Commonplacing the public domain: reading the classics socially on the Kindle

Abstract

Amazon leads the market in ebooks with the Kindle brand, which encompasses a range of dedicated e-reader devices and a large ebook store. Kindle users are able to share the experience of reading ebooks purchased from Amazon by selecting passages of text for upload to the Kindle Popular Highlights website. In this article, I propose that the Kindle Popular Highlights database contains evidence that readers are re-appropriating commonplacing – the act of selecting important passages from a text and recording them in a separate location for later re-use – while reading public domain titles on the Kindle. An analysis of keyness in a corpus of 34,044 shared highlights from public domain titles suggests that readers focus on words relating to philosophy and values to draw an understanding of contemporary society from these classic works. This form of highlighting takes precedent over understanding and sharing key narrative moments. An examination of the top ten most popular authors in the corpus, and case studies of Jane Austen's *Pride and Prejudice* and William Shakespeare's *Hamlet*, demonstrates variation in highlighting practice as readers are choosing to shorten famous commonplaces in order to change their context for an audience that extends beyond the original reader. Through this analysis, I propose that Kindle users' highlighting patterns are shaped by the behaviour of other readers and reflect a shared understanding of an audience beyond the initial highlighter.

Keywords

Amazon.com, Commonplace book, Digital reading, Ebooks, Highlighting, History of reading, Kindle, Kindle Public Highlights, Public domain books, Social reading

1 Introduction

The rise of digital media has led to the re-examination of historical literacy practices including the act of commonplacing.[1] The tradition of commonplacing refers to the historical practice of selecting important parts of a text to remember and reflect upon for their significance beyond their original textual context (see Blair, 1992, Lesser and Stallybrass, 2008; Throsby, 2012). In this article, I use keyword analysis of the Kindle Popular Highlights website (Amazon Inc., 2014) together with close analysis of variations between similar highlights to investigate how readers of public domain ebooks use shared reading functions to carry out a contemporary form of commonplacing. This analysis suggests that Kindle users' shared highlights that focus on shared wisdom or pivotal narrative moments are guided by the affordances of the platform and an awareness of an audience much larger than the individual reader.

2 Background: Kindle Popular Highlights and social reading

2.1 Reading together online

Digital social reading has been framed as a break from traditional forms of reading, and previous research has analysed several important sites for interaction such as the GoodReads website (Nakamura, 2013). One of the key sites for such interaction is the website maintained by leading online retailer, Amazon (see Allington, 2016: Section 1 for further discussion). The Amazon website enables readers to post reviews of books, and these have been the subject of a number of academic studies focusing on contrasts between 'popular' and 'highbrow' forms of literary culture (Gutjahr, 2002; Steiner 2008; Allington, 2016). As well as selling physical books, Amazon is the world's

leading ebook retailer, thanks to its Kindle range of tablet computers and dedicated e-readers, which are designed in order to direct sales traffic towards Amazon's online store. Like the Amazon website, the Kindle provides a social infrastructure facilitating interactions between Amazon customers. Cameron (2012: 86) posits a link between the Kindle's shared highlight feature, which offers the reader the chance to share a section of an ebook with other readers of the same title, and the aide-memoire of marginalia in print culture. Barnett (2014) argues that shared highlights extend beyond an aide memoire for a single reader, as Kindle Popular Highlights 'suggest[s] the promise of facilitating deeper reading conducted across social networks, of motivation for reading for reluctant readers, for productive readings in educational settings, and for guided readings through the involvement of teachers of literature and authors themselves'. While this seems optimistic, Cameron's (2012: 89) study of highlighting in a single electronic edition of *The Adventures of Sherlock Holmes* suggests that a pre-existing pattern of frequently highlighted passages can drive future behaviour as the 'hive mind' converges on already popular quotations focusing on 'love, method, and culture'. While 'method' is a core theme in *Sherlock Holmes* which might not extend to other titles, 'love' and 'culture' may feature as commonplaces in the Kindle Popular Highlights. In this article, I will use the corpus linguistic approach of indentifying keyness and collocates in order to test whether Cameron's findings extend beyond *Sherlock Holmes*.

2.2   *Accessing Kindle Popular Highlights*

In order to analyse large-scale datasets, researchers require access to data that is often proprietary, and Kindle Popular Highlights represented an uncharacteristic degree of

openness from Amazon. Users were able to view what their fellow readers had highlighted on titles they had not purchased, which provided a useful marketing tool for popular titles on the Kindle. Amazon made this data available as a public resource from early 2011 to mid-2014, when it obfuscated access to the database through removing a central list of the most highlighted passages, requiring users to instead crawl through four million individual records to collate the same data.[2] While the dataset was in circulation, it provided a cache of data that offered an insight into how various parts of the Kindle's infrastructure were used. For example, while Amazon boasts a catalogue of several million ebooks, it is unclear which titles users are purchasing, let alone reading. Kindle Popular Highlights cannot map the reading habits of all users, but it reveals which ebooks a subset of Amazon customers have purchased and subsections that a proportion of those readers found interesting.

