published with attribution in the *Sydney Morning Herald* and *Southern Australian*. 
13. In 1841 the United Kingdom had a population of 26.7 million (Mitchell with Deane 8–9), compared with 17 million in the United States (U.S. Bureau of Statistics 249) and only 220,968 in the colonies (Australian Bureau of Statistics). These same reference sources are used for the later population statistics.

14. Of the fiction where nationality is “inscribed,” 29 percent of titles are “Australian”; where nationality is “known,” 24 percent are by an Australian author.

15. Average proportions of both British and women’s writing were notably higher in Western Australian newspapers. Perhaps the high rate of immigration in this colony—the discovery of gold transformed Western Australia’s population from under 30,000 in 1880 to over 170,000 in 1899—explains this difference, in that newly arrived readers, accustomed to female-authored periodical fiction, were in the majority. But with only fourteen Western Australian newspapers included in this study, averages for this colony have little effect on overall results.

16. Fiction by known authors of other nationalities made up 3 and 5 percent of titles in metropolitan and provincial newspapers, respectively. While this other national fiction deserves further investigation, because of the relative scales of publishing, I focus on American, Australian, and British fiction.

17. The curated dataset contains 3,792 titles from metropolitan newspapers and 5,249 from provincial ones. An additional 206 titles were published in three suburban newspapers: the Elsternwick Leader, Oakleigh Leader, and Port Adelaide News (defined as periodicals located at least ten but not more than thirty kilometers from the center of the colony’s capital).

18. In fact, fiction by other national authors was the most male-dominated category in metropolitan newspapers (81 percent), but the small number of titles involved (73) compared with the number by British authors (1,514) makes the proportional result less significant.

19. While eighty titles by known American authors appeared in provincial newspapers in 1891, in 1892 this fell to forty-four, declining to only twenty-one titles in 1896.

20. For instance, in the 1880s and 1890s, Harte was the most widely published American author identified by this project in metropolitan newspapers. But he was responsible for only nineteen titles, compared to forty-five by Braddon in these same decades.

21. The Australian Journal (a magazine rather than a newspaper) generally did not pay for fiction (Campbell 56). But both Ada Cambridge and Marcus Clarke “gained more from Australian serial rights than from English publication” (Stewart 23), with the Age paying Cambridge “the extremely large amount” of £197 for Australian serial rights to “A Black Sheep” (Morrison, Introduction xxvi).
22. In metropolitan newspapers between 1865 and 1879, where both gender and nationality were inscribed, 88 percent of “Australian” fiction was “male” authored, compared with 75 percent of “British” fiction and 84 percent of “American.”

23. The prevalence of pseudonymous publication of local authors in metropolitan newspapers complicates the interpretation of attribution and its relationship to cultural value. A considerably lower proportion of Australian than either American or British fiction is classified as “attributed,” both before and after 1880 (the proportions are, for Australian fiction, 46 percent before 1880 and 67 percent after; for American, 58 and 73 percent; and for British, 60 and 80 percent). However, much of this local writing was published under pseudonyms that clearly identified writers and aligned them with oeuvres. Indeed, some of the most well-known colonial authors used pseudonyms, including “Rolf Boldrewood” and “Maud Jeanne Franc.” When pseudonyms are included, Australian fiction in metropolitan newspapers was significantly more likely to be attached to an author name than other national categories, before and after 1880 (this is the case with 84 and 96 percent of Australian titles, compared with 67 and 79 percent of American and 64 and 90 percent of British titles).

24. Well-known examples include Ada Cambridge, Maud Jeanne Franc, and Mary Hannay Foott.

25. Of titles by known authors in provincial newspapers, 70 percent of Australian fiction was attributed, compared with 57 percent of British and 59 percent of American. Fiction by Australian men was attributed in 73 percent of cases compared with 56 percent for Australian women. For British and American fiction the gender division for men and women was 62 and 49 percent and 63 and 54 percent, respectively.

26. Other nineteenth-century commentators contradicted Smith. As Docker notes, A. G. Stephens’s 1899 introduction to his Bookfellow magazine highlighted colonial readers’ interest in local writing and challenged Henry Lawson’s complaints regarding opportunities for Australian authors (239).

