

Introduction

1. Historical background and context
2. The Corpus
3. Definitions
4. Analysis
 - a. Word Cloud
 - b. Line charts
 - c. Bubbleline charts
5. Conclusion
6. References

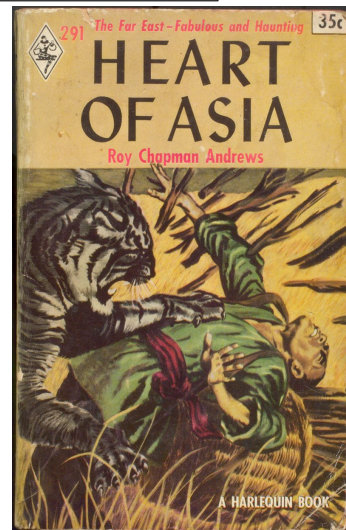


Figure 1: *Heart of Asia* by Roy Chapman Andrews.

The use of computational analysis to visualize and aid in the interpretation of literature and book history is a way of supplementing traditional literary criticism. The application of statistical algorithms on a large scale such as this allows a researcher to derive textual patterns and relationships that could not be determined without the aid of technology. Instead of only examining a small number of summaries this study will utilize a large corpus published by Harlequin Enterprises during its first ten years of operation, from 1949 to 1959. By examining stylistic changes in publishing language over time this project will be firmly situated as an effective example of a computational, cross-disciplinary approach to the research and analysis of Canadian book history.

Historical background and context

● Post-World War II Era

- More advertisement and promotion from booksellers (West, 2011)
- New retail avenues and new audiences (West, 1990; Kaestle, 2014)
- Modern printing technology and cheap materials (West, 1990; Radway, 1991)

● Harlequin Enterprises

- Founders: Doug Weld, Jack Palmer, Richard Bonnycastle
- Created in Winnipeg in 1949
- Initially a side-business publishing reprints for Canadians (Grescoe, 1996; Radway, 1991)
- Mills & Boon partnership led to the increased publishing of romances (Grescoe, 1996)

The rise of the paperback novel was an indicator of a change in the publication, marketing and selling of books. A widespread increase in literacy and improvements in printing technology from the Industrial Revolution onwards helped push North American booksellers towards advertisement and other methods of promotion. Books became available through new retail avenues - mail-order catalogues, book of the month clubs, department stores, grocery stores - to a new, diverse customer base. This reach into a wider audience, along with the use of modern printing technology and cheap materials becoming more common allowed for the growth of new paperback companies such as Harlequin Enterprises.

Harlequin's creation involved three major players: Doug Weld, Jack Palmer, and Richard Bonnycastle. Bonnycastle, a Canadian socialite and previously an employee of the Hudson's Bay Company joined up with Weld to assist in re-energizing a branch of Weld's family printing firm, which specialized in paperback novels. During this time Bonnycastle met Jack Palmer, a sales manager for a distributor of publisher Grosset & Dunlap. The three founded Harlequin Enterprises with the intention of having a small publishing company that specialized in mass-market paperbacks and catered to Canadians. They printed approximately two dozen novels in their first year, but returns were common, and the company struggled for several years.

After the death of Jack Palmer, Richard Bonnycastle's wife Mary and his personal secretary Ruth Palmour began to take major roles in the company. Ruth took on a managerial role and contacted publishers to find titles to reprint, while Mary held final approval and editing responsibilities. The two women were shrewd enough to recognize Harlequin's early successes took the form of romance novels, and were the two key players in creating a partnership with Mills & Boon, an English publishing company that sold pocket-books and romances. The Mills & Boon romances were a success with very little returns, and "from 1955 on, Harlequin published one romance title almost every month, nearly all of them with nurse-and-doctor themes".

The Corpus

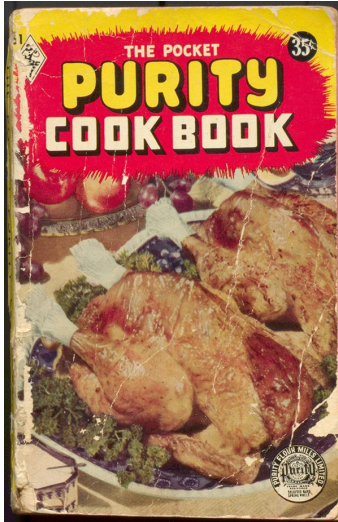


Figure 2. *The Pocket Purity Cookbook* by unknown front cover.