The dataset cannot be used to form hypotheses about small-scale reading practices as it only offers a macroscopic view of highlighting culture, such as the number of people who have highlighted a particular passage. There is no ability to break down the demographics of this mass readership or see how any one reader has highlighted his or her copy of the book. However, the dataset still provides a large body of evidence with regard to textual fragments that Kindle users wished to record and share. In response to these methodological limitations, I have been careful to focus primarily on the textual evidence without trying to reconstruct the intention of readers.

I will use two approaches to the Popular Highlights culture: first, a comparison of the keywords in public domain highlights against both the Kindle Popular Highlights dataset and a selection of Project Gutenberg texts to represent texts in the public

domain; and second, a close analysis of highlighting patterns in the works of the ten most popular authors and repetition across unique editions of the same titles to reveal the interests of readers. The combination of large and micro-scale analysis is instrumental in understanding new cultures of literacy. Analysis of screen reading offers methodological frameworks for understanding literacy, but these must be considered through an understanding of its historical precedents.

2.3   The historical context of commonplacing

Historical context can offer illuminating parallels between print and digital culture. The Kindle Popular Highlights infrastructure facilitates a behaviour analogous to the historical act of commonplacing, which Blair (1992: 541) describes as 'select[ing] passages of interest for the rhetorical turns of phrase, the dialectical arguments, or the factual information they contain'. These commonplaces were either noted in the margins of the book being read, or formalised into manuscript books to collate a single reader's activity.

Valenza (2009: 220) demonstrates how commonplacing helped transform narratives into wisdom, since the act of copying important passages from the text to the margins or separate notebook allows readers to focus on parts of the text that impart particular truths or values instead of teasing out the narrative as a whole. This practice was formalised with the rise of the anthology, which functioned as an extended form of commonplacing pre-approved by the publisher and 'trained readers to pace themselves through an unmanageable bulk of print by knowing when to skip and where to linger' (Price, 2000: 4). The commonplace book encourages readers to view texts as collections of aphorisms that can be manipulated in various ways through focusing on

decontextualized expressions of wisdom rather than on narrative (McGill, 2007: 357).

In this framework, the commonplace book can be seen as the database to complement

the narrative of the original text (Manovich 2001: 231). The collation of

commonplaces into a single cross-referenced repository in book form provides a clear

precursor to the popular highlights database.

Authors such as John Milton and Francis Bacon maintained commonplace books that

served as a record of their composition process. These manuscripts were later

preserved through publication to allow readers to track the development of these

famous authors' works. While this emphasises the individuality of each reader, Kindle

Popular Highlights transforms the highlighting behaviour of vast numbers of readers

into an amorphous whole, where it is difficult to distinguish the individual from the

group, creating a new type of commonplace that focuses on convergence rather than

individuality.

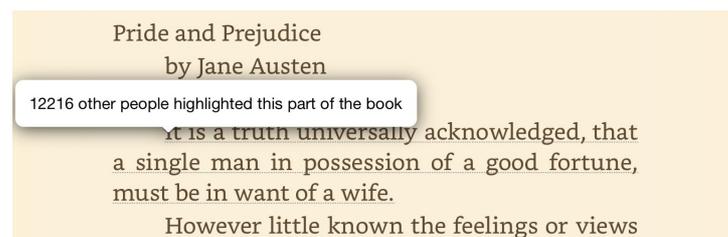## 2.4   Shared highlights on the Kindle

Figure 1. An example of a popular highlight from Jane Austen's *Pride and Prejudice* (2007). The shared highlight is lightly underlined and, when clicked, reveals a speech bubble stating the number of times the highlight has been shared

The Kindle's highlighting infrastructure is complex. Highlights can be used to

bookmark an important part of the text or can be viewed within a personalised

commonplace book that contains the accumulative highlights of the user's reading.

Readers can also choose to publicly share their highlights, which are then amalgamated

into the complete shared highlights dataset. The most popular passages have been shared over ten thousand times, but the majority have received fewer than a dozen highlights. Any highlight will be stored in Amazon's database, but will not be visible for readers of the ebook or database unless it is one of the top ten most popular highlights within the title. Each publicly shared quotation is the tip of an iceberg, as the bulk of shared highlights may be spread thinly throughout the ebook, particularly in the most popular titles where the threshold for inclusion is several thousand. Unfortunately, this means it is impossible to know the full extent of highlighting culture on the Kindle, which means that researchers must rather focus on homogenous reading practices that are publicly visible. The feature has enjoyed great popularity, as users have generated a million unique highlights from over one hundred thousand titles. Shared highlights then appear in two locations: (1) in the ebook, although users must purchase the title to see this; and (2) a publicly maintained database, which will be the focus of this paper.

