ARTICLE

'A pretty sublime mix of WTF and OMG'. Four explorations into the practice of evaluation on online book reviewing platforms

Peter Boot[1]

[1] Huygens Institute for History and Culture of the Netherlands

The article uses a corpus workbench (Sketch Engine) to investigate practices of evaluation in online book reviews. The reviews were taken from Goodreads, Amazon, bol.com and a number of Dutch online book discussion platforms. We look at tools that have been used to study online book reviews. Then we investigate our own collection of reviews. Findings suggest (1) that online reviews are not just centred on the reviewers' experiences but include solid discussion of the merits of books; (2) that reviewers of suspense prefer plot and character while reviewers of literary books prefer style and story; (3) that literal and metaphorical phrases referring to the body are often used in describing positive reading experiences; and (4) that positive reviews recount parts of the story, while negative reviews try to explain why the book was a disappointment.

Introduction

Online book reviewing platforms are a rich source of information for studying various aspects of reading, writing and other book-related behaviour (Rebora et al.). Reviews may be investigated in literacy studies, they give access to the reader's experience, but may also be informative of wider social issues such as social inequality (D. L. Miller; Driscoll and Rehberg Sedo; Dörrich). Reviews can be investigated to throw light on the books that they discuss and to learn about influence on the book market (Harris; Sutton and Paulfeuerborn). In this article we will use tools of corpus linguistics to look at various aspects of evaluation of literature as practised on online book reviewing platforms.

We will first give an account of the book reviewing sites and the reviews that their visitors write. Then we will introduce our corpus-based methodology, specifically Sketch Engine, a toolset developed originally for lexicographical research, and describe the data that we will use in the rest of the article. Then we will carry out a number of exploratory investigations to see what corpus approaches can teach us about reader response to literature and about reading more generally.

Many researchers have approached online book reviews as an alternative form of the newspaper literary review (Salgaro and Rebora; Bachmann—Stein and Stein; Neuhaus). It is true that to some extent both types of review fulfil the function of telling prospective readers what to expect from a given book. However, for us the value of the review is especially in what it can show us about readers' reading experiences and book evaluations. According to Alison Hegel, '[...] reviews are a rich source of first-hand insight into modern readers' expectations for and reactions to books,' (4-5) and Martin Rehfeldt writes 'Lay

reviews should not be seen as degenerate professional criticism [...], but as documents of reception and as potential objects for research [...]' (286).[1] We briefly discuss some limitations of using reviews for this purpose in the next section.

Some words must be said about the role of Sketch Engine in the preparation of this article. While there are many researchers who flock into Digital Humanities who have no coding skills themselves, there is an important school that believes that the intellectual benefits of computing only come to those who are code-literate.[2] And there is a strong belief in the DH community that the software that we write and use should be available as open source. By contrast, this paper uses a closed source lexicographical corpus workbench, Sketch Engine. This was a pragmatic choice.[3] It was selected for being interactive, user-friendly, and suited to the task, a choice further motivated by the resources available to the project. Usage of expert tools such as Sketch Engine should of course never be uncritical.[4]

Book reviewing platforms

There is a wide variety of online platforms where people discuss books (Boot, "Genre Analysis"). These include book sellers' sites such as Amazon, general purpose social media such as Facebook, YouTube, Instagram, Twitter and Tiktok, as well as weblogs and specialized web platforms such as Goodreads (Vlieghe et al.; Albrecht; Jaakkola; Gruzd and Rehberg Sedo; Driscoll). In this paper, we are interested only in platforms where visitors can leave reviews of books. We define a review as a usually evaluative textual response to a book, mostly from a personal perspective, that may range in length from a few words to a few hundred words or more. Platforms where people write reviews include the booksellers' sites and the specialized web platforms.

---

1 Translation mine.

2 See, for instance, Stephen Ramsay, "Who's in and who's out," *Defining digital humanities. A reader* Farnham/Burlington, Ashgate, 2013.

3 In fact, the author of this article could (perhaps) have hand-coded every query of which we will see the outcome in the following pages. That would have resulted in a number of Python programs that could have been made available for inspection through the journal's repository. Instead, I chose to use Sketch Engine, to which I only have temporary access thanks to the EU Elexis project, access that will end three months after writing these paragraphs. My readers will not be able to repeat my computations within an open-source software and neither of us will be able to check the algorithms applied to produce their outcome. If I am unlucky, and the paper's reviewers ask me to redo some of the computations with different parameters or different data, I may be unable to satisfy their request. Or I would have to take a subscription to Sketch Engine that will cost me or my employer at least € 500 per year (the price of a single-user academic license, including the required space for the corpora used in this study, but not much more. See https://www.sketchengine.eu/prices-for-academic-personal-accounts/, access date Dec 14, 2021). And even then, the algorithms might have changed, leading to different results, without me being able to do much about it. Is this an indefensible choice? I don't think it is. It would have cost me an inordinate amount of time to develop the required algorithms myself. But this is not just about time, it is also about trust. For some things, if you are not a professional, you have to trust professionals, in this case, the makers of Sketch Engine. My algorithms would certainly have contained bugs. They would have taken ages to run, as I have neither have the time, nor, probably, the skills, to optimize the algorithms and have no access to servers with powerful CPUs. It would have been impossible for me to dynamically query the data, proceeding by trial and error, running interesting queries, modifying them and running them again to gradually reach a better understanding. On the contrary, each idea would have had to be tried out running scripts that would have taken too long to wait for their results, forcing me to switch to other tasks in between and so taking the momentum out of the work.

