

Tracking Pirates through the Digital Archive: The Case of Dickens

MARY HAMMOND
University of Southampton

Introduction

Few nineteenth-century cultural historians can still be unaware of the mass digitization of post-1800 newspapers and periodicals which has been taking place over the last decade or so, and which has meant more primary materials are currently available to us than we have ever had access to before. At the time of writing, Wikipedia's list of these resources names digitization projects in over eighty countries around the world, and the total number is far greater when the entries by State and Territory and other internal sub-divisions are counted separately.¹ The list promises rich pickings: it suggests we now have access to thousands of newspapers, either free or via a subscription charge, which could potentially render the slow, painstaking process of scrolling through microfiche or turning over the fragile pages of large bound volumes in an archive all but obsolete. Even those of us who cling to the dusty isolation of archive searching in hopes of an occasional eureka moment must feel a certain amount of relief at the burgeoning accessibility of titles which we might otherwise never have had the time or resources to consult. Anglophone newspapers (my own current area of interest and the focus of this essay) are, perhaps predictably, given the uneven global distribution of research resources, among the most numerous and extensive in terms of online availability; between them the Nineteenth-Century British Library Newspapers and Periodicals databases, Welsh Newspapers online, the Nineteenth-Century Serials Edition, the Library of Congress's Chronicling America database, Australia's Trove project, the National Library of New Zealand's Papers Past project and a number of similar projects in Canada offer instant access to millions of pages of newsprint in English. At their best, these resources offer deep levels of searchability which should in theory make possible new ways of thinking about literature, social history, cultural change and many

I would like to thank James Mussell for his help in pointing me towards some recent work on digital history, and for sharing with me his own article, 'Digitization', to appear in a forthcoming Ashgate Companion volume.

¹ <http://en.wikipedia.org/wiki/Wikipedia:List_of_online_newspaper_archives> [accessed 10 July 2014].

other concerns of interest to Humanities scholars, and perhaps particularly to book historians. We can now easily pinpoint and study examples of ‘scissors-and-paste’ journalism, finding out which papers reprinted verbatim copies of news stories, fiction, poetry and other items, where they were copied from, and (if we pay attention to things like typesetting) using what material processes.² This means we can uncover hitherto unknown piracies of now-canonical works, and that we have begun to rediscover long-lost non-canonical texts which, based on their reprint histories, might once have been of equal popularity and importance. We can now also trace the history of a news story, an idea or even a joke over space and time by tracking the appearances of keywords, or what Digital Humanities scholars call n-grams (which can be phrases as well as single words when ‘n’ is replaced by the relevant number), and perhaps we might draw from the patterns of these appearances some conclusions about cultural drift, the influence of newspapers on taste formation, and the ideological roles of communication technologies more broadly. Through these means we might eventually come to understand both synchronic and diachronic nineteenth-century language systems more fully than ever before, able in effect to eavesdrop on the dialogic interactions of a lost world by extracting and analysing slices of its daily life, or comparing its shifting concerns over time and space. Whole projects have recently been dedicated to this new field of research, among them ‘Viral Texts: Mapping Networks of Reprinting in 19th-Century Newspapers and Magazines’, hosted and supported by Northeastern University in the USA and the National Endowment for the Humanities’ Office of Digital Humanities. This project’s laudable aim is to ‘develop theoretical models that will help scholars better understand what qualities — both textual and thematic — helped particular news stories, short fiction, and poetry “go viral” in nineteenth-century newspapers and magazines’.³ As this statement suggests, digitization has not only enabled what Franco Moretti has called ‘distant reading’, by which he means the analysis of patterns occurring across large numbers of texts.⁴ It has also — paradoxically — opened up new ways of *close* reading.

This has exciting possibilities for our teaching, as well as our research. At the University of Southampton, Year 3 undergraduates taking my ‘Victorian Bestsellers’ module have embraced these developments with gusto. In the module’s first presentation in 2011 I set them the task of tracking references to

² M. H. Beals, ‘Musings on a Multimodal Analysis of Scissors-and-Paste Journalism (Part 1)’, <<http://mhbeals.com/musings-on-a-multimodal-analysis-of-scissors-and-paste-journalism/>> [accessed 9 July 2014].

³ <<http://www.viraltxts.org/>> [accessed 10 July 2014].

⁴ See Franco Moretti, ‘Conjectures on World Literature’, *New Left Review*, 1 (Jan–Feb 2000), 54–68; *Graphs, Maps, Trees: Abstract Models for a Literary History* (London: Verso, 2005); and the regular pamphlets published by the Stanford Literary Lab (<<http://litlab.stanford.edu/>>).

one of the core best-selling ‘texts’ on the module (which include novels, poetry, news stories and cultural events) as they appear in the digital archive. I asked them to consider, first, what these appearances might be able to tell us — and what they cannot tell us — about a given text’s popularity and endurance. I then asked them to analyse how user-friendly the digital resources they investigated were, how these resources might invisibly direct the questions the researcher is tempted to ask, how the searchability and extent of various resources compared, and what pitfalls might await the unwary researcher tempted to read frequency of mention as a straightforward reflection of influence. Confident (overly confident, as it turned out) in my own relative familiarity with the basics of digital resource use, I prepared them with an interactive lecture in which I demonstrated several different digital resources, and I drew their attention to the cautionary notes already sounded by many Digital Humanities scholars. In 2008, for example, Daniel J. Cohen, one of the contributors to the forum ‘Interchange: The Promise of Digital History’, commented: ‘Machine-readable texts and metadata present interesting new possibilities for research since the computer can easily scan the entirety of a personal research collection for patterns, words, and other entities.’ But he also pointed out that the recognition of patterns is insufficient in itself as an answer to the kinds of questions Humanities scholars need to ask, and that such patterns actually pose their own type of methodological problem. As he put it: ‘Does it really tell us anything new about the Bible to discover that Jesus is the most frequently mentioned proper name?’⁵

