

A Database of Online Book Response and the Nature of the Literary Thriller

Peter Boot

peter.boot@huygens.knaw.nl
Huygens ING, The Netherlands

Introduction

The study of literature has traditionally focused on the literary work, and sometimes its author, rather than on the response that works evoked in their readers. The arrival of the computer in the study of literature has not really changed that – perhaps unsurprisingly, as reader response has never been systematically recorded. The fact that readers have begun to document their reading and reading response on websites is therefore very fortunate (Gruzd and Rehberg Sedo, 2012, Maryl, 2008: 390-406). Booksellers' sites such as Amazon and review sites such as Goodreads, as well as weblogs, forums and general-purpose social media sites provide access to first-hand reading reports. Though most research on these sites focusses on (behavior of) users (Nakamura, 2013: 238-243, Thomas and Round, 2016: 239-253), we are beginning to see them being used in literary research (Finn, 2011).

This paper presents the Online Dutch Book Response (ODBR) database, that was designed to facilitate research into book response. At present, the database holds reviews and other response items from an online bookseller as well as from four Dutch-language mass review sites, including, where available, the information about books, reviewers and sites necessary to put the reviews into context.

To show one type of research that the database supports, the paper displays a clustering of reviews by genre, based on frequently used words. I discuss the clustering and what it suggests for further research.

Content: mass review sites, reviews from booksellers' sites, weblogs

While the ODBR database was designed to hold any kind of (online) book discussion, at the moment it holds 280,000 response items from four mass review

sites: three from the Netherlands (hebban.nl, dizzie.nl and watleesjij.nu) and one from Flanders (lezerstippenlezers.be). It also holds 313,000 items from the main online bookseller in the Netherlands, bol.com, and 36,000 brief expert reviews. We are currently working on downloading about 400 book blogs.

Mass review sites are sites where the main focus is on book reviews uploaded by the sites' users. The best known example is Goodreads (goodreads.com) (Thelwall and Kousha, 2016: 972-983). Hebban.nl, for example, is currently the largest mass review site in the Dutch-language area. It is in many respects like Goodreads: users post reviews, they can follow other users and they can create book lists. Other users can respond to the reviews or vote them up or down. There are moderated reading clubs on specific books. Apart from the user-contributed content, Hebban also holds a fair amount of editorial material, among other things expert reviews, blog posts, interviews and giveaways.

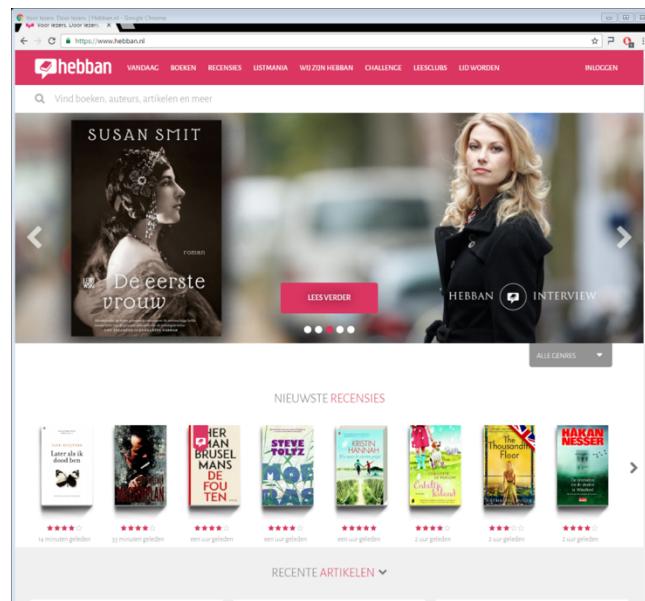


Figure 1. Front page of Hebban.nl

The other three sites are largely similar. Lezerstippenlezers.be ('readers tip off readers') is somewhat simpler in that it lacks the social functions of the other sites. Noteworthy about dizzie.nl is among other things that the site was downloaded for inclusion into the ODBR database only days before it was closed down. The site administrators explicitly stated in their final announcements that no archive would be kept. The ODBR database may be the most

complete record of the site's existence. Since then, watleesijj.nu has also been closed down.

