Toward An Epic Epigraph Graph

Francis Bond, Graham John Matthews
Linguistics and Multilingual Studies, English
School of Humanities
Nanyang Technological University
www3.ntu.edu.sg/home/fcbond/
www.soh.ntu.edu.sg/Programmes/english/FacultyandStaff/Pages/Graham-Matthews.aspx
bond@ieee.org, gmatthews@ntu.edu.sg

Abstract
We present a database of epigraphs collected with the goal of revealing literary influence as a set of connections between authors over time. We have collected epigraphs from over 12,000 literary works and are in the process of identifying their provenance. The database is released under an open license.

Keywords: epigraph, influence, digital humanities

The best ideas are common property.
Seneca the Younger,
“On Old Age”, Moral Letters to Lucilius

1. Introduction
An epigraph is a quotation at the beginning of some piece of writing. Genette (1987, Ch. 7, pp156–160) identifies four major uses: (i) commenting on and justifying the title of the work; (ii) commenting on the text of the work (the most canonical usage); (iii) claiming a relationship to the cited author (name dropping); (iv) signaling the genre of the work (romantic and gothic authors use more epigraphs, classical and realistic fewer). As a results authors use them both to set the theme and to link their work to the existing body of literature.

We have built a database of epigraphs, as a preliminary step to creating a network which maps epigraphs and their provenance (author, work, date, and country of origin) and consequently reveals literary influence as a set of connections between authors over time.

Epigraphs are indicative of the intended audience: for example, Evelyn Waugh uses epigraphs written in Ancient Greek, which suggest that the reader is expected to have some familiarity with the Classics, while Ali Smith utilises epigraphs from around the world, suggesting a more cosmopolitan readership. Most importantly however, they function as a clear marker of influence. From their choice of epigraph, we know that writers such as Jeanette Winterson and Margaret Atwood were influenced by Greek tragedy, Ian McEwan was influenced by Jane Austen, and that Roberto Bolaño’s influences range from Malcolm Lowry to Petronius. Reading literature as a communal network of ideas, philosophies, and artistic practices rather than a set of discreet objects signals the ways in which societies, institutions and structures of power represent themselves in art and language and are shaped in turn by representations.

The epigraph database has the potential to not only provide a new means of describing and analysing the development of the epigraph but to reshape our understanding of the dissemination and international reception of the novel more generally.

1https://epigraphic.tumblr.com/
Epigraphs also appear in more technical works, for example *The TeXbook* (Knuth, 1984) cites epigraphs both from the poet Byron and the Proceedings of the United Typothetæ of America. Although our database focuses on literary works, we also include a few technical works, and hope to add more in the future.

## 2. The Epigraph Database

The database contains the following records:

- **the epigraph**
  - text of the epigraph
  - image of the page it appears on or link to online text
  - title of original text
  - author of original text
  - country of origin or more detailed region if available
  - year of origin
  - language of the epigraph
  - original language of the epigraph if different from the language it is cited in
  - medium (novel, play, song, ...) if it exists
  - ISBN of the original text (if it exists)

- **the work**
  - title
  - author(s)
  - country of first publication or more detailed region if available
  - year of first publication
  - language of work (all currently English)
  - original language of work
  - genre
  - ISBN (if it exists)

- **Remarks** (for other information such as whether the cited work is fictional)

For example for this paper, the data would be:

- **Epigraph**
  - The best ideas are common property.
  - [image of page 1]
  - “On Old Age”, Moral Letters to Lucilius
  - Seneca the Younger
  - 65 AD
  - Rome
  - English
  - Latin
  - Letters
  - ISBN=9780674990845 (Loeb Classical Library)

- **Work**
  - Toward An Epic Epigraph Graph
  - Francis Bond, Graham Matthews
  - Miyazaki, Japan
  - 2017
  - English
  - Academic
  - ISSN=???

* shows data that was deduced, rather than explicit in the original. For works such as *Moral Letters to Lucilius*, for which many editions exist, we select one – the goal is to link to a controlled vocabulary to allow further look up of metadata.

The database can be thought of as a very unconnected graph. However, by backing off to the meta-data (e.g. group authors from a location or time or genre as one node) the graph can be made more connected.