Well-armed with such cautionary tales, my students were then let loose on the digital archive and charged with producing a paper reporting their findings. The results were illuminating. All the students, predictably enough, demonstrated acute sensitivity to the relative user-friendliness of the resources they encountered, and they were highly critical of those which did not measure up. The strongest students also seemed aware of the limits which the design of some digital resources might place on their research questions, and — still more impressively — some even speculated on the possible significance of the n-grams they found which were repeated verbatim or with only the smallest variants (usually typesetting errors) between newspapers. But others — a depressing majority — were so excited or overwhelmed by the number of hits their searches returned that in spite of my warnings they made the instant leap to taking numerical frequency as evidence of bestsellerdom, their ‘distant readings’ having apparently completely obliterated their ability (or their willingness) to close-read. As a cohort these students thus produced, on the one hand, genuine new

⁵ Daniel J. Cohen, Michael Frisch, Patrick Gallagher, Steven Mintz, Kirsten Sword, Amy Murrell Taylor, William G. Thomas III and William J. Turkel, ‘Interchange: The Promise of Digital History’, *The Journal of American History*, 95 (September 2008), 452–91 (p. 481).

research: some turned up evidence of how references to the Great Exhibition featured in printed popular discourse, for example, and for how long. Others were able to analyse the public disquiet caused by the publication of Mary Augusta Ward's *Robert Elsmere* in 1888. But, on the other hand, these results gave me an insight into how easy — and how dangerous — it is to accept digital results at face value, and they offered a salutary lesson when I went on to teach the module again the following year and to use digitized newspapers in more depth in my own research.

This essay aims to explore some of the pleasures and pitfalls of using digitized newspapers as primary evidence by offering as a case study my own experience of tracking a specific n-gram — the bigram 'Great Expectations' — through two large and very different databases: the British Library's Nineteenth-Century Newspapers and the Library of Congress's Chronicling America. One of my primary aims was to find out if there were nineteenth-century newspaper piracies of the novel in Britain and the USA which we did not already know about and, if so, where and when they appeared. In the event, I also found other unexpected uses for these resources which had equal quantities — though different types — of pleasures and pitfalls attached. It would be rash to claim that I offer solutions to any of the methodological problems habitually raised by the results of digital archive searches. My aims here are simply to share my recent experience of using digitized newspapers for book history research, and to draw attention to some of their benefits and limitations.

The Pitfalls

The temptation, as some of my students' work made clear, is to read the cyclical appearances and disappearances of texts, words and phrases in newspapers as straightforward evidence of patterns of culture or literary influence in action, and thus to view the spread of information through print as an organic, unmediated process driven by popular cultural consensus which is merely *reflected* in journalism. My students are not alone in these assumptions: in a panel on using digitized newspapers at a recent conference on Victorian Comedy, a presenter claimed to be able to use digital resources to pinpoint the genesis and track the spread of individual jokes in Victorian Britain, and to infer from these patterns how, and how fast, topical humour travels and linguistic fashions change. This may well be true — up to a point. But my own research has suggested that such conclusions might be premature, or at least that they might require some further research and analysis. For one thing, the relationship between the appearance of a given n-gram in print and its reception by and effect on readers are by no means clear or straightforward. How many of the readers of the newspapers tracked by the Victorian Comedy article presenter did not find the joke funny? How many of those who did find it funny when they read it

in a newspaper later spread it through other means which are not (and may never be) recoverable through the digital archive? We will probably never know the full answers to those kinds of questions, though readership historians have been working on them for years, but they serve here as a warning: it cannot be assumed that the simple repetition of a joke — or any n-gram — in numbers of newspapers is incontrovertible evidence of widespread acceptance or understanding.

In book-historical terms, individual n-grams are not always the result of editorial perceptions about readers' needs or potential receptiveness; they are sometimes the result of pragmatism or expediency. Many nineteenth-century newspaper proprietors — even those working in the pre-syndication period in the first half of the century — shared or pilfered information when necessary either via barefaced piracy or via the mutually agreed transfer of plates or sheets; for example, regional newspapers often slotted whole pages of national news or general articles into their own publications to fill space in slack news periods, and thus imported wholesale the language — including different word use — of other communities.⁶ Stories and poems lifted from other publications, sometimes without consent, were also often used to fill in blank spaces between local news and adverts. This does not mean that the receiving community understood or adopted (or laughed at) the linguistic immigrants; merely that it published, distributed and was confronted by them. And there, in terms of what the digital archive can tell us, the trail generally goes cold.

The sharing of printed and printable matter was not always attributable to piracy, either; nineteenth-century newspapermen and women created intricate matrices of familial and professional relationships which encouraged the sharing of printed material for reasons which went far beyond finance, politics or ideology. Dickens was only one among hundreds of powerful editors who regularly gave writing, sub-editorial and illustrating jobs to his family and friends and their wider connections, whose fame encouraged the development of textual cross-referencing as a kind of shorthand, and whose own journalistic practice commonly embraced the re-casting of material in new contexts. The essays in the recent collection *Charles Dickens and the Mid-Victorian Press, 1850–1870* demonstrate that while there is an element of truth to the long-established assumption that 'The press made Dickens — then Dickens made the press', there were vast global networks of professional and personal relationships in the nineteenth-century press of which his were but a part.⁷ The Viraltex project designers vividly illustrate this issue. Using 'computational linguistics tools and

⁶ See Meredith L. McGill, *American Literature and the Culture of Reprinting, 1834–1853* (Philadelphia: University of Pennsylvania Press, 2003).