- 249 books of 500 published between 1949 and 1959
- Different genres, authors, nonfiction and fiction
- Wrappers only: front covers and back covers
- Front cover quotation and back cover summary

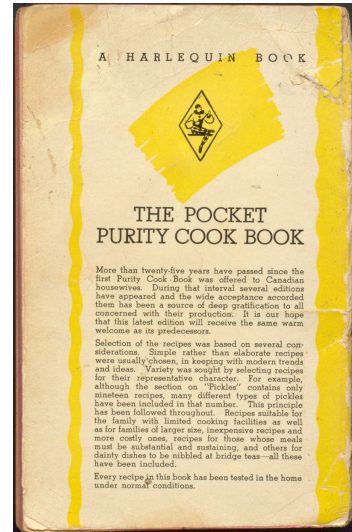


Figure 3. *The Pocket Purity Cookbook* by unknown back cover.

The corpus that is used for this analysis consists of two hundred and forty-nine novels published by Harlequin Enterprises from 1949 to 1959.

The corpus is comprised of peritextual elements rather than the interior text. Specifically, this analysis will use the front cover quotations (when available) and back cover summaries of each book in addition to publication date and author information. This is due to the fragility of the books; as they were made with poor quality paper and glue digitization of the interior pages was not possible. Publication dates are the only data to be taken from these pages, and were transcribed by hand.

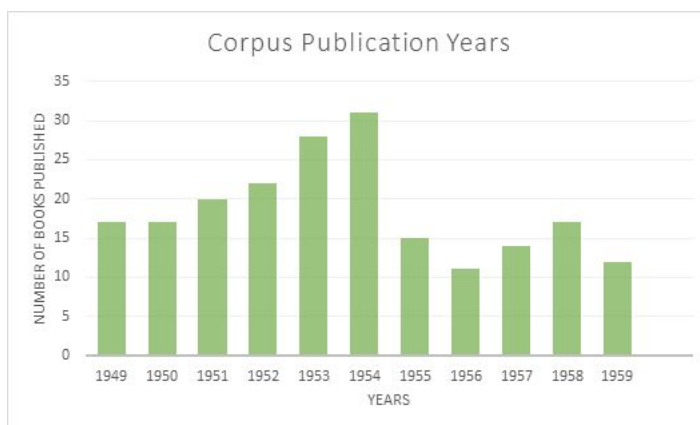


Figure 4. Corpus Publication Years graph. Shows the variation in number of books published within the date range of the corpus.

They are not spread evenly across publication years. 1954 has the largest presence within the corpus, having 31 books, while 1959 is represented with 11 books. The rest of the years within this date range vary from having 12 to 28 books published. The reason for this inconsistency is a matter of availability; the collection comes from the Bruce Peel Special Collections at the University of Alberta, and is currently in the process of being completed.

Author names have been taken from the wrappers and supplemented with further information found on the internet. Full names, pseudonyms and author gender were determined using a mixture of websites and databases. In five cases determining whether the author was male or female was not possible, and in 31 cases it was not clear whether the author was using a pseudonym. In total 134 author names are given, with 34 female and 95 male. Only 5 male authors and 8 female authors are confirmed to use pseudonyms.

Definitions

- Paratext: "A threshold... that offers the world at large the possibility of either stepping inside or turning back" (Genette, 1997, p. 2).
- Peritext: A subset of paratext pertaining to its location of being around or within a book in relation to the contents (Genette, p. 4-5).
- Voyant: Free text analysis tool that offers many ways to analyse and visualize a large body of text.