27. Unless researchers are mistaken about dynamics relating to gender and nationality in British and American periodicals, for conceptions of such phenomena are largely based on the same approaches that produced a view of Australian newspaper fiction as predominantly British: studies of specific—usually canonical—authors; contemporaneous anecdote; and sampling of particular (often “small” or literary) magazines.

Chapter 5

1. Identifying and reprinting relevant or interesting content was a central part of the nineteenth-century newspaper editor’s job. Although “scissor-and-paste” journalism was discussed in a pejorative sense, there was no “clear
professional consensus . . . about how much copying was too much, or how soon was too soon to reprint another paper’s material” (Nicholson, “You” 275). On reprinting in colonial Australian newspapers, see Kirkpatrick, *Sworn* 8–9.


3. On the inadequacy of most critical bibliographies for studying fiction reprinting, see Johanningsmeier, “Frank” 285.

4. Nicholson used keyword searches to identify reprinted American jokes and slang in British newspapers (“Looming”). As noted in chapter 2, the Viral Texts project employs a text reuse discovery algorithm to identify reprinted passages in multiple genres (Cordell and Smith; see also Cordell, “Reprinting”; Smith, Cordell, and Dillon; Smith, Cordell, and Mullen).

5. Although in geospatial models the position of nodes is determined cartographically, their size and the connections between them are an effect of data availability.

6. For instance, Cordell draws conclusions from the finding that “Brownlow’s *Knoxville Whig* has the highest betweenness centrality in this network” (“Reprinting” 432). His more recent work—outlined in a blog post (“Two”)—steps away from this approach in seeking alternative ways of rendering network models so as to “discern the links that truly seem indicative of historical connections rather than data artifacts” (np). The historical constraints he imposes—according additional weight to examples of reprinting with temporal or geographical proximity—are promising in terms of the capacity to use network models for exploratory purposes but do not overcome the broader problem, also noted by Cordell, of constructing networks based on the highly partial datasets derived from mining mass-digitized collections.

7. More recent scholarship has challenged transparency as an ideal for understanding and governing algorithmic operations (Ananny and Crawford).

8. This method constructs millions of networks, containing all possible combinations of causes, to explore and contrast the range of possible dynamics.

9. For instance, statistical measures of probability could be used to extrapolate from observed republications to calculate the probability that the 50 percent of titles appearing only once in the curated dataset would be republished if the approximately 80 percent of colonial newspapers not digitized by *Trove* were included; a “forest” network could be devised to explore the system dynamics that result when all possible causes of reprinting in nineteenth-century newspapers are considered.

10. Measuring the number of unique and reprinted titles within a calen-
dar year aims to limit inclusion of instances where two newspapers published the same story from different sources (this happened but was much less likely to occur in the same year). However, this approach understates reprinting in the curated dataset because it excludes the limited number that occurred in consecutive years (for example, when one newspaper began publishing a story in December of one year and another in January of the next).

11. I have discovered 69 titles shared by the *Brisbane Courier* and *Queenslander*, 102 by the *Evening Journal* and *Adelaide Observer*, 77 by the *Telegraph* and *Week*, and 31 by the *Evening News* and *Australian Town and Country Journal*. Strictly speaking, the daily *Evening Journal* and weekly *Adelaide Observer* were not companions: the latter was paired with the daily *South Australian Register*. However, the same proprietors published all three, meaning the same structure and rationale as companion reprinting underpins the stories shared by companion newspapers. Thanks to Elizabeth Morrison for noting this distinction.

12. The list of titles in digital appendix 3 is undoubtedly incomplete. As noted already, I have not identified all fiction in nineteenth-century Australian newspapers, and Law mentions authors in connection with Tillotson’s—including F. W. Robinson, George MacDonald, and Henry Lucy—without listing syndicated titles (Law, *Serialising* 77).

13. Provincial newspapers that published fiction by these authors prior to the 1880s include the *Benalla Ensign*, *Capricornian*, *Fremantle Herald*, *Goulburn Herald*, *Newcastle Morning Herald*, and *Northern Star*.


