

# Marked E-Books and Kindle's popular highlight culture

Simon Rowberry

[s.rowberry.13@unimail.winchester.ac.uk](mailto:s.rowberry.13@unimail.winchester.ac.uk)

@sprowberry

The transition to eBooks requires new methods for tracing readership.

The study of used eBooks relies on large datasets forming around reading patterns on various dedicated and general social networks.



Figure 1. Kindle Popular Highlights

## Why Study Amazon's Kindle?

- Despite ethical and proprietary concerns, they still dominate the eBook marketplace
- The data is publicly available
- The largest dataset of highlights and annotations
- Amazon integrate services across the Kindle platform

## Method

The top 100,000 popular highlights were extracted from [https://kindle.amazon.com/popular\\_highlights/](https://kindle.amazon.com/popular_highlights/) in October 2013 to analyze in a pilot study of the Kindle's popular highlight culture.

## Overview of corpus

- **21,000** books represented
- Predominantly **Young Adult, Popular Fiction, Biography** and **Self-Help** — little poetry
- Top 10 books with most unique highlights are all non-fiction
- *The Hunger Games* series has the most total highlights

## What do people highlight?

- Primarily classics or new titles
- Calendars, word searches and dictionaries receive shared highlights



Figure 2. Atypical Highlights

- Users frequently highlight titles, captions, URLs and email addresses in eBooks
- 25% of highlights include dialogue – indicating an investment in human interaction in fiction
- Highlights tend to be aphoristic or contain advice

## Highlight Lengths

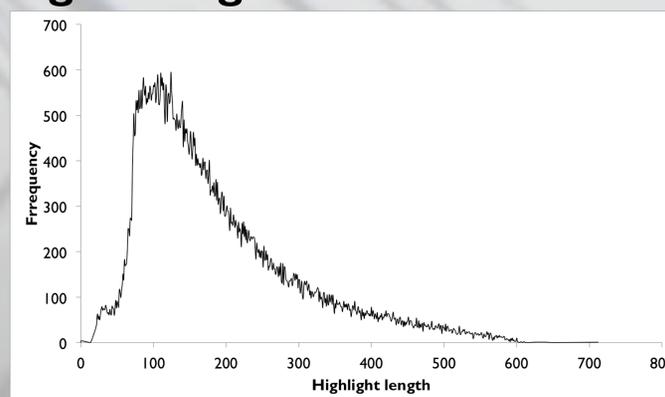


Figure 3. Frequency of highlight lengths

- Highlights are rarely shorter than a sentence or more than a paragraph
- The upper limit is a single page
- 40% of quotations are under 140 characters – users will often tweet popular highlights

## Discussion

- Highlighters represent a vocal minority of readers, not Kindle users altogether
- The highlights are heterogeneous and cannot easily be categorized into a single form of use
- Readers use the Kindle for much more than reading fiction.
- It is much easier to share an already popular highlight than trigger a new one. Equally, there is no innate reason to re-share an already popular highlight as it is shared regardless of volume once over the threshold of 3 highlights. Some activity may be accidental.
- Multiple instantiations of a single quote (e.g. "Winter is coming" in *Game of Thrones*) dilute sharing efficiency and result in exclusion from the dataset

## Limitations

- Amazon's complex infrastructure has no API
- The data is only on a large-scale level, there is no personal data for close analysis
- As sharing highlights is not the default setting, an unknown majority of the data is unavailable.

## Future Work

- Analyse entire dataset with diachronic development.
- Map integration with Twitter and other social networks
- Map eBook titles to Amazon's genre designations to assess dominant genres

For larger images, and further preliminary findings, please visit:

<http://www.sprowberry.com/kindle/>