4 For a look at alternatives, see a.o. Adam Kilgarriff et al., "The Sketch Engine: ten years on," *Lexicography* 1, no. 1, 2014; Laurence Anthony, "What can corpus software do?," in *The Routledge Handbook of Corpus Linguistics*, Routledge, 2022; Andressa Rodrigues Gomide, *Corpus Linguistics Software: understanding their usages and delivering two new tools*, Lancaster University (United Kingdom), 2020.

While the booksellers' sites generally do not offer the reviewers additional ways to interact with other readers, the specialized websites offer ways to make friends, to comment on posts, to have forum-based discussions as well as other interactive functionality, to the extent that sites such as Goodreads have been described as book-based social networking sites or as book communities (Thelwall and Kousha; Lukoschek, Katharina). I will describe the Goodreads-like sites as book-based 'affinity spaces', using the term coined by education scientist Gee to avoid the assumption of 'belonging' inherent in the word 'community' (Gee, *Situated Language*). Affinity spaces are 'loosely organized social and cultural settings in which the work of teaching tends to be shared by many people, in many locations, who are connected by a shared interest or passion' (Gee, "Affinity Spaces"). The booksellers' sites and the book affinity spaces also differ in demographics: while Amazon book reviewers are (as far as known) mostly male and over 50, participants on book affinity spaces are mostly young and female (Pinch and Kesler; Jessen). On both platform types most participants are well-educated.

Some research has been done on differences between reviews from the overtly commercial booksellers' sites and the non- or tacitly commercial book affinity spaces (Newell et al.). Reviews on commercial platforms are generally shorter (but not all studies agree), they use more vocabulary related to the act of buying, they devote less attention to the content, style and author of a book. However, since the mentioned studies vary in their design, with only two comparing the same pair of sites, it is hard to draw generalizable conclusions from their findings.

Some researchers have questioned the value of online reviews because the platforms are usually owned by 'big tech'. Already in 2013 Nakamura wrote of Goodreads that 'open access to a for-profit site like Goodreads has always exacted a price—loss of privacy, friction-free broadcasting of our personal information, the placing of user content in the service of commerce, and the operationalization and commodification of reading as an algocratic practice' (9). I believe this is an overly pessimistic view, which underestimates the agency and autonomy of readers who freely choose to use the site for their own purposes. That for Amazon a book is a commodity does not 'commodify' reading. A term like 'algocratic' is demagogic, fostering the idea that there is something inherently sinister about algorithms. More recently, Simone Murray wrote an essay about Goodreads and book history which seems to assume that if our reviews are useful to Amazon, that must necessarily be an evil thing, leading to 'enmeshing books in a world of vast profits, corporate IP and data-mining' (983; see also Walsh and Antoniak). For many researchers it is apparently hard to accept that a company such as Amazon, with all its ruthlessness towards its employees and competitors, can still maintain a site that is a reader's delight and a useful resource for the scholar at the same time.

That we must be aware, when using data in research, how the data came about, is evident, and the black box character of the platforms' algorithms can make that difficult. The platforms' commercialism, however, is certainly not the only influence on the evaluative content of the reviews. Book reviews, as all activity on social media, are also a performance of self. Book reviewing platforms can be seen as 'exhibition spaces where individuals submit artifacts [=the reviews] to show each other' (Hogan 377). Jaakkola writes about the objective of 'creat[ing] a public image of oneself as a reading person' that can influence book reviews, and self-presentation is an important motive for blogging (Tian). While book reviewing research so far only limited attention was devoted to this phenomenon, we should clearly be alert to it.

Methodology: Corpus approaches

Use of the term 'corpus' in this context derives from its use in corpus linguistics. McEnery and Hardie define a corpus informally as 'some set of machine-readable texts which is deemed an appropriate basis on which to study a specific set of research questions.' In linguistics, a corpus approach bases itself on the study of (large rather than small) collections of observed language data, rather than on the researcher's intuitions on what constitutes correct or appropriate language. Corpora and corpus linguistic tools are increasingly used outside of the linguistic community, as is the case in this paper (Goźdź—Roszkowski and Hunston). The tools used in corpus approaches include the analysis of frequency lists, possibly limited to certain types of words, key words (words that occur significantly more in one corpus than in a reference corpus), collocations and concordances (Jaworska). There is no reason, however, why corpus approaches should be limited to these 'canonical' tools. A distinction is sometimes made between corpus-based and corpus-driven approaches (Tognini—Bonelli). In the former approach, corpora are used to study predefined, theoretical linguistical features. The latter approach is inductive 'so that the linguistic constructs themselves emerge from analysis of a corpus' (Biber). In our analysis of book reviews, we limit ourselves to a number of corpus-based explorations.

The specific corpus tool that we will use is Sketch Engine (Kilgarriff et al.). Sketch Engine is a popular corpus analysis toolset developed principally for lexicographical research. Its main component is the word sketch, a tool for collocation analysis (Kilgarriff and Tugwell). Rather than working on the basis of frequencies of adjacent or near-adjacent words alone, word sketches take into account the grammatical relations that can exist between words. For a verb, word sketches can be used to display words that occur as subject or as object of the verb. For all word classes, word sketches can show words that are used to qualify the target word, that occur in 'and/or' constructions with the target word, etc. The candidate collocations are selected on the basis of a 'sketch grammar', a set of regular expressions over the sentences' POS-tags. It can display these collocations ordered by salience. This makes Sketch Engine a useful tool for investigating what are the contexts in which, in a certain corpus,

words are used.[5] It is also possible to create word sketch differences, a display of the difference in behaviour of two words in the same contexts. This tool can also be used to compare the behaviour of a single word in two different subcorpora.