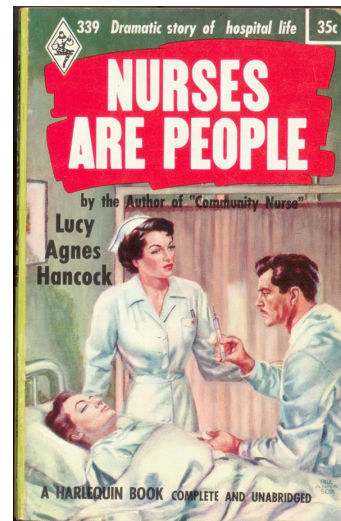


Figure 5. *Nurses are People* by Lucy Agnes Hancock.

G rard Genette first categorizes paratext in *Paratext: Thresholds of Interpretation* as being the text surrounding the contents of a book, defining this information as "a threshold... a 'vestibule' that offers the world at large the possibility of either stepping inside or turning back". This text conveys messages from the author or publisher to the reader, providing them with certain expectations about the book's contents.

Due to their placement in relation to the contents (that is, around or within the book) they are considered the subset of paratext, peritext. In the case of this study, the front cover quotation and the back cover summary are defined as peritext.

Voyant was chosen due to its user-friendly interface, its flexibility in analyzing the text and because it offers a variety of visualization outputs for interpretation. Most importantly, it contains several powerful features that are used in the processing of the text. It contains a list of common function words that are then auto-detected and removed to prevent their skewing of the text analysis. Finally, it allows for visualizations based on both relative and raw frequencies.

Voyant's relative frequency options accounts for uneven lengths of text, as the corpus has been split two different ways: into smaller documents based on

publication years, and male and female authors.



Harlequin did not begin publishing the romances for which they are now known immediately, but there is a marked progression in that direction within these first ten years. The word cloud shows that overall *doctor*, *hospital*, *dr*, and *nurse* are some of the most frequently occurring words in the entire corpus. Presumably *death*, *blood* and several other words may also be attributed to this genre, though due to its relation to others such as detective novels and thrillers the focus will remain on the more obvious terms.

The most frequent words used indicate the offering of character-focused, male-centric stories: the most frequent words are *life*, *story*, *man*, *love* and *men*. Character descriptors (*young*, *nurse*, *doctor*, *beautiful*, *girl*, *woman* and *women*) are also common, reinforcing the idea of Harlequin's early publications focusing on character-driven stories. The most common phrases in the corpus support this, consisting of "story of the..." occurring 16 times, and "is a tale of..." occurring 4 times. Several locations are also among the most frequent words, such as *London*, *town*, *hospital*, *island*, *England*, *country* and *valley*.

Thus, the Harlequin collection most frequently tried to appeal to its customer base through the prominence of specific characters, giving some detail to them through traits such as occupation and their placement within specific locales. Medical-related terms are of note due to their high frequency and the important

historical context in which the medical romance genre is placed.

Analysis: Line chart showing words to describe medical romance.

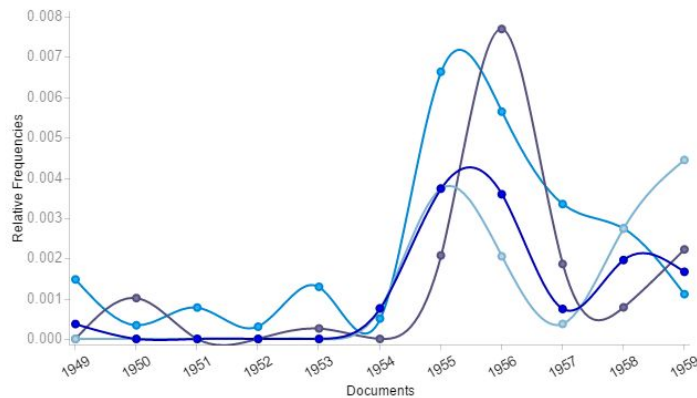


Figure 7. Medical romance term line chart.

Terms	Colours
Doctor	●
Dr	●
Nurse	●
Hospital	●

The line chart adds a chronological element showing an increase in frequency from 1954 onwards. While there is a dip in 1956 and 1957, all terms aside from *doctor* increase, and comparing 1949 to 1959 shows a general increase in the use of every term. Harlequin's history supports this development of the terms with the partnership with Mills & Boon to provide more medical romances to a Canadian audience in the mid-to-late 1950s.