Database design and content

The purpose of the database is to facilitate research into online book response. It should be able to store the response texts as well as information about the response's context. Response items are not just reviews. They include book lists, expert reviews, blog posts and review responses and other response types. Ratings and tags are also stored in the database. Figure 2 shows the main entities in the database.

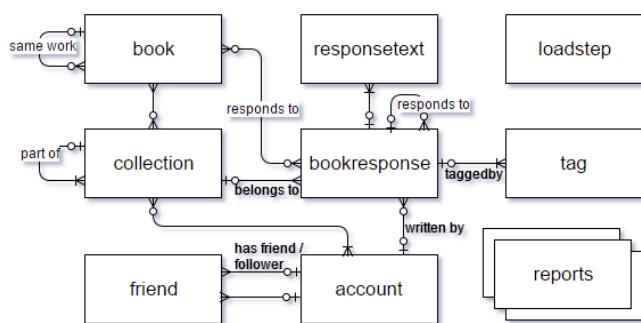


Figure 2. Main entities in ODBR database

The database now contains the following numbers of records: 146,800 books, 58,000 user accounts, 40,000 friendships/followers, and 628,800 book responses. The types of the responses are given in Table 1.

Article type	Number
article	4584
<u>articleresponse</u>	7445
blog	6122
<u>blogresponse</u>	2536
<u>bookdesc</u>	149002
<u>expertreview</u>	36634
list	10280
<u>listresponse</u>	2380
PM	14021
quote	3389
review	382998

Table 1. Article types in ODBR database. PM: private message; Bookdesc: book descriptions provided by sites.

The 'response' items (blogresponse, etc.) are responses to a response: a blogresponse is a response to a blogpost.

This large collection of book response items and context information creates many different possibilities for research. First of all, the response texts facilitate the investigation of response to individual books, authors and genres. The texts show the norms that readers apply as well as the way reading affects them. The availability of expert reviews alongside general readers' responses creates the possibility to study differences between (semi-)professional and lay reading. Other types of content enable different lines of research. Information about friends and followers can help investigate the influence of the social environment in reading choices and book appreciation. As many writers use the book review platforms to get in touch with their (potential) readers, the collected data can also provide insight into their marketing strategies. Information about the book lists that readers create (for instance 'to read', 'read in 2014') shows which books are perceived as similar, prompting the question whether they will also be rated similarly. The integration of the discussions and the context information from multiple sites in a single environment that facilitates integrated querying is something that, as far as I know, has not been done before.

Unfortunately, because of copyright and privacy concerns, the database is not accessible over the web. Researchers who are interested in accessing its content are asked to contact the author.

Clustering by genre

Of the research possibilities that the ODBR database offers, here I discuss just one example, an investigation in the word use in reviews by genre. This is an interesting subject, as word usage can be considered as an indication of how people respond to books. I will show a clustering of the genres by word use, which should give a first indication of which genres are perceived by readers as similar.

Dutch publishers use a shared system for classifying their books. This so-called NUR code (an abbreviation meaning Dutch-language Uniform Categorization) covers aspects of format as well as genre. On some of the downloaded sites, books were assigned NUR codes. As the load process of the database tries to merge the book information from multiple sites, NUR codes are available for 75% of the reviews. In the computations, reviews were merged by NUR code. Relative frequencies were computed for all

words, these frequencies were then transformed into z-scores. The Euclidean distances between the frequency vectors for the 200 most frequent words were computed and formed the basis for the clustering dendrogram in Figure 3. The figure only shows NUR codes for which there are more than 500 reviews available. For other choices for most frequent words and distance measure, the clustering is largely similar.