The technical configuration of the Kindle's shared highlight function encourages comparison with commonplace books. The link between commonplaces and Kindle Popular Highlights is part of a wider trend of skeuomorphism, or borrowing cues from other media cultures, within the Kindle infrastructure. The highlighting function uses skeuomorphisms through reference to 'My Notebook' as the location where users can read their highlights for each individual book, and their collation into the 'My Clippings' file, itself a riff on scrapbook aesthetics and a narrative of disassembly and reassembly. The Kindle Popular Highlights infrastructure requires a combination of local and remote storage to cater for audiences inside and beyond the individual titles.

The only way to add new quotations to the system is through those ebooks. Once shared, highlights are rendered distinct from the books, and as the database does not maintain a constant shape, they can be viewed in multiple contexts. In this way, the highlight has morphed from a stable entity in a single manifestation of a book to a fluid and promiscuous data point, which is disassembled and reassembled at will. It is important to note that Kindle Popular Highlights is *opt-in* – not all users will necessarily choose to either see or interact with these highlights – and this represents a further departure from the traditional approach to highlight culture as the potential audience becomes much larger. Jackson (2002: 96–97) proposes that the creation of marginalia is a 'four-way transaction' involving 'text, reader, target audience, unknown future reader'. Kindle users who choose to share their highlights extend the 'target audience' and 'unknown future reader' into a large potential audience which may discover the shared highlight through reading the book or exploring the separate popular highlights website.

2.5   The importance of public domain ebooks

The Kindle Popular Highlights database contains a hetereogenous selection of texts, since readers are naturally interested in a broad variety of texts. In order to mitigate this disorder, I focus on a subset of ebooks that have had a particular role in the development of ebook culture: public domain texts. The economics of public domain ebooks, which are often sold cheaply or released free of charge, have ensured a robust market, although users may not have necessarily read them. Focusing on these texts creates a more homogenous and manageable dataset, but it also provides several secondary benefits. Public domain books have a pre-formed reception due to their

survival and circulation as digital editions. While the deluge of new ebooks being published can be difficult to sift through, public domain ebooks have undergone a form of consecration, which has an ancillary effect on contemporary readers' interactions with them, as users may find their reading shaped by the wealth of material based upon the book's reception prior to their interaction with the text. This can lead to a convergence of thought and reaction to texts that are familiar, which is the perfect counterpart to the hive mind culture of Kindle Popular Highlights.

Unfortunately, it is not easy to identify the limits of the public domain. The precise limits of the public domain are ambiguous, because recent copyright legislation has left the status of many texts unclear (see Spoo, 2013: Epilogue). Consequently, prospective publishers have an uneasy notion of what counts as public domain, particularly as the ebook marketplace has become global and therefore has to abide by the most stringent copyright legislation across the world. Since this project is not concerned with solving the complex issues of copyright, works that are treated as if they are public domain are included due to their consecration and integration into this aspect of ebook culture. Due to the fuzziness of the public domain in Kindle Popular Highlights and the poor quality of Amazon's metadata for public domain ebooks, the project works with a sample of 34,044 highlights – a corpus of over one million words – from public domain authors rather than the complete database.

Through this context of the Kindle's popular highlights culture, I will assess how users are appropriating this feature as an aspect of both individual and social literacy. Analysis of the public domain works in Kindle Popular Highlights reveals a correlation between shared highlights culture and commonplacing practices. Linguistic analysis of

the dataset can facilitate a deeper understanding of digital social reading practices through historical analogies, and more importantly, an analysis of important topics that crop up throughout shared quotations. In particular, Kindle Popular Highlights offers a chance to analyse comparative elements of the same texts across multiple editions, which is difficult to achieve outside of the public domain due to the strictures of copyright for post-1923 publications leading to a single ebook edition for each title.

3   What do readers of public domain ebooks highlght?

Having identified public domain highlights as a potentially useful subset of Kindle Popular Highlights, I will now analyse whether they reveal any new insight into sharing practices on the Kindle. The public domain highlights reveal the diverse range of genres that interest ebook users. These extend beyond 'the classics' of literature and books that appear in print, as readers are offered a greater selection of public domain texts through the long tail (Anderson, 2009) of the large-scale digitisation projects that have shaped ebook culture in the last decade. Philosophical treatises, fiction, and the Bible are the most popular genres in the dataset. The digitisation of public domain books has led enterprising readers to dig into the 'great unread', the deluge of fiction from the Victorian era that accompanied the mechanisation of the printing press (Cohen, 2002).