⁷ John Drew, 'Introduction' to *Charles Dickens and the Mid-Victorian Press, 1850–1870* (Buckingham: University of Buckingham Press, 2013), pp. v–xxi (p. vii).

digitized corpora' they are developing large-scale national maps of the patterns formed by reprinted n-grams — exciting enough in itself — but by mapping them against other types of spatial data and census records they have also begun to see how some of the reprint patterns worked at the human level.

One community of reprinting partners [...] includes newspapers in Vermont, New York, Ohio, and Missouri, which was in the west of the United States during the period of our study. These connections suggest further research. When the network graphs suggested a close connection between the *Vermont Phoenix* (Brattleboro, Vermont) and the *Fremont Journal* (Fremont, Ohio), for instance, we investigated further, only to discover that the editors of these papers were brothers-in-law.⁸

Such findings suggest that while we can glean exciting new things from quantitative data, we need to continue to pay attention to more traditional methods of data recovery, and to remember that both personal and professional human interactions lie behind the statistics.

One more example serves to illustrate the point. Dallas Liddle has recently described some of the exciting new things that book historians might do with the amount and kind of information stored by PDF image files:

One kind of metadata to which researchers pay very little attention — except when our flash drives are getting full — is the size of the computer files required to store page images of newspapers. This figure would seem to have nothing to do with the characteristics of the original documents, but those used by *The Times Digital Archive* — comprising text, vector graphics, and compressed bitmap or raster graphics — actually has [*sic*] validity as an index of the information density of a typeset page.⁹

But the digitized, information-dense (or information-light) typeset pages whose traces appear in this metadata were created by human beings (sociable, emotional and error-prone as they are), and their inter-relationships and daily working practices are not usually recoverable via these means. We might well be able to make cautious assessments of patterns and trends in the ebb and flow of information as they appear in our computers' metadata, even identifying transitional moments in the evolution of journalism which we have never before noticed or been able to prove; but we cannot afford to jettison other means of literary archaeology quite yet. To do so would be to ignore the personal, the serendipitous and the examples of individual agency which so often lay behind the particular characteristics of nineteenth-century newspapers and periodicals.

⁸ David A. Smith, Ryan Cordelly and Elizabeth Maddock Dillony, 'Infectious Texts: Modeling Text Reuse in Nineteenth-Century Newspapers', *Proceedings of the Workshop on Big Humanities*, IEEE Computer Society Press, no date or page numbers. Text reproduced at <<http://www.viraltexts.org/>> [accessed 11 July 2014].

⁹ Dallas Liddle, 'Reflections on 20,000 Victorian Newspapers: "Distant Reading" *The Times* using *The Times Digital Archive*', *Journal of Victorian Culture*, 17 (June 2012), 230–37 (p. 231).

Relying exclusively on computer-generated metadata would be a bit like trying to determine the colour of an Archaeopteryx from the size of its footprints.

Equally, digital archives, however extensive, are still far from complete. While, for example, the British Library holds over 52,000 separate newspaper, journal and periodical titles which in total occupy over 45 kilometres (28 miles) of shelving, its digitization project — funded by the Joint Information Systems Committee — has focused on ‘London national newspapers, English regional newspapers, home country newspapers from Scotland, Wales and Northern Ireland, and titles in specialist areas such as Victorian radicalism and Chartism’, a selection which tells us more about current intellectual trends than about Victorian newspapers.¹⁰ As Laurel Brake warns us, ‘We need to remember that presentism is the dangerous companion of the digitized title’.¹¹ Databases are still subject to the often arbitrary decisions of digitizers and funding bodies. They are also subject to the variable abilities of programmers, cataloguers and search engines which often make cross-checking datasets difficult. Daniel J. Cohen has memorably claimed that ‘historians planning a digital project should think like architects, not like plumbers’, but it is increasingly clear that not all of them have done so.¹² And finally, as Melodee Beals points out, some newspapers — perhaps key newspapers — have been lost altogether, so ‘even if a newspaper has been selected for preservation, multiple editions and non-surviving issues mean that true certainty will always remain elusive, even with manual examination’.¹³

So, bearing in mind these caveats, what can digitized newspapers offer us, how might we supplement our findings from other sources (while remaining cognizant of their limits) and what might the results of such a synthesis look like?

Great Expectations as Bigram: The Piracies

While researching for *Charles Dickens’s Great Expectations: A Cultural Life, 1860–2012* (2015), I spent several months amassing all the references I could find to the novel in digitized nineteenth-century English-language newspapers by performing a search in the British Library’s Nineteenth-Century Newspapers and Periodicals and the Library of Congress’s Chronicling America databases. My initial aim was very basic: I wanted to find out if previous work on *Great Expectations*’ cultural lives had missed anything. There are many years of painstaking scholarship on stage adaptations, reviews and piracies of this novel, but digital collections give us new ways of searching for information and

¹⁰ Edmund King, ‘Digitisation of Newspapers at the British Library’, *Serials Librarian*, 49 (2005), 165–81 (p. 166).

¹¹ Laurel Brake, ‘Half Full and Half Empty’, *Journal of Victorian Culture*, 17 (June 2012), 222–29 (p. 225).

¹² Daniel J. Cohen et al., ‘Interchange: The Promise of Digital History’, p. 461.