Beyond the word sketch, Sketch Engine has other tools that one would expect of a corpus analysis toolset, such as support for metadata at the document level and support for CQL (Corpus Query Language), as well as tools that are specifically useful in a lexicographical context, such as automatic dictionary drafting. Beyond its home base in lexicography, Sketch Engine is also beginning to be used in wider Digital Humanities contexts (Moreton; Mpouli and Ganascia).

Data

To examine the possibilities of Sketch Engine for investigating reviews, we use two corpora. The first corpus that we use is in English and contains N= 373,854 reviews taken from Amazon (52%) and Goodreads (48%). We selected reviews from existing corpora (Ni et al.; Wan et al.). The corpus contains two subcorpora defined by rating (four or higher and three or lower). The second corpus contains N= 200,410 Dutch-language reviews taken from the Online Dutch Book Response (ODBR) corpus (Boot, "Online Book Response"). About half of them come from bol.com, The Netherlands' largest online bookseller. The other half comes from various online book reviewing sites, comparable to Goodreads. The ODBR corpus as uploaded into Sketch Engine has subcorpora defined by rating as well as by genre. The genre subcorpora include one containing only reviews of suspense novels (including the so-called 'literary thriller') and one containing only reviews of (general) literature. The genre is assigned by the publisher.

When ingesting the data, Sketch Engine does lemmatisation and part-of-speech tagging. While no systematic evaluation of pre-processing quality was performed, repeated inspection suggests wrongly labelled words to be very infrequent, both for English and for Dutch.

As we saw above, both the demographics of the reviewers and properties of the reviews depend on the site type. Our corpora therefore include about equal quantities of reviews from overtly commercial (bookseller) sites and from reader-oriented sites where commercial interest is usually more tacit.

Aspects of evaluation

What follows are four exploratory investigations. In each of them, we very briefly discuss the theoretical relevance of the topic that we investigate and formulate some questions. Then we will use one or more of the tools available

---

5 Sketch grammars are available for many languages besides English. The grammar for Dutch is described in Carole Tiberius and Adam Kilgarriff, "The Sketch Engine for Dutch with the ANW corpus," in *Fons verborum*, Leiden: Instituut voor Nederlandse Lexicologie/Gopher, 2009.

within Sketch Engine in an attempt to answer these questions; in a concluding paragraph, we will summarize our preliminary findings both with respect to the questions we formulated and with respect to what Sketch Engine can do. To be clear, none of the investigations pretends to provide final answers to the questions we discuss: their purpose lies in showing what is possible with review data utilizing an off-the-shelf corpus tool as much as in the initial answers we find.

It is important to note that not all readers are reviewers. Even on the reviewing sites, the reviewers are in a minority compared to those who use the platforms to find information or just to track their reading (Jessen). As far as I am aware, there exists no research comparing reviewing and non-reviewing readers with respect to their reading behaviour. It is likely that those who write longer reviews, being forced to put their thoughts into words, will have more explicit opinions. They are probably also those who are comfortable with revealing personal data (Blank and Reisdorf). Studies that explicitly compare the reading behaviour of these two groups are sorely needed.

Exploring the alleged inanity of online reviews: plot

In their classic introduction into literary evaluation, Von Heydebrand and Winko write disapprovingly of ordinary readers who 'usually bring their own needs and values to or into the work regardless of the text's intention' (186), and summarily write about a reader's 'heteronomous and therefore inadequate reading' (213). Following this traditional suspicion of lay readings harboured by literary scholarship the prejudice against lay reviews (for examples see Stein; Ernst) has been widely shared. According to Bachmann-Stein, 'Lay reviews do not orient themselves towards the standards of literary criticism. Rather, the reviewing is guided by psychological criteria referring to the impact of the individual reading experience' (89). This distrust of online book discussion went into overdrive in Ronán McDonald's pamphlet on *The death of the critic*. He saw it as a form of 'people power' that 'decks out banality and uniformity in the guise of democracy and improvement' (17).

To check whether this widely held assumption that reader reviews talk mostly about the reading experience is true, we will look at a word sketch of 'plot' (but might as well have discussed the writing, or the dialogues, or other issues – see below). Story or plot is a central concern in literary studies and a well-crafted plot is important for readers of all types of narrative (Boyd; Tobin). In online book reviews, the plot is one of the most frequently discussed aspects of the book (Kutzner). This is true for literary as well as for popular books (Daniels). What does a word sketch of 'plot' show us about how it is discussed? If the reviews are mostly about the readers' experiences, we would expect the collocates of 'plot' to refer more often to the readers' impressions, and in general to avoid technical discussion of books' merits.