Pulp fiction from older eras has found a natural home with the Kindle as a result of the economics of the public domain, where it is possible to create a free, or cheap, copy of a text which users can purchase through the Kindle Store without the requirement to exit Amazon's ecosystem. Project Gutenberg, HathiTrust, and other non-profit digitisation projects have been instrumental in shaping the titles available on the

Kindle store, as several publishing projects have emerged to create Kindle ebooks automatically from new titles available from these archives. The following analysis focuses on those texts that have been published automatically in multiple editions in order to sample how different groups of readers semi-independently choose the same elements to highlight. These repeated elements represent the commonplace, as they indicate parts of the text are deemed important by multiple groups of Kindle readers.

3.1   Locating the commonplace

It is essential to identify what might count as a commonplace in the public domain highlights. The aphorism, which is a section of text that 'must stand by itself (unconnected), be brief, and treat a moral topic' (Morson, 2003: 409), is an important literary device connected to the commonplace, as it separates knowledge from narrative. Some of the most memorable aphorisms use humour to effectively transmit knowledge. While aphorisms are one of the most common types of commonplace as the genre encourages re-use outside of the source text, they remain only one genre. Given the popularity of highlighting these bite size chunks of text, such as the opening to *Pride and Prejudice*, it is likely that readers are mindful of an audience outside of the book who might peruse the Kindle Popular Highlights list and discover the passage. This secondary readership of shared highlights might see the quotation on Twitter or the Kindle Popular Highlights website and need to understand the relevance of the quotation without the added context of the ebook. Aphorisms, by their nature, can be treated in isolation, as they make sense without any surrounding context. This makes them ideal shared highlights for readers who might only see the quotations through an ancillary website. Both the choice of highlights and the way that readers

choose to present them can offer an insight into how the Kindle has shaped reading practices in the early twenty-first century.

3.2 Identifying keywords

In order to understand how users are responding to the public domain titles through the popular highlights system, I used AntConc to identify keywords in the public domain Kindle Popular Highlights. Keyness assesses which words occur significantly more frequently in a particular corpus compared to a reference corpora (Baker 2006: 121–149). In corpus linguistics, it is conventional to treat a word with a log-likelihood of over 6.63 as a keyword for the corpus under analysis. I used keyness to identify clusters of words that readers focus on when sharing highlights from public domain texts compared to all other highlights. If commonplacing is an important part of this shared literacy, these words are likely to represent the basis of any aphorisms or discussion of universal themes outside of the narrative specificity of any individual work.

In order to reveal how users react to the public domain subset of Kindle Popular Highlights, I used two reference corpora: the complete Kindle Popular Highlights dataset, and Lahiri's (2013) sample of 3,036 public domain English books by 142 authors from the Project Gutenberg collection. The Kindle Popular Highlights dataset is useful for finding differences between the larger set of highlights and those in the public domain, while the sample of public domain texts can help reveal what features users are emphasising in texts that are no longer protected by copyright.

The public domain Kindle Popular Highlights feature many arcane and foreign words not present in either reference corpus. This is a clear consequence of comparing a multilingual public domain corpus to a selection of post-1923 works and Lahari's English language-only selection of public domain titles. After removing foreign and arcane words, as well as function words, leaving only contemporary English lexical words (Baker, 2005: 54), it is possible to focus on the most pertinent keywords for commonplace culture in the public domain Kindle highlights dataset. This focus provides evidence of semantic content of the highlights rather than reader's potential interest in syntactic features and style more generally. While the literary style may interest readers, commonplaces turn out to focus mostly on the knowledge that can be extracted from a work's semantic value. Table 1 shows the top 10 nominal and verbal keywords in the public domain Kindle Popular Highlights as compared to the two reference corpora.

All Highlights		Public Domain texts	
Word	Keyness	Word	Keyness
shall	1352.647	love	4814.379
soul	*1133.030*	human	3829.434
labour	955.408	god	3636.368
nature	*949.255*	world	3192.113
ought	694.389	*soul*	*2933.648*
heaven	640.586	happiness	2827.975
virtue	*571.624*	*nature*	*1941.157*
vanity	570.443	self	1602.828

fortune	502.481	people	1590.350
honour	399.525	*virtue*	*1560.509*

Table 1. Keywords in public domain Kindle Popular Highlights [shared words in italics].