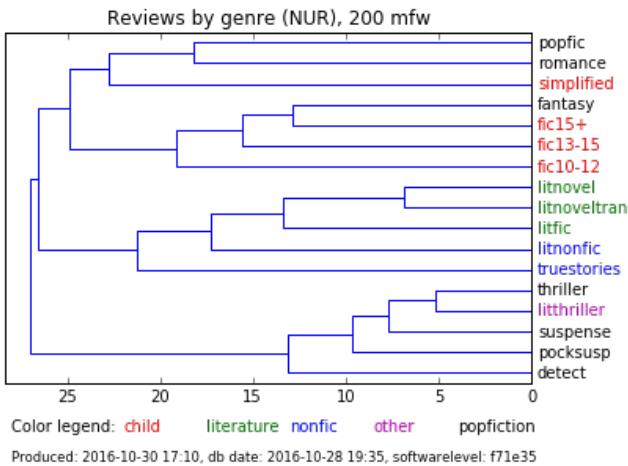


Figure 3. Dendrogram of NUR codes, based on distances between word usage in the corresponding reviews. The colors of the NUR labels reflect a higher-order grouping of genres (see legend)

A number of interesting groupings appear: the literary novel, original and in translation, groups nicely with other literary fiction. At a higher scale literary fiction groups with literary nonfiction and, interestingly, with true stories. Fantasy groups with youth literature, maybe a reflection of its popularity among younger readers. A group of general popular fiction and romance is also close to youth literature. Perhaps the most remarkable in the clustering is the location of the literary thriller. It sits squarely on one branch with the other suspense books such as the regular thriller and the detective. Those who doubt whether the 'literary' in the literary thriller is more than a marketing label, will see their views confirmed by this clustering.

Looking for an explanation, we can drill down to look at the individual words underlying the dendrogram and see for example how readers of (literary) thrillers talk about plot and plot lines, as one would expect. They also speak about 'main characters'(plural), while people who discuss literature use the singular 'main character'. That suggests an interesting difference between literature on the one hand and (literary) thrillers on the other.

To praise a book, readers of literature use 'beautiful' and 'nice', readers of (literary) thrillers use 'good' or 'great'. For other observations I am still looking for an explanation. Why, for example, do people who discuss literature use more personal pronouns? That question, like many others, requires further investigation.

Conclusion

Until now, most humanities-oriented researchers that have worked on online book discussion sites and communities have taken a qualitative approach (Fister, 2005: 303-309, Foasberg, 2012). The ODBR database is meant to facilitate quantitative research into online book discussion and, through the lens of online book discussion, into literature, both with respect to its effects on readers and as a social phenomenon. The rich data model and the large quantity of available data should provide support for both language and network oriented research approaches.

Bibliography

- Finn, E.F.** (2011). *The Social Lives of Books: Literary Networks in Contemporary American Fiction*. PhD, Stanford University.
- Fister, B.** (2005). "Reading as a contact sport". *Reference & User Services Quarterly*, 44 (4): 303-09.
- Foasberg, N.M.** (2012). "Online Reading Communities: From Book Clubs to Book Blogs". *The Journal of Social Media in Society*, 1 (1).
- Gruzd, A. and Rehberg Sedo, D.N.** (2012). "# 1b1t: Investigating Reading Practices at the Turn of the Twenty-first Century". *Mémoires du livre*, 3 (2).
- Maryl, M.** (2008). "Virtual Communities – Real Readers: New Data in Empirical Studies of Literature" in Auracher, J. and Van Peer, W. (eds.), *Virtual Communities – Real Readers: New Data in Empirical Studies of Literature*. Cambridge: Cambridge Scholars Publishing, pp. 390-06.
- Nakamura, L.** (2013). "Words with Friends: Socially Networked Reading on Goodreads". *PMLA*, 128 (1): 238-243.
- Thelwall, M. and K. Kousha** (2016). "Goodreads: A social network site for book readers". *Journal of the Association for Information Science and Technology*, 68 (4): 972-83.

Thomas, B. and J. Round (2016). "Moderating readers and reading online". *Language and Literature*, 25 (3): 239-53.