### 2.1. Selection Criteria

The acquisition of data followed a combination of opportunistic and guided data collection strategies. First the data collection started at Nanyang Technological University’s Humanities and Social Sciences Library. Research assistants (RAs) photographed epigraphs and the novels’ metadata using smartphones and then entered the data into the database. They used https://isbnsearch.org in order to accurately record the ISBN and https://www.bookbrowse.com to record the date of first publication.

We took advantage of the locally hosted Singapore Literature in English Bibliography (Koh, 2008) to ensure that all Singaporean literary texts were checked for epigraphs.

The RAs then systematically worked through the novels in all of the public libraries in Singapore. At this stage, it became clear that Young Adult Fiction would dominate the database since these writers typically publish in large quantities and have a high proportion of epigraphs. The RAs added a tag to these texts so that they could be identified as such their effects on the database accurately recorded.

Since public libraries display a preference for contemporary literature, the RAs also developed a list of influential literature from the seventeenth century to the present (based on Wikipedia lists). They then sourced electronic copies of these texts using the Eighteenth Century Collections Online (ECCO),2 the Literature Online Database (LION),3 Google Books,4 and Project Gutenberg5 to record the epigraphs. The literature list records whether or not each novel contains an epigraph. This gave us a picture of the history of the epigraph and its spread. The RAs later returned to the dataset for quality control: removing repeat entries, double checking the accuracy of the metadata, and normalizing the data.

2 https://quod.lib.umich.edu/e/ecco/index.html
3 http://www.proquest.com/products-services/literature_online.html
4 https://books.google.com/
5 https://www.gutenberg.org/
2.2. The State of the Database

There are currently 10,921 records in the database, although not all fields are complete. Many epigraphs do not identify the source at all, or only the author’s name, it is rare to give the work and date. In addition misquotation (or paraphrasing) is common.

The most common works cited in epigraphs are shown in Table 1, the most common authors in Table 2, the most common types of works in Table 3 and finally the most common quotations used in epigraphs are shown in Table 4. The most cited author (if we exclude the bible) is Shakespeare, by an extremely wide margin. Gothic and fantastic authors are common, reflecting the fact that these genres use epigraphs more. Finally, it is interesting to see that poetry is the most common source of epigraphs, with plays, songs, and proverbs all also popular.

| # Cites | Work Cited            |
|--------|-----------------------|
| 129    | Bible                 |
| 35     | The Tempest           |
| 32     | Hamlet                |
| 23     | Paradise Lost         |
| 23     | Macbeth               |
| 22     | Romeo and Juliet      |
| 20     | As You Like It        |
| 19     | The Book of Counted Sorrows |
| 18     | King Lear             |
| 15     | Alice’s Adventures in Wonderland |

Table 1: The most common works quoted in epigraphs

| # Cites | Author Cited          |
|--------|-----------------------|
| 323    | Shakespeare, William  |
| 74     | Dickinson, Emily      |
| 62     | Poe, Edgar Allan      |
| 45     | Wilde, Oscar          |
| 44     | Whitman, Walt         |
| 43     | Blake, William        |
| 37     | Nietzsche, Friedrich  |
| 36     | Carroll, Lewis        |
| 35     | Thoreau, Henry David  |
| 35     | Milton, John          |
| 34     | Emerson, Ralph Waldo  |
| 33     | Einstein, Albert      |
| 32     | Twain, Mark           |
| 32     | Frost, Robert         |

Table 2: The most common authors quoted in epigraphs

The most commonly cited quotations are all from famous literary works, and show a wide spread from Ancient Greece and China (Aristotle and Sun Zu) to modern novels (such as Faulkner). The most common works are dominated by the Bible and Shakespeare, and include one originally non-existent work The Book of Counted Sorrows. This comes from a young adult series where the author cites poems from a fictional book at the start of each book.6

Later, the author wrote the book (Koontz, 2001). Another book cited by the same author, The Book of Counted Joys, remains non-existent.

2.2.1. Access

The epigraph database is released under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, which allows you to share and adapt in any medium or format for any purpose, so long as you give appropriate credit, provide a link to the license, and indicate if changes were made. A snapshot is currently available at http://compling.hss.ntu.edu.sg/projects/epigraph/, and future releases and analysis will be made available there.

3. Future Work

In future work, we intend to both increase the size and richness of the database, further analyze it and add support for visualizing patterns. First we intend to add another 10,000 or so epigraphs, and fill in as much missing information as possible. This work will be completed within a year. The revised network will be published as an open-access online resource available to other scholars and researchers as well as members of the general public.