As explained above, a word sketch displays the collocates of 'plot' in several contexts ('grammatical relations', in Sketch Engine terminology). Figure 1 shows how Sketch Engine presents the adjectives most frequently associated with the word 'plot' in the Goodreads corpus,[6] followed by the absolute frequency of the collocation and its (LogDice) typicality.[7] These are words such as 'predictable', 'interesting' or 'intriguing' that we would indeed expect to be used with 'plot'. Table 1 gives the full list of these adjectives. Some are mostly evaluative ('excellent', 'awesome', 'bad'), some are mostly descriptive ('twisty', 'straightforward'), others are evaluative as well as descriptive ('predictable', 'compelling', 'confusing'). Many of the words do indeed imply a reading experience ('intriguing', 'engaging', 'compelling'), and are necessarily subjective. But taken together, they provide a rich and balanced vocabulary for qualifying plots.
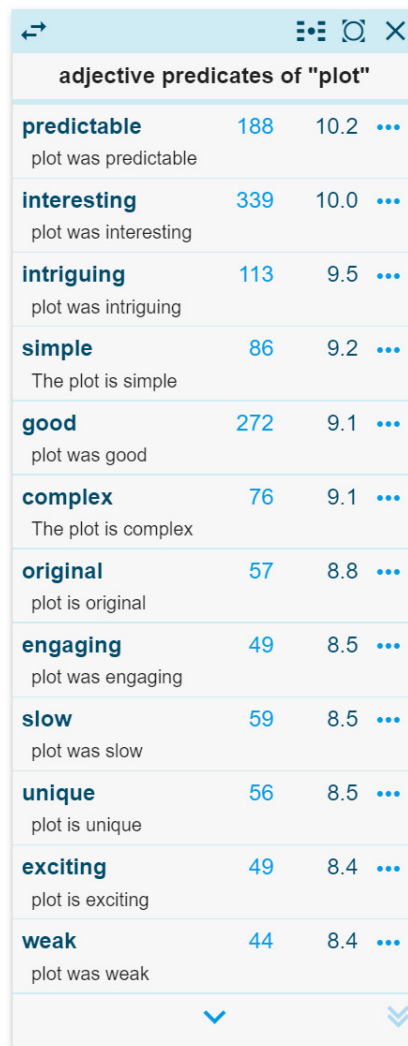
Table 1. Adjectives co-occurring with 'plot'

predictable interesting intriguing simple good complex original engaging slow unique exciting weak tight believable great full thin intricate excellent compelling entertaining decent solid easy fast-paced amazing clever fantastic ridiculous suspenseful similar boring realistic okay thick awesome gripping strong unbelievable plausible creative straightforward confusing unpredictable obvious riveting implausible basic different fascinating contrived hard brilliant captivating enough secondary rich fresh simplistic wonderful stupid formulaic bad new other shallow unrealistic enjoyable fine complicated unusual nice superb twisty unoriginal fast silly outstanding dull much flimsy sound intense unexpected familiar dark funny perfect

Table 2 presents nouns modified by 'plot' in the reviews. Some are compounds denoting aspects of plots in the books ('plot twists', 'plot devices', 'plot direction'), some are about the discussion of plot ('plot summary'). The algorithm is clearly imperfect: 'plot nothing' does not seem to make sense. Yet it is impressive that simple statistics and a few language rules can lift these patterns out of a few hundred thousand reviews. It is relevant because it shows that, at least collectively, online reviewers have a subtle and extensive vocabulary to talk about plot, even if individually they don't use all these collocations. If they were only interested in 'their own needs and values' there would be no need for this extensive vocabulary. This does not imply that the majority of reviews should discuss the plot – for one thing, most reviews are too short for that.

---

6 The list of frequently used adjectives is one of the grammatical relations that Sketch Engine shows for nouns. Others include 'prepositional phrases with plot', 'plot is a ...' and the 'nouns modified by plot' and 'plot and/or ...' that we will see later in this section.

7 See for an explanation of the LogDice statistic: Lexical Computing Ltd, *Statistics used in Sketch Engine*, Lexical Computing Ltd, Brighton, UK, 2015, https://www.sketchengine.eu/wp-content/uploads/ske-statistics.pdf.

Figure 1. Sketch Engine table showing adjective predicates of plot

---

Table 2. Nouns modified by 'plot'

twist line point device hole development thread element summary wise structure detail arc turn move idea progression center character spoiler contrivance advancement movement synopsis resolution description action dialogue theme issue change flow convenience characterization event strand advance lot writing hinge direction complication surprise pace premise devise building inconsistency sort tool setting right outline subplot complexity progress tension everything choice deal nothing something work conflict kind aspect perspective romance piece situation ending part time thing story

---

Plot is certainly not the only aspect of books that reviewers discuss. This is shown by another entry in the 'plot' word sketch, the 'Plot and/or ....' list. Table 3 shows an edited version (there were too many irrelevant words here). These words are used by reviewers in a conjunction with 'plot', so they write 'plot and writing', 'plot or theme', etc. The words in this list show that online reviewers are not merely interested in plot. We could look at the Work Sketch for each of these terms and find a similarly nuanced vocabulary for discussing

that aspect. For reasons of space, we can't go into these words' collocations, but taken together, these examples suggest that online book reviewing practices are certainly not simplistic – they are not merely based on feeling and identification – but display solid discussion on a variety of aspects related to the merits of books.[8] The corpus analysis tool allows us to perceive that by zooming out and looking at the patterns that emerge when we look at thousands of reviews collectively.

Table 3. Edited list of 'Plot and/or …'-collocates

character writing characterization setting story storyline action dialogue subplot theme pace style world premise structure mystery prose humor concept conflict idea characterisation sex intrigue cast language narrative tone adventure flow motivation tension drama storytelling

---

Exploring what constitutes 'beauty' in reading

Let us move from the noun 'plot' to an adjective, 'beautiful'. According to an empirical study by Knoop et al., 'beautiful' (or rather it's German translation 'schön') is the adjective that is most frequently suggested to describe an aesthetic literary experience. What does the way it is used suggest about that experience? Table 4 shows the nouns modified by 'beautiful' in the Goodreads corpus.