15. The sources linking these authors to syndication agencies or agents are Colby; Jones; Johanningsmeier, *Fiction*; Law, *Serialising*; Turner. While Colby, Jones, and Turner focus on Tillotson’s, Johanningsmeier considers American syndication broadly, and Law explores a number of Tillotson’s competitors, including individual agents and companies.

16. The top twenty most published authors in metropolitan newspapers between 1865 and 1899, including the number of publications, were M. E. Braddon (64); Dora Russell (44); James Payn (33); Adeline Sergeant (31); B. L. Farjeon (29); Wilkie Collins (28); William Black (27); Ada Cambridge and George Manville Fenn (25); Margaret Oliphant and W. Clark Russell (24); Walter Besant (23); J. Monk Foster, Bret Harte, and W. E. Norris (22); Henrietta Eliza Vaughan Stannard (writing as “John Winter Strange”) (20);
F. W. Robinson, G. A. Henty, and Henry Herman (18); and David Christie Murray (17). Cambridge, an Australian writer, and Herman, a British author, are the two exceptions in this list: highly published authors not associated with well-known syndicators in the sources I have consulted. In Herman’s case, this is probably an omission of the sources, given his long-standing collaboration with Murray, who was syndicated by Tillotson’s and represented by A. P. Watt. Herman and Murray wrote several novels together, including three published in colonial newspapers. Also associated with these well-known agencies and among the top forty most published authors in colonial newspapers are Hawley Smart (16); S. Baring-Gould and Joseph Hatton (15); H. Rider Haggard, William Le Queux, and Eliza Lynn Linton (14); Robert Buchanan and John K. Leys (13); Hall Caine and Thomas Hardy (12); and Margaret Hungerford (11).

17. Of the eight newspapers centrally involved in reprinting fiction within the colonies, five are the most prolific metropolitan publishers of fiction in this study: the Queenslander (322 titles), Leader (302), Adelaide Observer (274), Evening News (265), and South Australian Chronicle (251). The other newspapers I have identified as heavily involved in reprinting—the Evening Journal, Telegraph, Week, and Brisbane Courier—are in ninth, tenth, eleventh, and thirteenth place, respectively.

18. As in figure 10, I assessed rates of reprinting among provincial newspapers based on the number and proportion of nonunique titles per year. While essential for comparison, this approach particularly understates provincial reprinting, which was more likely to occur in consecutive years than metropolitan reprinting was.

19. This project uncovered 124 titles published by both newspapers.

20. The Goulburn Herald published fourteen titles in common with the Cootamundra Herald and twenty-five with the Hay Standard.


23. Price Warung’s “An Endorsement in Red” appeared alongside Hennessy’s “The Mystery of Sea-Cliff Towers” in Hennessy and Harper’s 1898 Christmas annual and was subsequently republished in the Western Grazier in 1898.
24. The broader database also contains multiple short stories (completed in a single issue) published by these provincial newspaper syndicates.

25. Other features of the analog collection sometimes responsible for this same patchy publication of a sequence of syndicated titles include minimal availability of issues of a newspaper and poor microfilm quality, leading to digital pages so illegible that manual title correction was impossible.

26. Titles that in the 1880s were published by Cameron, Laing and Co. and appeared in New Zealand newspapers include “Denis Devine,” “In the Folds of the Serpent,” “The Mystery of Major Molineux,” “Marc Grecli,” “Dora Dunbar,” and “Days of Crime and Years of Suffering.” These publications predate the involvement of the Australian syndication agency S. & D. Reid, with New Zealand newspapers in the 1890s (Harvey). Such reprinting suggests the value of a future, comparative study of nineteenth-century Australian and New Zealand newspaper fiction.

27. I have decided, on the balance of evidence, that syndicate 1 ceased operating in 1892, but it is also possible it continued, publishing fiction I have allocated to syndicate 6. Supporting the first interpretation are the different newspapers involved (more than half of those associated with Cameron, Laing and Co., up to and including 1892, no longer published the same fiction after 1893); the different locations of these newspapers (syndicate 6 worked mostly with Victorian rather than New South Wales publications); and the different type of fiction published (syndicate 6 contained a large number of titles by unknown authors). Supporting the second interpretation are the involvement, in both syndicates, of many of the same authors (including Kenneth Hamilton, Harold M. MacKenzie, Atha Westbury, and James Joseph Wright) and many of the same newspapers (almost half of those associated with syndicate 1 appear in syndicate 6, although two-thirds of the newspapers in syndicate 6 did not feature in syndicate 1). A change in ownership might explain such dramatic shifts in publishing and business practices. But Cameron, Laing and Co. was acquired by S. & D. Reid in 1888 (Harvey 84), so the timing seems to discount this explanation.