¹³ M. H. Beals, ‘Musings on a Multimodal Analysis of Scissors-and-Paste Journalism’.

cross-referring it and have the potential to reveal printed references in places most of us would never have thought to look. As Patrick Leary pointed out some years ago: ‘What is most striking, and often quite useful, about this sort of fishing expedition is how often the sources in which one finds a “hit” are utterly unexpected [...] No amount of time spent in the library stacks would have suggested to me that any of those sources would be an especially good place to look’.¹⁴

The first results on potential piracies were promising. Dickens scholars have long been aware of the main, very publicly pirated version: in 1955 K. J. Fielding wrote a short but informative article on a piracy by the editor of the *Eastern Province Herald*, Port Elizabeth, South Africa, who began serializing the novel on 5 March 1861, before being forestalled when Dickens took legal action.¹⁵ The editor of the *Herald* seemed surprised that Dickens had prevented his circulation of the novel, but anyone more familiar with Dickens’s media profile in this period (and anyone who has performed the right targeted search in digitized newspapers since) would not have been surprised at all. While (I discovered during the search process) British newspapers frequently and usually disapprovingly reported the prices Dickens was paid for American and other foreign rights, and often tacitly equated his active and legally informed pursuit of a fair price for his works with poor taste, a targeted search in British newspapers focused exclusively on appearances of the name *Eastern Province Herald* between March and August of 1861 revealed that in this case they took his side. The piracy was a newsworthy item countrywide and engendered numbers of outraged responses from correspondents in papers ranging from *Lloyd’s Weekly* to the *Falkirk Herald*. In this case, it seemed, however denigrated his zealous copyright-watching might be at home, Dickens was perceived as performing a useful public role as legal watchdog overseas.

This was illuminating in itself, but there was more buried piratical treasure to come. Dickensians have long known that extracts from his novels were often reprinted in regional newspapers. Indeed, George Ford asserts that the phenomenal success of the *Pickwick Papers* in 1836 after a slow start was as much due to the extraction and reprinting of key Sam Weller episodes by William Jerdan, editor of the high-circulation weekly journal the *Literary Gazette*, as to the invention of Sam himself. Ford also tells us that Disraeli claimed he had never read any of Dickens’s novels in their entirety, only extracts in a newspaper.¹⁶ To

¹⁴ Patrick Leary, ‘Googling the Victorians’, *Journal of Victorian Culture*, 10 (Spring 2005), 72–86 (p. 77).

¹⁵ K. J. Fielding, ‘The Piracy of “Great Expectations”’, *Notes and Queries* (November 1955), 495–96.

¹⁶ George Ford, *Dickens and his Readers: Aspects of Novel Criticism since 1836* (New York: W. W. Norton, 1965), p. 91.

date I have found no scholarly discussions of reprinted extracts from *Great Expectations* appearing in newspapers outside the practice of reviewing, but in fact there were dozens and they can tell us a great deal about Dickens's presence in both British and American culture in the 1860s, the period of his career so often characterized in both countries as one of reputational decline. Between March and August 1861, during the last few months of the novel's serialized run in Dickens's weekly *All the Year Round*, many British papers including the *Worcestershire Chronicle*, the *Salisbury and Winchester Journal*, the *Bath Chronicle and Weekly Gazette*, the *Leeds Times* and the *Kentish Chronicle and General Advertiser* reprinted extracts varying between a column and three columns in length. Almost invariably, perhaps in a desperate attempt (repeated throughout the Anglophone world) to rescue vestiges of the early, playful Dickens from what was generally felt to be the somewhat relentless gloom of his later period, they select the humorous episodes: Wemmick's marriage to Miss Skiffins; Pip and Estella taking their laborious tea in a London hotel; Mr Wopsle playing Hamlet; Joe visiting Pip's bachelor apartment in the metropolis and repeatedly dropping his hat. The extracts are given no preamble and are seldom commented on: Dickens's name seems to speak for itself. This inevitably poses the questions: was he aware of this practice? Did he sanction it? Or was he simply powerless to stop it? A name search in the multi-volume 'Pilgrim' edition of *The Letters of Charles Dickens* revealed no mentions of any of these publications.¹⁷ Either he was unaware of the extent of the piracy, or he knew it was worth turning a blind eye to it because any publicity is good publicity. Having spent five years getting intimately acquainted with Dickens's remarkable ability to negotiate the complexities of the Victorian print marketplace, I am inclined to believe the latter.

The digital archive offers some further useful possibilities in relation to these extracts: the surrounding material (also keyword-searchable) makes possible some fascinating intertextual readings. The *Kentish Chronicle and General Advertiser* (1859–67) is a particularly interesting example, since it circulated in Dickens's own community during his country sojourns at his Kent house, Gad's Hill, in this period. A search for *Great Expectations*-themed words in the issue for Saturday 16 March 1861 (the issue in which the extract 'Mrs Belinda Pocket' appears on p. 4) turned up some extraordinary coincidences (if indeed coincidences they were). Surrounding this stolen extract from the novel were stories of a convict who had escaped from a chain gang in Chatham and later been recaptured (p. 3), a drowning in Ramsgate harbour (p. 1), a breach of promise trial (p. 2), and a man who had been sent to Van Diemen's Land for machine-breaking and afterwards disappeared without trace, never learning that

¹⁷ *The Letters of Charles Dickens* (Pilgrim Edition), ed. by Madeline House, Graham Storey and Kathleen Tillotson, 12 vols (Oxford: Clarendon Press, 1965–2002).

not only had he been pardoned, but he had also come into an unexpected fortune (p. 3). We will probably never know whether or not these stories were selected in order to pick out some of the themes of *Great Expectations* as a way of supplementing the understanding of readers who were also reading (or intended to read) the novel, though I seriously doubt it, since they are fairly typical of the content of many regional newspapers. But they certainly give us a glimpse into the astonishing topicality of Dickens's thirteenth novel for British readers in the 1860s (despite its setting in the 1810s and 1820s). It does not require much of a stretch, either, to imagine Dickens (an inveterate literary magpie) reading just such stories in the *Chronicle* at the breakfast table at Gad's Hill before scuttling off to write the next chapter of Pip's life.