Table 4. Nouns modified by 'beautiful' in the Goodreads corpus. Italicised words probably refer to the story world rather than to an aspect of the book.

writing story prose *woman* illustration cover picture *girl* language artwork imagery setting tale *place* book art word passage piece *daughter* job poetry scenery work *city thing* style ending romance moment novel journey soul *world heart* poem *man creature landscape* message *wife* tapestry *friendship* collection writer *lady princess* sentence image edition *heroine* relationship *voice island country home child sister* conclusion read *house dress* portrait *boy mind people* scene *music eye* example *land* quote detail *color gown couple* line *area* narrative gift *countryside* drawing *wedding painting* character idea storytelling *name face love star* tribute epilogue *mansion widow horse stranger beach* depiction letter *maiden*

As we see, the list contains many of the words that in a book review would be expected to be qualified as beautiful ('writing', 'story', 'cover'). But the list also contains many words that do not refer to the book as a reviewed object,

---

but to the story world ('woman', 'city', 'gown').[9] Some could be used in both ways ('gift', 'drawing'). This presents us with a problem of 'opinion target identification' (Catharin and Feltrim).

Within the Sketch Engine context, there does not seem to be a way to differentiate the two domains. One option might be running a word sketch difference of the word 'beautiful' in subcorpora of positive and negative reviews. We could hypothesize that beautiful women occur in both high- and low-rated books, while beautiful writing probably occurs more in high-rated books. Then, in a word sketch difference of the word 'beautiful' between the high- and the low-rated reviews, the book domain collocates of 'beautiful' should appear on one end of the spectrum, the story-word collocates should be unmarked. But experiments show this doesn't work. Beautiful scenery, beautiful cities and beautiful women occur more in positive reviews than in negative ones. Because it seems unlikely that books about beautiful scenery (or cities, or women) are rated more highly, the findings imply that in negative reviews, reviewers may see no reason to mention these aspects of the story world in their reviews. We will come back to this when we use a key word analysis to investigate differences between positive and negative reviews, below.

Of the collocates that *do* refer to reviewed aspects of the book, some refer to the linguistic aspects of the book, to the writing ('writing', 'prose', 'language'), others refer to the story ('story', 'ending', 'narrative'). If we assume with Jakobson that the poetic function is the essence of literary art, and distinguish distanced readers (who focus on form) from identifying readers (who focus on character and plot), it seems likely that the form-oriented words appear mostly with literary books, the story-oriented words with suspense-driven books. We can test that hypothesis by creating a word sketch difference by genre (Heydebrand and Winko). For that, we move to the Dutch corpus, as there we have subcorpora based on genre.[10] We create in Sketch Engine a word sketch difference for the Dutch word 'prachtig' (one possible translation of English 'beautiful') between the literary and the suspense subcorpus. Sketch Engine shows the output as displayed in Figure 2. The words in green occur typically in reviews about literary books, the words in red occur typically with suspense novels. The numbers after the words give the frequency and (LogDice) typicality in the respective subcorpora.

The strongest indications for differences in how 'prachtig' is used for evaluating the two genres, can be seen in the columns '"prachtig" and/or …' and 'subjects of "be prachtig"'. Table 5 gives the translations of the Dutch words in Figure 2. Except for 'house', all words come from the reviewing

---

9 From a gender perspective, it is striking how many of these words refer to feminine characters: woman, girl, daughter, wife, lady, princess, heroine, sister, widow, maiden. 'Man' and 'boy' are the only ones referring to masculine characters. This would be worth a study in itself.

10 Subcorpora can be defined in Sketch Engine based on metadata of the corpus texts, known in Sketch Engine as 'text types'. For most reviews in the Dutch corpus we know the reviewed book's NUR code, a publisher-assigned genre label.

WORD SKETCH DIFFERENCE

prachtig (frgenlit) 15,281× | 6.0 4.0 2.0 0 -2.0 -4.0 -6.0 | (frgensuspllt) 4,335×

👁 modifiers of "prachtig" 👁

"prachtig" and/or ...				
indrukwekkend	52	0	10.1	— •••
poëtisch	16	0	8.7	— •••
ontroerend	135	12	10.7	7.7 •••
meeslepend	89	11	9.9	7.3 •••
aangrijpend	37	5	9.4	7.3 •••
prachtig	37	8	9.7	8.5 •••
beeldend	24	6	8.9	7.9 •••
boeiend	24	11	8.4	7.8 •••
spannend	28	34	7.5	8.0 •••
levensecht	5	5	7.1	8.5 •••
gruwelijk	0	5	—	8.2 •••
beklemmend	0	8	—	9.0 •••

subjects of "be prachtig"				
stijl	9	0	9.0	— •••
taalgebruik	11	0	9.0	— •••
zin|zinnen	7	0	8.8	— •••
schrijfstijl	21	0	8.5	— •••
einde	11	0	8.4	— •••
verhaal	69	7	8.8	5.6 •••
boek	74	14	8.5	6.2 •••
cover	7	25	8.6	11.0 •••
plot	0	5	—	7.2 •••
karakter	0	10	—	8.6 •••