28. While Johanningsmeier provides a long list of these companies (Fiction 96), he notes the difficulty of investigating even the major American syndicates—Bacheller’s and McClure’s—due to “the paucity of available manuscript and secondary materials” (71).

29. Syndicate 7’s serialization of four stories by American periodical author “Bertha M. Clay” could indicate an American company or an Australian agency that acquired fiction from American sources; alternatively, its inclusion of advertisements for colonial companies in its partly printed pages—for instance, for “Australian Explosives” and a Melbourne dentist—could indicate a locally based agency or an overseas syndicate producing partly printed pages especially for the colonial market. Intriguingly, a number of stories published by this syndicate appeared previously in either the
Evening Journal and/or the Adelaide Observer, raising the possibility that these metropolitan companions syndicated fiction for provincial newspapers. For syndicate 11, the mixture of international fiction and inclusion of miscellaneous American materials could suggest an overseas syndicate operating in the colonies or a local company extracting content from international newspapers.

Chapter 6

1. Dolin later adapted this argument to focus on nineteenth-century Australian readers rather than fiction (“Fiction”).

2. I consider only these three national categories because the samples available for other national literatures are too small for the integrated method used in this chapter.

3. The claim that nineteenth-century men and women wrote distinct types of fiction is foundational to much twentieth-century feminist literary scholarship (for example, S. Gilbert and Gubar; P. Gilbert). Numerous stylometry projects focus on distinguishing male and female authors based on word frequencies (for example, Olsen; Rybicki).

4. Examples abound, within and beyond digital humanities, and include Blei; DiMaggio, Nag, and Blei; Underwood, “Topic.”

5. The stop-words list also includes some common OCR errors, although ultimately, I elected to deal with this issue by excluding topics comprised primarily of such words.

6. Other strategies for relating topics and documents—for instance, considering documents in terms of the major topics they contain or deeming a topic present when it constitutes a certain percentage of words in a document—acknowledge the presence of all topics in each document. But the thresholds they implement (such as three main topics or 5 percent of words) introduce random and universal divisions into a spectrum and omit possible implications of the relationship between topics and documents: for instance, that the degree of presence of a topic—or all topics—is important for characterizing documents or that the threshold for significance for a topic is different depending on its prevalence in the corpus or its word associations.

7. I use regression trees, which express decisions as numerical choices; classification trees are also decision trees and offer true or false choices.

8. The code is written in MATLAB. Although proprietary, the software is commonly employed in academic research and was used in this case because it contains the necessary libraries for creating classification trees. Equivalent methods are offered in open-source numerical software such as Python, but the MATLAB implementation was superior.

9. Random samples were composed of equal numbers of titles from the
relevant categories, equivalent in size to 80 percent of titles in the smallest category: so, for exploring gender the method trained with random samples of men’s and women’s fiction, where both were equivalent in size to 80 percent of the available titles by women, and for exploring nationality it trained with random samples of American, Australian, and British fiction, where all were equivalent in size to 80 percent of the available titles by American authors.

10. To take an extreme example, predicting that titles with more than 0.001 percent of topic $x$ are by male authors might well be accurate in 99.99 percent of cases. But this high rate of predictive success would almost certainly come at the expense of very low accuracy in identifying female authors: the predictive success would simply indicate that most documents contained more than that very low level of topic $x$.

11. For instance, this integrated method could aid in categorizing fiction by unknown authors or, trained on an appropriate sample, in predicting fictional genres.