There is evidence from a keyword comparison that readers highlight parts of public domain texts that relate to commonplaces as compared to their reaction to other ebooks and public domain books in total. For example, the keywords often relate to values (for example, 'honour', 'vanity', 'virtue' and 'fortune'), and feature two modal verbs ('ought' and 'shall') associated with moral directives, an important form of imparting wisdom. This cluster dominates the keywords on both lists, demonstrating that readers are not only focusing on the importance of values within the public domain subset of Kindle titles, but that they are also statistically more likely to highlight parts of public domain works which focus on values and themes of morality and spirtuality ('self', 'heaven', and 'soul'). Overall, the list represents themes that have concerned the authors of both literature and philosophical works regardless of historical context.

Three words – *soul*, *nature*, and *virtue* – appear with greater frequency in both comparisons. I searched for collocates of 'soul' and 'nature' to find out whether the two words are associated with values rather than narrative markers – i.e. parts of the text that denote a pivotal part of the narrative's development – in the public domain Kindle Popular Highlights. 'Virtue', by definition, refers to values rather than to narrative, as the word refers directly to morality except through metaphor. The collocates of 'nature' reveal the term refers more frequently to humans and essence than flora and fauna. Only 15% of instances are capitalised to refer to '[Mother] Nature', while 'human nature' and its variants, appears in 25% of all instances of 'nature'. In the remaining 1012 instances of 'nature' in the corpus, 57% collocate with

prepositions – including 'of' (713 instances) and 'by' (165 instances) – indicating an interest in nature as essence. Out of 1119 instances of 'soul', the word collocates with 'man' or 'his' in 114 instances, which can refer to a generic everyman as well as a specific character. There is also a smaller cluster of collocates that refer to the soul and self-improvement, including 'nourish' (10 instances), 'save' (10 instances), 'mutiny' (13 instances), 'peace' (18), and 'cure' (20). The patterns of use for the two keywords demonstrate that these words are used in a context where they can be extracted from the text for reflection. As practised through the Kindle, the contemporary act of commonplacing therefore seems to focus on the same chunks of knowledge that were targeted in its print culture counterpart.

4   Popular authors in public domain highlights

Author	Highlights
William Shakespeare	1671
Charles Dickens	1031
Mark Twain	932
Friedrich Nietzsche	860
Oscar Wilde	727
Jane Austen	708
Leo Tolstoy	708
Arthur Conan Doyle	649
Plato	617
Fyodor Dostoevsky	613

Table 2. The top ten Kindle Popular Highlights authors.

Due to the prevalence of commonplace culture in Kindle Popular Highlights, it is no surprise that authors who craft aphorisms and ironic inversions that can readily be removed from their surrounding context dominate the most popular quotations list. Table 2 shows the top ten public domain Kindle Popular Highlights authors, who are also likely to be the most popular authors throughout the whole database, since no licence is required to publish their works leading to multiple editions. I used a word frequency count of the top ten authors (table 3) to investigate whether these authors are favoured for their literary and philosophical qualities.

Word	Frequency
man	1608
love	885
life	822
good	674
people	529
world	488
great	435
time	414
die	371
god	329

Table 3. The top 10 most frequently occurring words in the top 10 authors.

Table 3 shows that values and philosophical topics once more emerge as common themes. While these words have been drawn from texts initially published prior to the twentieth century, it is clear that the concepts remain relevant to contemporary society. These words coalesce into themes that remain a concern today, such as love, religion,

mortality and humankind's integration into the world. These commonplace themes confirm Cameron's assertion that 'love' is an important part of highlighting culture, but her other two categories of 'method' and 'culture' are replaced by a broader philosophical approach to the texts. Shared highlights are used as a way to glean wisdom through revisiting classic texts. However, this preference of reading public domain texts for their universal truths is not the only way readers are highlighting the text, as many are still interested in the plot.

4.1 Highlighting pivotal moments in the narrative

Beyond the commonplace, there is a second genre of annotation that is pervasive in shared highlights culture: pivotal narrative moments. In this genre, readers use the shared highlights function to note the affective climaxes of a narrative rather than any perceived wisdom in the highlighted passage. *Pride and Prejudice* provides one of the clearest examples: 'In vain I have struggled. It will not do. My feelings will not be repressed. You must allow me to tell you how ardently I admire and love you' (Austen, 2007: loc. 2205). While this is not aphoristic and needs further context to be understood, it provides other readers with a teaser which reveals key narrative information without being a spoiler, as there is a notable absence of naming actors. This form of *in medias res* highlighting, or the selection of text that takes place in the middle of the action, is replete with unidentifiable actors and decontextualised dialogue. As *in medias res* highlights take place in the middle of the action, they are identifiable through typographic patterns, such as a highlight which starts in the middle of a sentence, or quotation. These patterns demonstrate that users are slightly more likely to highlight *in medias res* segments of public domain titles than those still in

copyright. For example, highlighted passages in the public domain titles are marginally more likely to start with a lowercase letter (see table 4). Likewise, quotations, both complete and incomplete, are more likely to appear in public domain works, particularly at the start of a highlight. In other words, *in medias res* highlighting is as important a part of the kind of social reading studied here as commonplacing and shows the range of literacy practices at work on the Kindle.