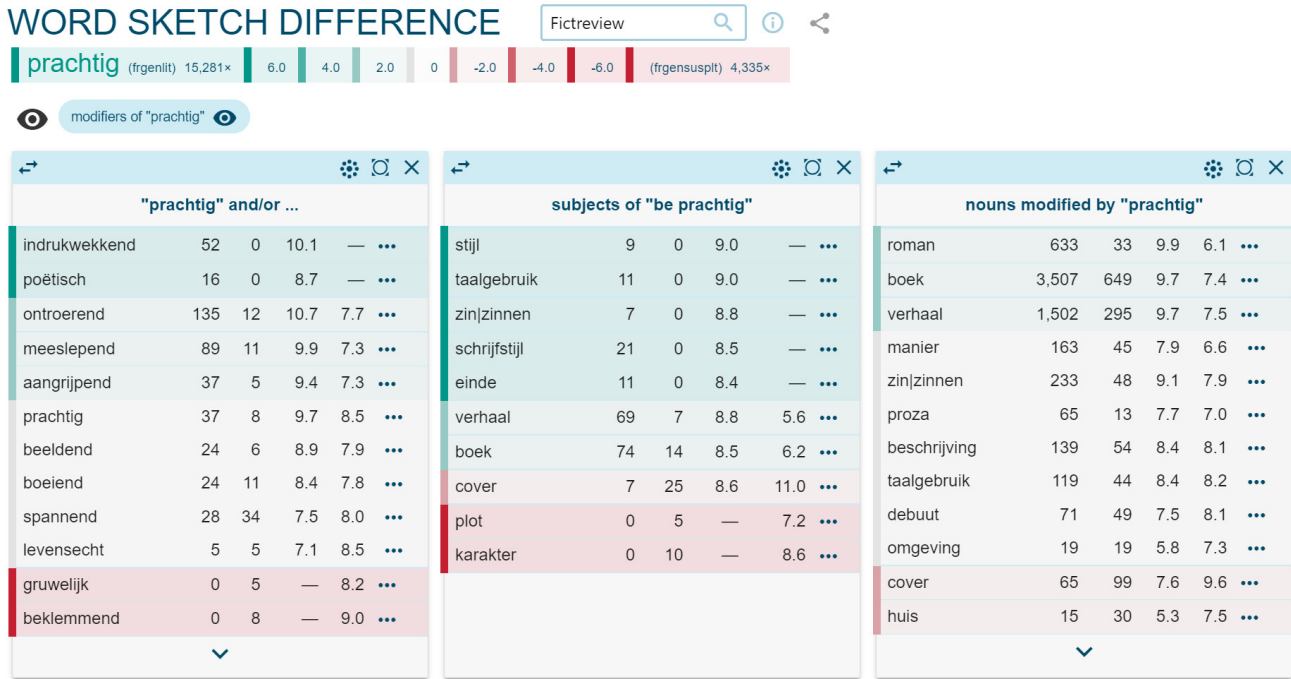| nouns modified by "prachtig"				
roman	633	33	9.9	6.1 •••
boek	3,507	649	9.7	7.4 •••
verhaal	1,502	295	9.7	7.5 •••
manier	163	45	7.9	6.6 •••
zin|zinnen	233	48	9.1	7.9 •••
proza	65	13	7.7	7.0 •••
beschrijving	139	54	8.4	8.1 •••
taalgebruik	119	44	8.4	8.2 •••
debuut	71	49	7.5	8.1 •••
omgeving	19	19	5.8	7.3 •••
cover	65	99	7.6	9.6 •••
huis	15	30	5.3	7.5 •••

Figure 2. Part of the Sketch Engine word sketch difference for 'prachtig' between the subcorpora literature and suspense in the ODBR corpus.

Table 5. Translation of the words in [Figure 2](#), Sketch Engine word sketch difference for 'prachtig' between the subcorpora literature and suspense in the ODBR corpus. Strongest collocates at top of first row and at bottom of second row.

	"prachtig" and/or ...	subjects of "be prachtig"	nouns modified by "prachtig"
Literary reviews	impressive poetic moving compelling, gripping moving, poignant	style language use sentence(s) writing style ending story book	novel book story
Suspense reviews	horrible oppressive	cover plot character	cover house

domain, not the domain of the story world. In the subjects-of column we see that what is described as beautiful in suspense reviews are character and plot[11] while in the literature reviews, the first four collocates refer to style. Still, the combination of 'beautiful' and 'story' also occurs more typically in literary reviews than in suspense reviews. Moving to the and/or column, we see that close associates of 'beautiful' in the suspense genre are 'horrible' and 'oppressive', words that are clearly connected to the effects of the story. In the literary reviews, 'poetic' and to a lesser extent 'impressive' seem clearly related to style; the other three words are all (near-synonyms of) 'moving'. While stylistic aspects undoubtedly are partly responsible for a text being moving,[12] what moves us are the characters and what happens to them.

---

11 And 'cover', also interesting.

12 For example Winfried Menninghaus et al., "The emotional and aesthetic powers of parallelistic diction," *Poetics* 63, 2017.

The modified-by column contributes further support to this discussion, 'story' again appearing as a collocate of 'beautiful' in the literary reviews, and 'house' as an element of the story world coming up in the suspense category, which might indicate that the setting of the story world be important to this type of reader. Summarizing, we could say that what suspense readers consider beautiful when evaluating a book are plot and character (and the cover); what literary readers consider beautiful is style and story. This seems to confirm the conclusion of Riddell and Van Dalen-Oskam that there is an overlap between distanced and identifying readers: distanced readers are identifying readers with more reading techniques at their disposal (with the corollary that all readers are identifying readers).

Exploring the role of the body in reading

That reading is an embodied activity is widely accepted among researchers:[13] '[P]eople imaginatively project themselves into text worlds via embodied simulations,' writes Raymond Gibbs (222). If this is true, we should see a reflection of that in the review texts.