Analysis: Line chart showing words to describe women.

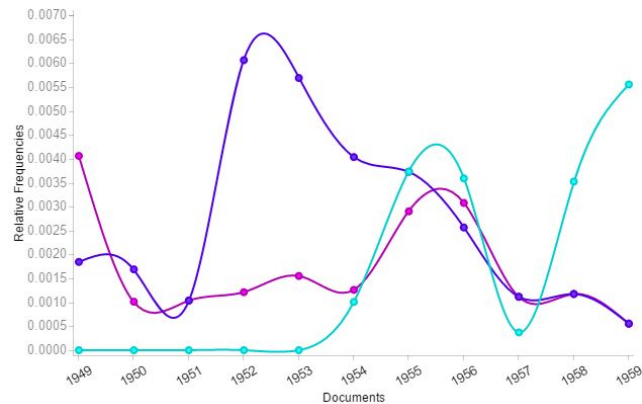


Figure 8. Women descriptor line chart.

Terms	Colours
Girl	●
Woman	●
Nurse	●

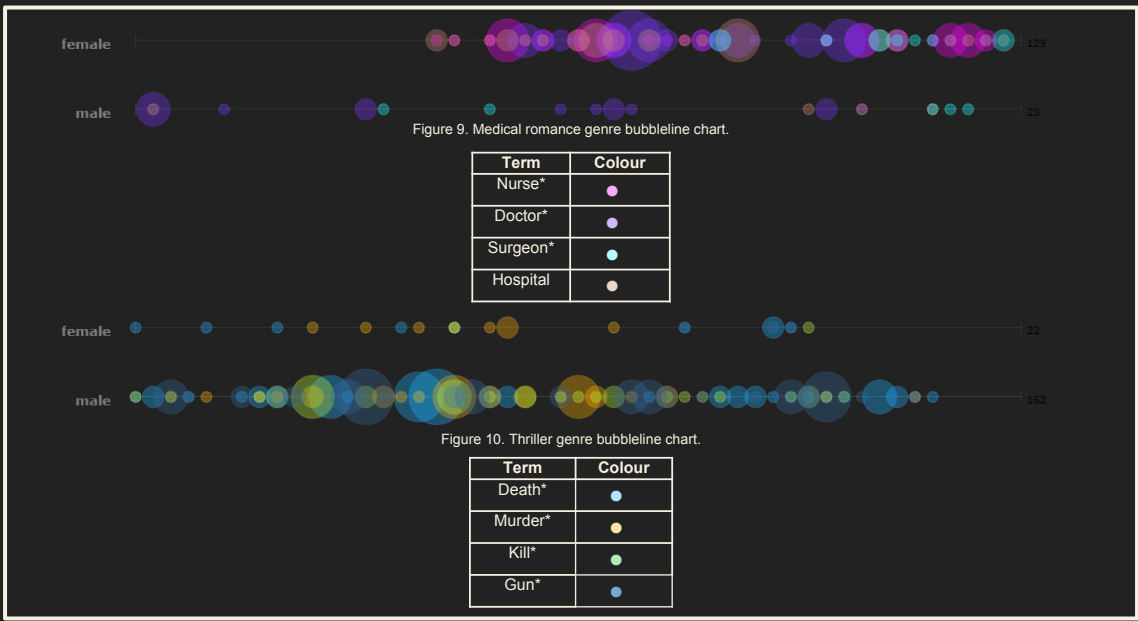
Further interpretations can be made from this data, such as the shift in audience. If Harlequin expected these romances would appeal to women, the argument can be made that the company's target audience began to shift from men to women. Thus, there should be a corresponding shift in language in general in the mid-to-late 1950s. Harlequin would have to refer to female characters in a manner that would both appeal to women and allow them to identify with the characters (namely, the heroine) to successfully market to them, so taking a page from the medical romance genre this analysis begins with *nurse*.