Typographic feature	Public domain highlights	Other highlights
Highlight starts with a lowercase character	14.96%	13.75%
Highlight contains a closed set of quotation marks	8.73%	5.93%
Highlight starts with a quotation mark	5.26%	2.16%
Highlight contains at least one incomplete set of quotation marks	3.96%	1.17%

Table 4. Typographic features indicative of *in medias res* highlighting

The early technical infrastructure of Kindle Popular Highlights encouraged *in medias res* highlighting involving the wider context, as the webpage for each title contained each quotation in its context. Figure 2 shows how the early interface of the Popular Highlights website encouraged *in medias res* highlights through displaying the shared

highlight in its immediate context from the original ebook. *In medias res* narrative highlights are more common in fiction and poetry than non-fiction. Many of these highlights involve speech as dialogue is a common method of delivering exposition, and five percent of all Kindle Popular Highlights contain at least one quotation mark.[3] Figure 2 demonstrates that commonplacing and *in medias res* highlights are not mutually exclusive as aphorisms can be integrated into narrative developments.

" Vanity and pride are different things, though the words are often used synonymously. A person may be proud without being vain. **Pride relates more to our opinion of ourselves, vanity to what we would have others think of us."** "If I were as rich as Mr. Darcy," cried a young Lucas, who came with his sisters, "I should not care how proud I was. I would keep a pack of foxhounds...

Highlighted by 405 Kindle users

Figure 2. *In medias res* highlighting in the early Kindle Popular Highlights.

4.2   Analysis of patterns in Jane Austen's Pride and Prejudice

The clearest evidence for commonplace culture appears in titles with multiple editions in Kindle Popular Highlights. Austen is one of the most frequently highlighted authors in the dataset and the opening of *Pride and Prejudice* – 'It is a truth universally acknowledged, that a single man in possession of a good fortune, must be in want of a wife' (2007: loc. 1) – is the second most shared quotation. This commonplace has remained popular since the novel's publication as it parodies a certain understanding of class, upbringing and self-interest: themes that remain as pertinent in 2015 as in 1813. In 2012, 9,260 users had highlighted the opening sentence of *Pride and Prejudice*, while by October 2015, this had risen to 12,216 (despite the edition's unavailability for download since 2012). This represents a fifth of all highlights in the novel and presents a useful lower boundary for the readership of the novel: at least twelve thousand users, and probably far more, have downloaded this edition of the novel. Austen's readers require an understanding that they are dealing with generations of readership and

interpretation and still fall into the common trends predicted by Austen: as Halsey writes with regard to an earlier stage in history, '[t]hose taking part in discussions about Austen are, consciously or unconsciously, rehearsing debates already implicit (or, occasionally, explicit) in the novels, and often enact in their conversations with and about Austen the concerns over reading, writing and femininity that her works address' (2013: 212). Due to the burden of prior reception apparent for any reader of Austen, the novel provides an ideal case study to explore patterns of convergence and repetition within social highlighting culture.[4]

There are forty-one editions of *Pride and Prejudice* in the dataset, ignoring the three instances of Seth Grahame-Smith's infamous remix, *Pride and Prejudice and Zombies*. Outside of the most popular edition, which has received 40,000 highlights, the remaining books have only received 750 highlights altogether. The interest of these other titles instead comes from looking at patterns and repetition. 'It is a truth universally acknowledged…' appears as the most popular highlight for sixteen of the titles. Since various editions often share similar highlight locations, it is possible to search for patterns and differences in highlighting practice. If Kindle users were guided only by the public highlights of other Kindle users, each edition would have a unique set of most popular highlights, but this is not the case, as some highlighted passages seem almost universal across versions of a text.

In this framework, it is possible to look at patterns and the development of repetitive commonplaces. 'It is a truth universally acknowledged…' remains intact, other than on two occasions, when the latter half of the aphorism – 'single man in possession of a good fortune, must be in want of a wife.' – is all that remains. To some extent, this