Extracts also appear in US newspapers, though they are understandably rarer here due to the priority given to news from the Civil War battlefields. There are, however, some real treasures among digitized US newspapers. On 7 October 1861, for example (after the serialization had ended in August of that year and the first volume editions had been published on both sides of the Atlantic), the *Cincinnati Daily Press* printed the extract of Wopsle playing Hamlet. There is far less obvious topicality to this reprinting, since in this paper the extract rubs shoulders with reports on gunpowder supplies, adverts for soldiers' cough mixture, and short jokes and humorous sketches. This provides us with further strong evidence that American and British readers encountered very different versions of Dickens's works, something that Dickensians have previously had to infer from a comparative analysis of British and American volume editions and criticism.

But the newspapers in the *Chronicling America* database had more important treasures in store. In the course of looking for more of these reprinted extracts I located an unusually long one of a whole chapter in length (the first, in which Pip meets Magwitch) appearing on 3 June 1861 in the working-class New York 'penny' morning daily, the *Sun* (circulation around 60,000). Why a full chapter, and not a particularly funny one at that? I pulled up the next issue — and found the second chapter. Further searches showed that subsequent issues also had their pockets full of ill-gotten gains; I had stumbled on a hitherto unknown pirated version of the full text, and in fact this turned out to be my major digital discovery. The *Sun* printed an unillustrated chapter from *Great Expectations* almost every day between 3 June and 11 September 1861, without attribution to anyone except Dickens, without mention of any 'ready-print' arrangement (an early form of syndication) with other papers or magazines, and seemingly unconcerned about copyright infringement. Copyright legislation was certainly a hit-and-miss affair in the USA in this divisive period,¹⁸ but since Dickens had

¹⁸ Charles A. Johanningsmeier, *Fiction and the American Literary Marketplace: The Role of Newspaper Syndicates in America, 1860–1900* (Cambridge: Cambridge University Press, 1997), p. 36.

carefully arranged for simultaneous publication of the serial in *Harper's Magazine* and volume publication was through an authorized publisher, T. B. Peterson's of Philadelphia, how was this particular piracy possible, and how did it go unpunished? Internal textual evidence indicates that the pirated version follows the *Harper's* printing, not that of *All the Year Round* (there are small but significant textual differences resulting from the fact that Dickens — working at top speed — sent the *Harper's* copy by steam-ship the moment he had finished it, but had a few more days' grace in which to correct the copy for publication in *All the Year Round* in Britain). But the most surprising thing is that Dickens's American publishers seem neither to have known nor cared about it, though it must surely have cut into the sales of the first American volume editions in the summer of 1861. There are no letters in Dickens's correspondence about it, no letters to the Press from *Harper's* and nothing in the firm's official history.¹⁹

The Chronicling America website itself, which gives useful potted histories of all its digitized titles, simply added a new layer to the mystery. It provided the *Sun's* operational dates (1833–1950) and suggested that, by 1834, it 'had the largest circulation in the United States', a fact which suggests an exciting extension of the known readership for Dickens's novel, and circulation among a class of reader not normally associated with his later, more complex works. But the website also claimed: 'Its rising popularity was attributed to its readers' passion for the *Sun's* sensational and sometimes fabricated stories and the paper's exaggerated coverage of sundry scandals.'²⁰ These are not exactly qualities one normally associates with Dickens, and certainly not with *Great Expectations*, which (despite Margaret Oliphant's characterization of it in *Blackwood's Magazine* as a 'sensation' novel akin to Wilkie Collins's *The Woman in White*)²¹ tends to be thought of now as a complex and rather bitter novel from his 'darker' late period rather than a sensational popular success.

An online search for information about the *Sun* finally led me (as such searches so often do) to an old print source: an official history of the paper published in 1917. Its discussion of this piracy (though the author is careful not to call it that) was illuminating; it was the result of the temporary takeover of the *Sun* by an evangelical editor who wanted to clean up its act and felt Dickens was respectable enough to do the job. Unfortunately for the well-meaning editor, the *Sun's* readers were unimpressed and the paper's circulation plummeted so drastically that by the end of 1861 the management had to change again.²²

¹⁹ J. Henry Harper, *The House of Harper: A Century of Publishing* (New York: Harper's, 1912).

²⁰ <<http://chroniclingamerica.loc.gov/lccn/sn83030431/>> [accessed 8 July 2014]

²¹ [Margaret Oliphant], 'Sensation Novels', *Blackwood's Edinburgh Magazine*, 91 (May 1862), 564–84. See Deborah Wynne, *The Sensation Novel and the Victorian Family Magazine* (Basingstoke: Palgrave, 2001) for a discussion of *Great Expectations* as 'Sensation Novel'.

²² Frank M. O'Brien, *The Story of the Sun, New York, 1833–1928* (New York: George H. Doran, 1917), p. 137.

Dickens might have been considered relatively respectable, and the digital evidence had initially suggested that his thirteenth novel had a greater circulation among working-class American readers than we had previously assumed. But by supplementing this with a little more research in other sources, we are able to put some detail on the discovery, which tells us that in 1861 his status among such readers was far from assured and that he was in fact experiencing fierce competition for readers from newer and more sensational authors — and in some cases, apparently, he was losing.