Two other generic descriptors (and two of the most frequent words in the corpus) for female are added: *girl* and *woman*. Over those ten years a pattern can be seen: *girl* begins as the most common descriptor for female characters, but is replaced by *woman* in the mid-1950s, which is in turned replaced by *nurse* in the mid-to-late 1950s. To compare, a similar analysis was made using terms for men found in Appendix A, with no visible pattern arising. For Harlequin's female characters (quite possibly though not necessarily the heroines) they are at first described diminutively, then generically, and finally by a common occupation for women during that era.

Their role in this peritext becomes more important over time; *girl* is a term that in this context denotes either feelings of innocence and helplessness or

sexuality, standing in stark contrast to *nurse*. Accompanying adjectives support this notion: "... the gentle Quaker girl from Martha's Vineyard..." (*The Manatee*), "... the story of a sensitive young girl whose shyness and insecurity cause her to seek escape in a career as a dancer" (*Flame Vine*), and "a fiery passionate Spanish girl, who knows all the tricks and the answers too" (*Honeymoon Mountain*).

Analysis: Bubbleline charts comparing medical and thriller genre terms between male and female authors.



Further, there is a clear distinction between female authors writing this genre, and male authors gravitating towards more violent genres. Using a bubbleline chart, the terms *nurse**, *doctor**, *surgeon** (using an asterisk to include any plural forms) and *hospital* all become increasingly common from female authors over time, while use of these words by male authors only occur in a few, small, discrete groupings. The exact opposite of this pattern can be seen when entering several terms commonly used in thriller, detective or western genres.

For this second bubbleline chart, *death**, *murder**, *kill** and *gun** were all chosen as they all represent violent aspects of a novel that can be applicable to all three genres. Asterisks were included to increase the range of terms to others that would fall within one or two of these genres, but not all three. For example, *gun** could include *gunslinger*, which is found in several western back summaries but would be an unlikely occurrence in a detective novel. Due to the era in which the collection was published it was unlikely back summaries that portrayed violence would have been targeted towards women, and just as with the medical romance terms there is a clear distinction in which authors are writing more violent novels.

Men vastly outnumber women in this instance, with these terms only appearing in a few places on the female author chart. Additionally, unlike the

medical romance bubbleline chart there is no increase in the use of these terms over time but a slight decrease. It is not as prominent as the medical romance increase though it is still visible, which can be attributed to Harlequin's partnership with Mills & Boon and their change in direction towards more romantic-themed novels. These charts also indicate that along with the change in genres Harlequin began to publish more female authors, though this is likely due to female authors being more common in romance genres.

Conclusions

- Marketing language patterns can be seen in the front cover quotations and back cover summaries
 - Describe the stories in straightforward language
 - Language used to describe female characters has changed over time to draw in women readers
- Men and women authors use differing language
 - This has decreased over time
- Harlequin has targeted women readers more and more over time

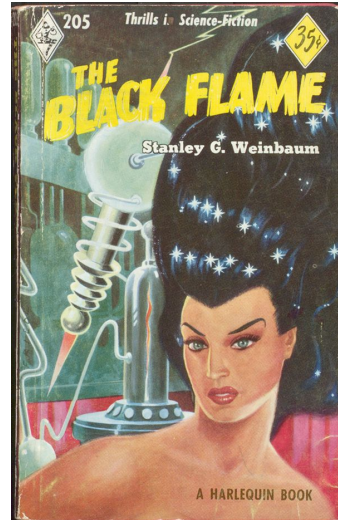


Figure 11. *The Black Flame* by Stanley G. Weinbaum.

The analyses presented in this study have allowed for the interpretation of the marketing language used to appeal to readers, in an overarching sense and specifically related to arguably Harlequin's most important genre, the medical romances. The back summaries have a strong focus on describing the stories the books offered to readers in a plain, straightforward language.

The language used to describe female characters, meanwhile, changed over time as the medical romances became more prominent and Harlequin sought to draw in more women readers. There was also a sharp delineation between the medical romances and female authors, and more violent language used by male authors, the former of which increased over time while the latter decreased. Taken together, these analyses can be interpreted as the company targeting a female audience to a greater degree over those ten years.