may be a technological problem – the user may not necessarily know exactly how to highlight the section they want – but equally, without a chain of influence of other readers, users converge towards the heart of their interests. Figure 3 shows a more complex example of variable highlighting, as users have altered the quotation's length. The final sentence receives the most overall highlights (8,437), while the others are all less than 100, other than *Pride and Prejudice and Zombies*, which features the exact same quotation without zombies. In total, a quarter of highlights relating to *Pride and Prejudice and Zombies* ignore the zombies in favour of the classic aphorisms from Austen's original, demonstrating how readers converge upon these statements. The clearest evidence of commonplacing is subtle and appears in only a single edition: the full sentence 'A person may be proud without being vain' is not direct enough, so the first three words are removed to transform the aphorism into an unmodalised imperative ('be proud without being vain') to share with fellow readers. A similar form of clipping a single sentence to focus on a specific sentiment occurs four thousand times in the corpus reflecting a level of transformation required to commonplace through the removal of contextual markers to instead focus on the universality of a particular quotation.[5]
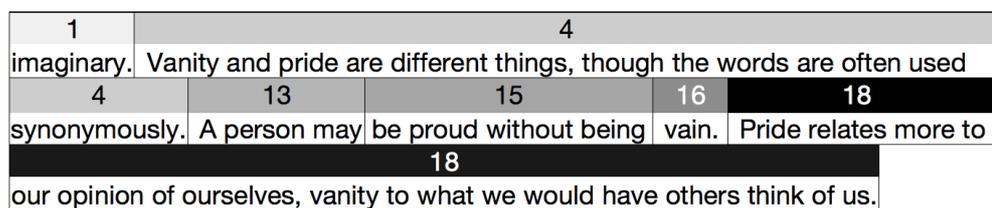
1	4			
imaginary.	Vanity and pride are different things, though the words are often used			

4	13	15	16	18
synonymously.	A person may	be proud without being	vain.	Pride relates more to

18
our opinion of ourselves, vanity to what we would have others think of us.

Figure 3. Heat map of a single quotation's appearance in unique editions of *Pride and Prejudice* [produced by author through examining variants of quotation].

4.3   Highlighting patterns of Hamlet's soliloquy

Hamlet's most famous soliloquy is the third most highlighted passage in the dataset after the preceding two examples from *Pride and Prejudice* and provides a rich example of variation in highlighting across editions, as the potential source material is much longer, as the monologue continues for 35 lines. There are no highlights of the whole speech, since Kindle users can only highlight a single screen of text, and Hamlet's speech extends beyond that for most typographic configurations of the Kindle. Likewise, it is rare for just the famous first couplet ('To be or not to be, that is the question') to be highlighted, as less than half just highlight the introduction, and once with the character's name too. The range of concluding points for the highlights between the first and last line are of greatest interest, as two of the highlights stop at the third line, 'Whether 'tis nobler in the mind to suffer', six continue further to 'Or to take arms against a sea of troubles, And by opposing them?' and two conclude with 'To die, to sleep'.

These highlighting practices occur across at least two editions, which demonstrates a convergence towards particular elements of the soliloquy. The shared highlighting patterns across editions may represent a shared aesthetic appreciation of elements of the soliloquy, although this is difficult to deduce without further knowledge of individual highlighters. At least three readers must highlight exactly the same passage without knowledge of the others' highlights for it to be classified as a popular highlight, leading to a degree of unseen convergence. These readers' common understanding of the importance of elements of the speech extends its historical role as

21

a commonplace, as readers choose to separate elements of the speech into chunks of knowledge rather than focus on its significance as a complete passage.

5 Conclusion

In this article, I have argued that the shared highlights of public domain books on the Kindle reveal that readers are re-appropriating the historical literacy practice of commonplacing. These public domain ebooks represent some of the most frequently shared titles as readers are mining the classics for wisdom on mortality, religion and humanity's place within the universe. Highlighting important narrative moments (a feature that is more popular in titles published post-1923) is combined with an interest in extracting universals and expressions of values from the text. Further to this, my analysis showed that readers undertake an act of curation when forming their highlights, as they may share a quotation that starts in the middle of a sentence to turn the highlight into an imperative. Examples from *Pride and Prejudice* and *Hamlet* demonstrate that readers edit and curate the content of their highlights to emphasise themes and messages that may interest individual readers. Since each popular highlight initially requires three users to independently select exactly the same passage, and this can happen across multiple editions of the text, users may be focusing on specific aspects of these public domain works. Once the highlight is shared more than three times, users are more likely to follow pre-existing trends that modify the pre-existing commonplace.

Through focusing on just the public domain titles in the Kindle Popular Highlights database, I have demonstrated nuances in the literacy practices of ebook readers that cannot simply be explained as the result of aide memoire or of the re-sharing of already

22

popular highlights. The advent of large-scale datasets of contemporary reading habits presents exciting new frontiers – and pitfalls (Davis, 2015) – in the quantitative and qualitative study of reading. This article presents one specific framework for such research. In so doing, it begins to show how readers make use of the affordances provided by commercial reading technologies: in this case, by hunting for wisdom in pre-1923 texts.

Notes

[1] See Barton (1994) for a discussion of literacy practices.