The Cultural Phenomenon

The next phase of my research took a different tack. In tracking piracies of Dickens's novel, I had been struck by the number of references in newspapers to the phrase 'great expectations' itself both with and without initial capitals. Clearly, it had meant something before October 1860 when Dickens had chosen it as his title.²³ Indeed, as Edgar Rosenberg points out in his masterly Norton Edition of the novel (1999), the phrase was very well known to Victorians and — perhaps more significantly — Charles Lever had used it in passing in number 7 (Ch. 9) of his serial *A Day's Ride*, just a couple of days before Dickens began the novel that bumped Lever's off the top spot in *All the Year Round*. As editor of the journal, Dickens had, of course, read the proofs of those chapters while contemplating what to call his new serial.²⁴

Lever's use of the phrase by his ne'er-do-well hero Potts is a throwaway one: in trying to explain some drunken bad behaviour to people whom he wants to impress, Potts writes to them that he is forced to make his apologies by letter as his uncle has died unexpectedly and he must leave in a hurry for the funeral. In order to create a still better impression, Lever's Potts writes: 'I vaguely hinted at great expectations, and so on.' This — as I will shortly show — was a common usage of the phrase in this period; it tended to mean expectations of money or success which carried a vague hint of the vainglorious. But from 1860/61 onward, popular usage of the phrase 'great expectations' was so consistently attached to mentions of Dickens or to assumptions about knowledge of his story, that I began to wonder whether he had changed perceptions of its meaning in some way. This was a whole new area made possible by digital archive searching, although it offered a whole new set of potential problems also.

My experience of tracking piracies and marking students' work had already warned me not to read too much, too uncritically, into numerical patterns apparently emerging from lists of 'hits'. But online newspaper archives, like hard-copy archives, can surely tell us *something* when an n-gram achieves critical mass

²³ See letters to John Forster throughout October 1860 in *Letters of Charles Dickens*, vol. ix.

²⁴ Charles Dickens, *Great Expectations*, ed. by Edgar Rosenberg (New York: Norton & Co., 1999), p. 422.

through repetition and begins to demonstrate potentially significant patterns. We might not — without concrete evidence — be able to pinpoint the exact moment when a given n-gram multiplies, or changes shape or location; there are simply too many pieces of the puzzle missing to enable such precision (a problem with which most historians — book historians included — are depressingly familiar). But trends are a different matter; in the time-honoured book-historical tradition, the idea of trends offers us sufficient looseness and sufficient scope for addition, correction and adjustment as new information becomes available to be useful up to a point. They can also offer us navigational aids as we continue our search, pointing us to potentially rich seams of material or to potentially significant periods of time.²⁵

With such hopes in mind, I performed a bigram search for ‘Great Expectations’ in the British Library’s Nineteenth-Century Newspapers database in twenty-year blocks, starting at 1 January 1780 and ending at 31 December 1919. The numerical results were as follows:

TABLE 1

Date range (1 Jan–31 Dec)	Number of hits	Number of newspapers	Average references per newspaper
1780–1799	122	21	5.8
1800–1819	198	38	5.2
1820–1839	628	89	7.05
1840–1859	1650	127	12.9
1860–1879	4171	179	23.3
1880–1899	3723	115	32.3
1900–1919	2017	104	19.3

The table makes it fairly clear that, even allowing for the explosion in overall numbers of newspaper titles in the nineteenth century, the phrase ‘great expectations’ went through a significant set of changes in frequency of use during this period. Immediately apparent is the fact that its use was on the increase even *before* Dickens chose it as his title, numbers of references to it more than doubling in the two decades 1840–1859 and the average use per title also increasing almost twofold. If these figures are to be believed, Dickens seems to have had his finger on a cultural pulse. Predictably enough, given his fame, references per newspaper almost doubled again in the twenty years following the novel’s publication in 1860/61. More significantly, perhaps, a closer look at the specifics of these appearances reveals that reprinted extracts, announcements of

²⁵ See, for example, Simon Eliot’s *Some Patterns and Trends in British Publishing, 1800–1919*, Occasional Papers, 8 (London: Bibliographical Society, 1994) which has proved seminal for a number of subsequent studies using bibliometrics.

new editions, mentions of additions of the novel to library stock, reviews and reprints do account for some of this increase, but by no means all; the phrase clearly took on a new level of currency of a different sort in this period. The twenty years after that — somewhat surprisingly — see another large increase; the possible influence of Dickens's text not only rumbled on, but it even burgeoned for a while before beginning to peter out in the early decades of the twentieth century, just as his critical and popular reputation went into one of its periods of decline. Further targeted searches which I have yet to perform might put still finer detail on these statistics. For example, it might be significant that while there were fewer references to the phrase 'great expectations' overall in the period 1880–1899 than there had been in the previous two decades, on average those newspapers that did still use it were doing so more frequently. The British Library's database enables us to search by title and by region, which potentially enables us to find out which digitized titles in which region of the UK clung on the longest to its use of the phrase.