12. Variables can be nonlinear, so the calculations involved in producing decision trees are more complex than those for finding the topic with the highest—or lowest—levels in a particular category. For example, almost all of topic $x$ could be in titles by women. But this might mean that titles by American and Australian women tend to contain large proportions of words associated with topic $x$, while in titles by British women, these same words are barely present. In that case, the majority of topic $x$’s words would appear in women’s fiction, without the presence of that topic characterizing the majority of women’s fiction in the corpus. Rather, the root node signifies the topic that, above or below a set proportion in documents, most effectively categorizes—to continue the above example—the largest number of titles by women on one side of the threshold and by men on the other. The calculations underpinning secondary and subsequent nodes are more complex still, in that they involve identifying the topics and thresholds that most successfully predict a category after titles have been initially divided by the decision specified in the primary node.

13. Topic 16 is also prominent in fiction by the well-known British author of juvenile fiction George Manville Fenn: in order of its presence, his “Aboard the Sea-Mew” (#1570), “Iron Trials” (#6183), and “In Marine Armour” (#14014) are seventh, eighth, and tenth, respectively.

14. By contrast, the method had very low rates of success in predicting fiction—overall or Australian—published in the various colonies, suggesting that, for newspaper fiction and at the level of word patterns at least, the similarities between colonial literary cultures were significantly more pronounced than their differences.

15. In models with more splits, American and British stories are sometimes aligned with relatively high levels of topic 80. In such cases, they are
distinguished from Australian writing by the relative presence of military allusions (topics 96 and 7). This outcome suggests that, when representing rural colonial settings, British and American authors tended to do so in terms of military pursuits rather than agricultural or social ones.

16. While 14 percent of titles analyzed for this chapter were published prior to 1880, 21 percent of those categorized as Australian by the presence of topic 80 appeared before this time. In contrast, fiction first published in the 1890s comprises 51 percent of the titles analyzed and 44 percent of those identified as Australian by the presence of topic 80 (for the 1880s, the respective percentages are basically equivalent—34 and 35 percent). These results are the average proportion of titles categorized as Australian by the presence of topic 80 based on the thresholds in figures 6.3 and 6.4 (0.61 and 0.20 percent, respectively).

17. Other expressions of this argument include P. Carter; Gelder and Jacobs; Steele; and Trigg. Although dominant, the view that Aboriginal characters were absent from colonial fiction is not total. For exceptions, see Hadgraft 10–14; and Allen.

18. When topic 80 is removed, the accumulated average reduction in accuracy at four, six, and eight splits is 17 percentage points for Australian, compared with 22 for British and 25 for American fiction.

19. But there are suggestive precedents in Australian literary history for this view. For instance, Judith Wright’s 1965 *Preoccupations in Australian Poetry* ascribed nineteenth-century writing a “double aspect” or split “inner reality; first, and persistently, the reality of exile; second, though perhaps we now tend to forget this, the reality of newness and freedom” (xi). Thanks to Leigh Dale for alerting me to this association.

20. Of the one hundred titles ranked highest and lowest, respectively, in terms of the presence of topic 41, 69 and 47 percent were published in metropolitan newspapers and 59 and 25 percent were by women. The same trends occur when considering all Australian titles above the higher (4.67 percent) and below the lower (4.28 percent) thresholds for topic 41 defined in figures 6.6 and 6.7. In this case, 70 and 52 percent of titles, respectively, were published in metropolitan newspapers and 50 and 26 percent were by women.

**Conclusion**

1. As already noted, James “Skipp” Borlase’s proposed syndication of the title never eventuated (Sussex 105).

2. As noted in the introduction, beyond the curated dataset, the database also publishes all of the fiction discovered in analyzing *Trove*, including over twelve thousand titles not discussed in this book: a rich and virtually unexplored resource for future research. The database also allows users to
interact with Trove’s digitized newspaper collection to identify and index new fictional titles and installments and to edit the bibliographical fields of records not included in the curated dataset, and all of the text files. If used, this facility for text correction means that the text files in the database (though not those held by the University of Michigan Press) will become different from those analyzed for this project. I consider that the poor quality of the OCR-rendered textual data, and the opportunity to improve it, makes the ambiguity introduced by potential changes to this aspect of the curated dataset worthwhile in the interests of providing a better resource for future research.

3. While finalizing the book I discovered that, following extensive protests regarding cuts to Trove, its funding has been secured for another four years. This news is very welcome, but the point still stands that mass digitization does not solve all problems of access to the documentary record.