We wish to take advantage of linked open data to link locations to the geonames database and the works, through ISBN, to further metadata, with the help of NTU’s librarians. This helps both with normalization and checking of the data. Having an ISBN number allows us to link to the library catalogue’s controlled vocabulary.

We also aim to display computer-generated visualisations of the map of literary influence in a manner accessible to a lay audience at venues such as the British Library and the Art-Science Museum in Singapore: we give sample visualizations in Figures 1 and 2. This will involve both clustering and visualization.

---

6In a letter dated August 10, 1992, Koontz stated: “Actually, there is no such book. I made it up. The way you

5-real-books-inspired-by-fake-books

4.0/

http://geonames.org/
Further, we will use the network of literary epigraphs to test connections and densities. For instance, we can map the influence of key literary authors such as Charles Dickens or Jane Austen by seeing how much of the network they capture. We can similarly track the influence of women writers or writers of a particular nationality or movement. For example, this project would allow us to determine the extent of the influence of the key Russian realist writers, Turgenev, Tolstoy, and Dostoyevsky on European literature of the late-nineteenth century. When are the Ancient Greek and Roman writers in their ascendancy? Is eighteenth-century satire an influence on postmodern writers of the late-twentieth century? Are lines of literary influence intercontinental or do they cluster around Europe? Singapore has a burgeoning literary movement and this project can determine where in the world Singapore writers take their inspiration from: is it predominantly the English literary canon, the Chinese classics, Indian literature, Malay writers, self-contained, or derived from more eclectic sources? We can also determine the influence of Singaporean literature on world literature.

A figure like William Shakespeare is widely acknowledged as important by literary critics because his works are meaningful and there is a great deal to be said about them. However, if we remove Shakespeare from the network, we can then explore the extent to which the network is altered: is Shakespeare such a focal point that it breaks apart or do the connections occur with a high enough frequency that its shape is unaffected? Conducting a series of case studies such as these presents us with fresh insight into literary and cultural influence, the dissemination of ideas, and the circulation of people.

Finally, the data can also be used to track influence though miscitation. Scholars have found the academics often copy citation information (including mistakes) without checking it, and possibly without reading the paper (Simkin and Roychowdhury, 2003). Similarly in our database, in addition to six citations of The past is never dead. It’s not even past.
from Faulkner (1951), we have several variants, shown in Table 5. We would like to investigate whether the change from never to not, which happens twice, reveals a connection between the two citing works. We add a link in the database between a variants of the same text.

4. Conclusions

We have created a database of 10,000 epigraphs, which we released under an open license. It allows us to study how works influence each other, as well as the use of the epigraph itself.

Acknowledgments

This research was supported by the Singapore MOE Tier 1 Grant: Digital Mapping the Literary Epigraph: Quantitative Analysis of Literary Influence Using Network Theory and Thousands of Epigraphs (RG158/16) and the NTU COHASS Cluster on Digital Humanities. We would like to thank the reviewers for their insightful comments.

5. Bibliographical References

Faulkner, W. (1951). Requiem for a Nun. Random House.
Genette, G. (1987). Paratexts: Thresholds of Interpretation. Cambridge University Press. (trans Jane E. Lewin).
Knuth, D. (1984). The TeXbook. Addison-Wesley.
Koh, T. A. (2008). Singapore Literature in English: an Annotated Bibliography. National Library Board Singapore and Centre for Liberal Arts and Social Sciences, Nanyang Technological University. (now online and updated: https://eps.ntu.edu.sg/client/en_US/SingaporeLiterature/).
Koontz, D. (2001). The Book of Counted Sorrows. Charnel House.
Paul Poplawski, editor. (2017). English Literature in Context. Cambridge University Press, second edition.
Simkin, M. and Roychowdhury, V. (2003). Read before you cite! Complex Systems, 14:269–274.
Variant | Cited in
---|---
The past is never dead. It’s not even past. | cited by 6 authors
The past is never dead, it’s not even past. | Carey, Peter (2012)
The past is not dead. It is not even past. | Wolff, Isabel (2014)
The past is never dead. It isn’t even past. | Ohanesian, Aline (2015)
The past isn’t over. It isn’t even the past. | Poulson, Christine (2016)

Table 5: Variants of *The past is never dead.* . . .
Differences underlined