There is another level of searching possible here, though, beyond the numerical patterns, and this is where my primary interest lies. Exactly *how* is the phrase being used in each twenty-year period? Does the publication of Dickens's novel contribute to any changes in its semantic range? In fact, a closer-grained search reveals that before Dickens appropriated it for the title of his novel in 1860 it did indeed tend to be used in a slightly different way, carrying with it both high hopes and a certain frisson of anxiety about their possible failure, and a touch of hubris. It was used in this way in Lever's novel, as we have seen; it was also used in this way in announcements of forthcoming entertainments, as in: 'Great expectations are entertained that the approaching Music Meeting at Worcester, from the respectability [*sic*] of the stewards, will be as numerously attended as any former Festival for some years past.'²⁶ It was also used this way in political commentary, as in: 'The commencement of a session of Parliament is often looked to with a certain anxiety, if not with great expectations, of legislative benefits.'²⁷ It was often used retrospectively for reviewing entertainers whose performances ended up being a disappointment: 'The great expectations raised of [the singer] Mademoiselle Nau have not been realised [...] she was heralded with too much pomp'.²⁸ And it was also often used mid-century in reports on the Great Exhibition of 1851, where disappointment seemed inbuilt. 'I had great [financial] expectations of the Great Exhibition', the Director of the Midland Railway told the company's half-yearly meeting in 1851, 'and I must now confess that, if we had looked at the subject with more calmness, we should have arrived

²⁶ *Gloucester Journal*, Monday 7 August 1797, p. 3.

²⁷ *Dundee, Perth, and Cupar Advertiser*, Friday 24 January 1851, p. 2.

²⁸ *The Era*, Sunday 20 October 1850, p. 11.

at a different opinion'.²⁹ Less frequently, the phrase appears with better results: 'Last evening, the celebrated Berlin male choristers [...] fully realised the great expectations to which the preliminary announcements of the executive committee had given rise.'³⁰ In all the digitally available British newspapers I have examined, the naivety and the sense of riding for a fall which hovered over the term 'great expectations' were already clearly part of cultural consciousness before 1860. But the sharper edges of snobbery, greed and final bitter comeuppance seem to be almost entirely inventions of Dickens's, and they quickly spread way beyond the confines of his novel.

Among the most common passing references to the novel post-1860 are reports of court cases titled 'Great Expectations', in which the defendant has been caught out lying about his or her financial prospects with the intention to deceive.³¹ It also appears as the title to articles and stories about greed and naivety, as in this one, reprinted from *Our Old Home: A Series of English Sketches* (1863) by Nathaniel Hawthorne, which tells of his time as US Consul in Liverpool (1853–57). The newspaper, not Hawthorne, appends Dickens's title to the tale:

GREAT EXPECTATIONS

One day a queer, stupid, good-natured, fat faced individual came into my private room, dressed in a skye-blue [*sic*], cutaway coat and mixed trousers, both garments worn and shabby, and rather too small for his overgrown bulk. After a little preliminary talk, he turned out to be a country shopkeeper, (from Connecticut, I think,) who had left a flourishing business, and come over to England purposely and solely to have an interview with the Queen [...] the shopkeeper, like a great many other Americans, had long cherished a fantastic notion that he was one of the rightful heirs of a rich English estate [...] I had never had so satisfactory a perception of a complete booby before in my life; and it caused me to feel kindly towards him, and yet impatient and exasperated on behalf of common sense [...]³²

Dickens's attachment of greed, laziness and self-aggrandisement to the phrase seems to have caught on. As far as digitized newspapers are able to tell us, this was not universally the case before December 1860.

Such usage continued through the 1870s and 1880s. Then, in the 1890s, came the cartoons. In the General Election year of 1893, *Punch* used the title 'Great Expectations' for a cartoon depicting Chamberlain and Balfour falsely acting like friends over a potential political alliance between the Liberal Unionist Party and

²⁹ Report on the half-yearly meeting of the Midland Railway, *Leicestershire Mercury*, Saturday 23 August 1851, p. 3. See also 'The Great Exhibition', *The Era*, Sunday 10 August 1851, p. 11.

³⁰ *Morning Post*, Saturday 2 November 1850, p. 2.

³¹ See, for example, 'Great Expectations', *Portsmouth Evening News*, Saturday 23 September 1882, p. 2; 'Great Expectations: A Captain's Reversionary Interest', *Daily News*, Thursday 7 July 1898, p. 8 and 'Great Expectations not Complimentary to Clifton Folk', *Western Times*, Saturday 14 May 1904, p. 3.

³² *John o'Groat Journal*, Thursday 15 October 1863, p. 4.

the Conservatives which might somehow benefit Ireland (which, as the title now suggests, it was unlikely to do).³³ In 1895, the magazine *Pick-Me-Up* published a cartoon strip called 'Great Expectations' which shows a young man receiving a summons from a sick elderly Uncle from whom he expects to inherit a fortune. The young man arrives excitedly on a Sunday and dutifully begins to nurse the old man, evidently keen to safeguard the terms of the will. A week later the Uncle is fully recovered and the young man an exhausted, disgruntled (and still poverty-stricken) wreck.³⁴ This particular joke ran in cartoon form well into the twentieth century, and the patterns suggest that Dickens's thirteenth novel did not just adopt the concept of 'great expectations' as it was understood in the nineteenth century, but that it may also have added significantly to the range of its cultural meanings. There is a sense after the novel's publication that to entertain 'Great Expectations' is to court disaster — and that it is, in the final analysis, no more than the entertainer deserves.

The Reviews — and Some Conclusions

One final benefit emerging from these digital searches was the number of reviews of the novel I was able to find which appeared in regional British and American newspapers. Previous surveys of the novel's critical heritage have concentrated on the big periodicals such as the *Atlantic Monthly*, *The Times*, the *Saturday Review* and the *Spectator*. But my search turned up more than 20,000 words of reviews in other papers, some of them no more than a paragraph and some (inevitably) simply quoting verbatim from the big journals, but many of considerable length, originality and detail. These reviews demonstrate that even in the final decade of his career Dickens was clearly still thought to be of interest and importance to readers far beyond the urban middle classes whom, conscious now of competition from 'serious' literary novelists such as George Eliot, he was increasingly trying to target and impress. Despite his constant iteration of his love for the wider 'public', Dickens's late career is marked by a desire to exert new levels of control over his reception, both in the dramatic re-enactment of his most famous scenes in a series of public reading tours, and in the markedly different flavour of his late works. Digitized newspaper reviews give us an exciting glimpse of how this attempt was received by his diverse public.