In this connection, it is interesting to look once more at the word sketch for 'beautiful'. The top modifiers of beautiful (in LogDice order) are given in Table 6. While some of these modifiers are neutral, many have a clearly corporeal semantic component.[14] While for 'heartbreakingly' or 'painfully' the reference to bodily perceptions is quite explicit, we would like to suggest an embodied dimension also for 'surprisingly' and its stronger variants 'amazingly' and 'stunningly', for 'hauntingly' and others (C. R. Miller). It appears that bodily simulation in processing and understanding a fictional text, as described by Gibbs, transpires to the evaluation of that text.

Table 6. Top modifiers of 'beautiful' in the English corpus.

hauntingly stunningly heartbreakingly absolutely achingly amazingly simply incredibly tragically devastatingly truly utterly strangely painfully exquisitely yet equally surprisingly very

To look more systematically into the role of embodiment, we would like to search for references to body parts. While Sketch Engine does not know which words refer to body parts, we can find many examples of the body's involvement in reading thanks to the fact that 'you' in book reviews usually refers to the reader. A search for 'your' followed by a noun brings up many

---

13 See also Marco Caracciolo, "Degrees of Embodiment in Literary Reading," in *Expressive Minds and Artistic Creations: Studies in Cognitive Poetics*, 2018; Ben Morgan, "Embodied Cognition and the Project of the Bildungsroman: Wilhelm Meister's Apprenticeship and Daniel Deronda," *Poetics today* 38, no. 2, 2017, https://doi.org/10.1215/03335372-3869287; Christian Benne, "Tolle lege. Embodied reading and the "scene of reading"," *Language sciences (Oxford)* 84, 2021, https://doi.org/10.1016/j.langsci.2021.101357.

14 I removed the ones which upon inspection turned out to be used more with story-world phenomena such as trees or heiresses, 'breathtakingly', 'strikingly', 'so', 'exceptionally' and 'extremely'.

(2,990 items, 31,372 total frequency)

		Lemma	Frequency ↓	Relative ?	% of conc. ?		
1	☐	heart	2,224	47.75	7.09 %		•••
2	☐	life	1,609	34.54	5.13 %		•••
3	☐	mind	1,053	22.61	3.36 %		•••
4	☐	seat	955	20.50	3.04 %		•••
5	☐	time	909	19.52	2.90 %		•••
6	☐	attention	881	18.91	2.81 %		•••
7	☐	book	803	17.24	2.56 %		•••
8	☐	interest	631	13.55	2.01 %		•••
9	☐	eye	610	13.10	1.94 %		•••
10	☐	head	609	13.07	1.94 %		•••

Rows per page: 10 ▼   1–10 of 2,990 |< <   1   / 299   > >|

Figure 3. Most frequently occurring nouns after 'your' in the English corpus.

Many of the words on this and the following screens refer to body parts.

relevant phrases. We can do this in Sketch Engine using the CQL option in the concordance tool, and then creating a frequency graph for the noun. The top results are shown in Figure 3. The most frequently occurring word in Figure 3 is 'heart'. The context menu allows us to jump to a concordance of the phrase 'your heart'. Then we find phrases from reviews such as 'the book is going to break your heart (in the best way possible)', 'a great historical romance to warm your heart', 'deep moments, the ones that rip your heart out', or 'has your heart ever exploded with so much joy?'

Table 7 shows some verbs and expressions occurring with 'your' + body part. Some of these are literal expressions, others are metaphorical. That a book 'brings a smile on your face' or 'tears to your eyes', or that you 'shake your head' over a character's action – all these are literal expressions about bodily response to reading. However, most of the expressions in Table 7 involve the body in a metaphorical sense. With Gibbs and Matlock we assume that people might 'understand metaphors by creating an imaginative simulation of their bodies in action that mimics the events alluded to by the metaphor.' Expressions stating that a book makes you 'open your heart', that it 'blew off your head', that it 'hits you in the face' or 'gets under your skin' evoke bodily metaphors for what reading does to the reader.

Thus, the role of the body in these expressions is in most cases distinct from the bodily involvement in comprehending and experiencing (metaphors in) fictional language, which is the focus of Gibbs. They are rather comparable to embodied metaphors used for the process of reading, such as 'reading is a journey' or 'reading is eating' (Nuttall and Harrison; Ross; Herrmann and Messerli). What we see here is another way in which the body is likely to be

Table 7. Verbs and expressions occurring with 'your + body part' in the English corpus

heart	open, rip, break, reach, pull, ache, capture, shatter, touch
head	shake your head, have your head spinning, scratch your head, blow off your head
eye	cry, open, close, glued (to the page), tears in, in front of, roll
face	smile on your face, hits in your face, make you fan your face, blush to your face
toe	keeps you on your toes, dip your toe, get your toes wet, makes your toes curl
hand	get your hands on, raise your hand in prayer, hands over your mouth
breath	hold your breath, take your breath away, steal your breath away, catch your breath
mouth	read with your mouth open, leave you with your mouth hanging open, shape of your mouth smiling
skin	make skin crawl, get/creep under your skin, make skin itch
foot	sweeps/knocks/blows you off your feet, keeps you on your feet, put your feet up

inextricably involved in reading, as documented by lexis referring to body parts and processes. It is interesting to note that almost all of these expressions are used to describe positive reading experiences.