Future considerations

- Redoing the analysis when all 500 books are available
- Further examination of individual genres
- Comparisons between Harlequin and other reprint companies

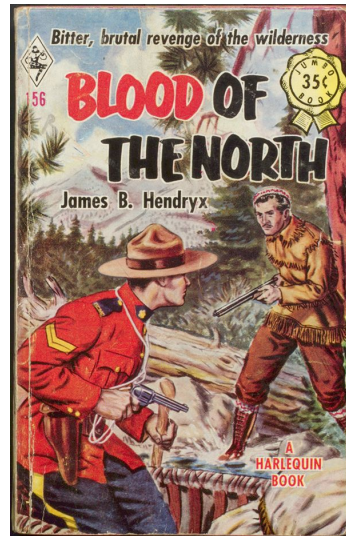


Figure 12. *Blood of the North* by James B. Hendryx.

Moving forward, there are many facets of the collection that can still be explored. Acquiring the full 500 novels to complete the collection would allow for a full digitization of the front cover quotations and back cover summaries, which would allow for these analyses to be repeated and their patterns to be confirmed. Further analysis can be taken in examining more individual genres, and questions regarding patterns and differences between these genres, and shifts in popularity could be explored. Incorporating paratext from publications outside of Harlequin could expand upon changing genre and marketing methods used during that era for paperback novels in a more general sense. Comparisons could be made between Harlequin and other reprint companies – do they follow a model like Harlequin, or is the Canadian company unique?

These are only a few ideas in which this topic can be further explored, lending support to using an interdisciplinary approach to interpret book history in a new light.

References

- Bruff, N. (1949). *The Manatee*. Winnipeg, Manitoba: Harlequin Enterprises.
- Cameron, D. (1952). *Dig Another Grave*. Winnipeg, Manitoba: Harlequin Enterprises.
- Fantastic Fiction*. (2016). Retrieved from <https://www.fantasticfiction.com/>
- Genette, G., Lewin, J., & Macksey, R. *Paratexts: Thresholds of Interpretation*. Cambridge, England: Cambridge University Press
- Grescoe, P. (1996). *The Merchants of Venus: Inside Harlequin and the Empire of Romance*. Vancouver: Raincoast Books.
- Kaestle, C.F. (2014). Seeing the Sites: Readers, Publishers and Local Print Cultures in 1880. In C.F. Kaestle & J.A. Radway (Eds.) *A History of the Book in America: Volume 4: Print in Motion: The Expansion of Publishing and Reading in the United States, 1880-1940*.
- Miller, H.T. (1949). *Flame Vine*. Winnipeg, Manitoba: Harlequin Enterprises.
- Radway, J.A. (2009). *Reading the Romance: Women, Patriarchy, and Popular Literature*. Chapel Hill: The University of North Carolina Press.
- Rockwell, G., & Sinclair, S. (2016a). *Creating a Corpus*. Retrieved from <http://voyant-tools.org/docs/#!/guide/corpuscreator>
- Rockwell, G., & Sinclair, S. (2016b). *Stopwords*. Retrieved from <https://voyant-tools.org/docs/#!/guide/stopwords>
- Rockwell, G., & Sinclair, S. (2016c). *Type Frequencies Chart*. Retrieved from <http://docs.voyant-tools.org/tools/typefrequencieschart/>
- Wees, F.S. (1949). *Honeymoon Mountain*. Winnipeg, Manitoba: Harlequin Enterprises.
- West, J. L. W. (1990). *American authors and the literary marketplace since 1900*. Philadelphia: University of Pennsylvania Press.
- West, J.L. (2011). Twentieth-century publishing and the rise of the paperback. In L. Cassuto, C.V. Eby, & B. Reiss (Eds.), *The Cambridge History of the American Novel*. <https://doi.org/10.1017/CHOL9780521899079.052>
- Title images (in order of appearance):
- Stark, Michael. (1951). *Run for your Life*. Winnipeg, Manitoba: Harlequin Enterprises.
- Martin, Chuck. (1957). *Circle F Cowboy*. Winnipeg, Manitoba: Harlequin Enterprises.
- Randall, Rona. (1953). *Girls in White*. Winnipeg, Manitoba: Harlequin Enterprises.