On balance, *Great Expectations* was given a mixed reception by these papers. Most of the reviewers — who are, interestingly, assiduous in noticing individual numbers as well as reviewing the whole — think the story begins well, but then deteriorates. Many of these early instalment reviews express the opinion that it represents a return to his early form; there is real hope here for a late revival after

³³ *Punch*, Saturday 18 March 1893, p. 132.

³⁴ *Pick-Me-Up*, Saturday 19 January 1895, p. 254.

what, clearly, had been general disappointment over *A Tale of Two Cities*. The *Leamington Spa Courier* was certainly hopeful: ‘Dickens’s new tale in All the Year Round, “Great Expectations,” which opened in the last week’s number, promises well. So far as it goes, it has all the freshness of his early writings, and calls back to the memory “Oliver Twist” — decidedly the best book he ever wrote’ (Saturday 8 December 1860, p. 2). The *Dundee Advertiser* was reminded of a different early tale: ‘His recent story “Great Expectations,” is, we are glad to observe, thoroughly Pickwickian’ (Wednesday 29 January 1862, p. 4). But by March of the following year, by which time Dickens had written Pip off the marshes and moved him to London for the second phase of his expectations, the disappointment had set in — and numbers of newspapers used the novel’s title to underpin its failure. ‘Mr Dickens’s “Great expectations” have decidedly disappointed the *great expectations* raised by themselves’, the *Hereford Journal* complained; ‘the work does not improve with succeeding numbers’ (Wednesday 13 March 1861, p. 5).

By the time the one-volume edition was released in June 1861, however, two months before the serialization was complete, many reviewers had reassessed their position — testament, perhaps, to Dickens’s ability to create something that was greater than the sum of its parts. *The Literary Examiner* announced: ‘All who have read this story from week to week, as it appeared in the vigorous and entertaining pages of Mr. Charles Dickens’s popular journal, should join those who now read it for the first time as a finished work. It is a finished work in the best sense’ (Saturday 20 July 1861, pp. 4–5). The *Saturday Review* concurred: ‘Mr. Dickens may be reasonably proud of these volumes. After the long series of his varied works — after passing under the cloud of *Little Dorrit* and *Bleak House* — he has written a story that is new, original, powerful, and very entertaining [...] It is in his best vein, and although unfortunately it is too slight, and bears many traces of hasty writing, it is quite worthy to stand beside *Martin Chuzzlewit* and *David Copperfield*’ (Saturday 20 July 1861, pp. 69–70). American papers were more hopeful still: ‘In our opinion, “Great Expectations” is a work which proves that we may expect from Dickens a series of romances far exceeding in power and artistic skill the productions which have already given him such a pre-eminence among the novelists of the age’, claimed the *Atlantic Monthly* on September 1861 (pp. 380–82). The *New York Daily Tribune* agreed: ‘No one of the former productions of Mr. Dickens is richer in the peculiar qualities which have given him such a wide and solid popularity than the present remarkable creation of his genius’ (25 August 1861, p. 2).

Somehow, sick and emotionally exhausted though he was by the summer of 1861, Dickens had pulled it off. But just this brief glimpse at the wider pool of reviews demonstrates that the ‘classic’ novel *Great Expectations*, so often now glibly described as a timeless masterpiece, was in fact a slightly baffling experience for most Victorians reading it week by week. They largely loved Magwitch and

Joe and hated Miss Havisham. They thought it an improvement on some previous novels, but still falling far short of his first works. They preferred the marsh to the London scenes, and believed in Pip the boy but not Pip the adult. We have long known that it received a mixed reception from the heavyweight literary critics, but these other reviewers give us a slightly more complex picture. From them we begin to get a sense that *Great Expectations* was not just a new novel being assessed for its literary merits; it was also, as part of an iconic author's *oeuvre*, a cultural phenomenon with a huge weight of hopes and expectations attaching to it. There is a sense here that it is being anxiously scoured for traces of a fondly remembered literary past that was already, by 1860, fast receding into the distance. Dickens, media-savvy though he was, seems to have belonged in the popular imagination to an older world driven by public consensus rather than marketing; his celebrity was thus a delicate thing, perceived as created by his public and carrying considerable responsibility: 'An outrageously absurd book by a great writer may be regarded as a breach of the public confidence', opined the disappointed *Morning Post*; 'it is in literature what a gigantic fraud is in commerce' (Wednesday 31 July 1861, p. 3). The juxtaposition here of literature and commerce, the clash between honour, duty and profit, is a reminder of just how complex, dynamic and modern the media marketplace in which Dickens was working in the final decade of his life had become.

I remain aware that professional reviews are not representative of the feelings of ordinary readers, and that — if I am to heed my own warnings — there is much more work to be done. But the recent development of digitized resources which has largely made possible the story I have begun to tell has had one huge benefit despite its many limitations: restored to a fuller historical context, for me *Great Expectations* has receded from the high-definition glare of its post-Victorian canonization, climbed off its pedestal, and moved back into the rich tapestry of intertextual references and fierce competition to which it more properly belongs. Digitized newspapers have not provided me with all the answers, but they have enabled me to ask new questions. As William G. Thomas III puts it, at its best 'To do digital history [...] is to create a framework, an ontology, through the technology for people to experience, read, and follow an argument about a historical problem'.³⁵ In the final analysis, that can only be a good thing.

³⁵ William G. Thomas III, 'Interchange: The Promise of Digital History', p. 454.