[2] No official reason was given for restricting access to the database, but it was likely due to the value of the data in an age where publishers are beginning to become more interested in the reading habits of their customers (Kobo 2014). To share data freely would be to give up the possibility of selling it.

[3] Since readers will not necessarily highlight the beginning or end of a quotation, this form of search is only a lower limit of the presence of quotations in the corpus.

[4] See Thomas and Round (2016) for more on digital public reading of Austen.

[5] The clippings were identified through running a 6-gram analysis of the top 10 most popular authors to find the most commonly recurring phrases, followed by tracking the 6-grams variation within the concordance.

References

Allington D (2016) 'Power to the reader' or 'degradation of literary taste'? Professional critics and Amazon customers as reviewers of *The Inheritance of*

Loss. *Language and Literature* 25 (3)

Amazon Inc. (2014) Kindle popular highlights. Available at: http://web.archive.org/web/20141020062409/https://kindle.amazon.com/most_popular (accessed 9 June 2015).

Anderson C (2009) *The Longer Long Tail*. London: Random House.

Austen J (2007) *Pride and Prejudice [Kindle Edition]*. [s.l]: Public Domain Books.

Baker P (2006) *Using Corpora in Discourse Analysis*. London: Continuum.

Barnett T (2014) Social reading: the kindle's social highlighting function and emerging reading practices. *Australian Humanities Review*. 56, Available at: http://www.australianhumanitiesreview.org/archive/Issue-May-2014/barnett.html (accessed 9 June 2015).

Baron N (2015) *Words Onscreen: The Fate of Reading in a Digital World*. Oxford: Oxford University Press.

Barton D (1994) *Literacy: An Introduction to the Ecology of Written Language*. Oxford: Blackwell

Blair A (1992) Humanist methods in natural philosophy: the commonplace book. *Journal of the History of Ideas*. 53(4): 541–551.

Cameron L (2012) Marginalia and community in the age of the kindle: popular highlights in the Adventures of Sherlock Holmes. *Victorian Review*. 38(2), 81–99.

Cohen M (2002) *The Sentimental Education of the Novel*. Princeton: Princeton University Press.

Davis M (2015) E-books in the global information economy. *European Journal of Cultural Studies*. 18(4–5), 514–529.

Gutjahr PC (2002) No longer left behind: Amazon.com, reader response, and the changing fortunes of the Christian novel in America. *Book History*. 5(1), 209–236.

Halsey K (2013) *Jane Austen and her Readers, 1786–1945*. London: Anthem.

Jackson HJ (2002) *Marginalia: Readers Writing in Books*. New Haven and London: Yale University Press.

Kobo (2014) Publishing in the era of big data. Available at: http://cafe.kobo.com/_ir/159/20149/Publishing%20in%20the%20Era%20of%20Big%20Data%20-%20Kobo%20Whitepaper%20Fall%202014.pdf. (accessed 28 September 2015).

Lahiri S (2013) Complexity of word collocation networks: a preliminary structural analysis. *arXiv:1310.5111 [physics]*, Available at: http://arxiv.org/abs/1310.5111 (accessed 9 June 2015).

Lesser Z and Stallybrass P (2008) The first literary *Hamlet* and the commonplacing of professional plays. *Shakespeare Quarterly*. 59(4), 371–420.

Manovich L (2001) *The Language of New Media*. Cambridge: MIT Press.

McGill ML (2007) Common places: poetry, illocality, and temporal dislocation in Thoreau's 'A week on the Concord and Merrimack rivers'. *American Literary History*. 19(2), 357–374.

Morson GS (2003) The aphorism: fragments from the breakdown of reason. *New Literary History*. 34(3), 409–429.

Nakamura L (2013) 'Words with friends': socially networked reading on goodreads. *PMLA*. 128(1), 238–243.

Price L (2000) *The Anthology and the Rise of the Novel: From Richardson to George Eliot*. Cambridge: Cambridge University Press.

Spoo R (2013) *Without Copyrights: Piracy, Publishing and the Public Domain*. Oxford: Oxford University Press.

Steiner A (2008) Private criticism in the public space: personal writing on literature in readers' reviews on *Amazon*. *Particip@tions*, 5(2) Available at: http://www.participations.org/Volume%205/Issue%202/5_02_steiner.htm (accessed 16 November 2015)

Thomas B and Round J (2016) Moderating readers and reading online. Language and Literature 25 (3)

Throsby C (2012) Byron, commonplacing and early fan culture. In: Mole T (ed.), *Romanticism and Celebrity Culture, 1750–1850*, Cambridge: Cambridge University Press, pp. 227–244.

Valenza R (2009) How literature becomes knowledge: a case study. *ELH* 76(1), 215–245.