Exploring differences between positive and negative reviews

Research into reading has often focussed on the good things that reading can result in: empathy, self-reflection, an interest in others (Koopman and Hakemulder). But not all reading experiences are positive and not all people like all books. We can distinguish the positive from the negative reviews on the basis of the accompanying rating (number of stars). But what are the textual differences between positive and negative reviews? It stands to reason that they can be distinguished based on positive and negatively valenced words.[15] But what other differences are there? Taboada and others look at words and constructions that can be used to convey negativity and illustrate those, among others, with quotes from movie and book reviews. Among the negative constructions they mention are rhetorical questions, sarcasm, certain suffixes ('-let'), certain words ('actually'), direct quotations, the juxtaposition of a very positive intensifier (such as 'hilariously') with a negative adjective ('bad'), and why-questions. Crossley and others use LIWC[16] and their own sentiment analysis tool (SEANCE) to compare positive and negative reviews. They find that negative reviews use more negative emotions (particularly anger-related terms), negations, exclusion terms, spatial terms, understanding and certainty (as defined by SEANCE). Positive reviews use more positive emotion terms, social and certainty (as defined by LIWC) terms, as well as terms related to power and respect. They also note that certain POS-tags are useful for discriminating positive from negative reviews.

---

15 See for example Tom De Smedt and Walter Daelemans, ""Vreselijk mooi!" (terribly beautiful): A Subjectivity Lexicon for Dutch Adjectives." Paper presented at the Proceedings of the 8th Language Resources and Evaluation Conference (LREC'12), 2012.

16 "The development and psychometric properties of LIWC2007," 2007, accessed 2017-01-24, http://hdl.handle.net/2152/31333.

We will use a keyword analysis to investigate the characteristics of negative vs. positive reviews, complemented by a brief analysis using LIWC. Keyword (or keyness) analysis renders the statistically distinctive words of a focus corpus as compared with a reference corpus (Bondi and Scott). Rather than examining content, we will initially look at function words and POS-tags. We use our subcorpora defined by rating: a subcorpus *pos*, containing all reviews with rating four or five, and a subcorpus *neg* containing ratings one, two and three.[17] The reason for including three among the negative reviews is partly practical, as the distribution of ratings is heavily skewed towards the higher ratings. In the English corpus, the ratings 1 and 2 together account for less than 10 percent of the reviews, while the ratings 4 and 5 account for 74 percent. But that also means that a 3 represents a distinctly below-average reading experience. A book with rating 3 is typically described as 'okay' or 'decent', but generally still something of a disappointment.

Figure 4 shows the top 20 keywords defined by comparing the positive versus the negative English subcorpora.[18] The Sketch Engine keyword tool uses a parameter that has the effect of displaying either the rarer or the more common words in the keyword list. That parameter was set at its highest value (1,000,000) to select the most common words. Even so, we notice that, due to the nature of the corpus, some content words appear among the top twenty keywords ('read', 'great' and others). We will only look at the function words and notice that 'you', 'he' (and third-person possessive pronouns) as well as 'a' occur more frequently in positive reviews. Interestingly, when we reverse the comparison and look at keywords in negative versus positive reviews (results not shown here) we find 'I' (the other singular personal pronoun) and 'the' (as well as 'not', as predicted by Crossley et al.).

Before trying to answer why these words occur in a positive or negative context, we will step up the level of abstraction and look at key POS-tags. Figure 5 shows the key POS-tags in the positive vs. the negative corpus.[19] There is much to note here, for example that third-person singular present-tense verbs occur more in positive reviews (tags VHZ ('has'), VBZ ('is') and VVZ for other verbs), as do nouns (NN) and proper nouns (NP). In the reverse comparison (not shown here) we observe that past tense forms are more frequent in the negative reviews. Table 8 summarizes the observed differences.

An explanation of these differences begins with the observation that most NPs in the review texts refer to characters in the books, and only to a lesser degree to authors, places and book titles. One explanation would be that positive reviews

---

17 We ignore reviews with rating zero, which presumably represents a missing value.

18 After the first ten to twenty keywords, the differences in relative frequency quickly get so small that, while still statistically significant, it becomes hard to ascribe meaning to these differences. For the algorithm that Sketch Engine uses, see Adam Kilgarriff, "Simple maths for keywords." Paper presented at the Corpus Linguistics Conference, 2009.

19 The meanings of the tags are explained here: https://www.sketchengine.eu/english-treetagger-pipeline-2/ (access date Dec 14, 2021).

	Word	Frequency per million[?]				Word	Frequency per million[?]		
		Focus	Reference				Focus	Reference	
1	and	29,142.88	25,031.82	•••	11	life	2,120.74	1,322.17	•••
2	love	5,118.83	2,336.70	•••	12	to	23,204.42	22,404.09	•••
3	you	5,851.01	3,889.34	•••	13	his	3,948.55	3,177.78	•••
4	read	6,795.19	5,558.52	•••	14	can	2,839.75	2,073.88	•••
5	great	2,138.06	999.29	•••	15	story	5,620.59	4,967.74	•••
6	in	13,345.27	12,294.21	•••	16	of	20,901.72	20,281.42	•••
7	her	6,700.10	5,697.01	•••	17	book	12,021.79	11,436.20	•••
8	this	12,568.89	11,585.10	•••	18	he	4,370.56	3,793.80	•••
9	will	3,092.02	2,226.26	•••	19	wait	861.76	294.76	•••
10	a	26,784.53	25,920.75	•••	20	series	2,077.33	1,517.99	•••

Rows per page: 20 ▼   1–20 of 1,000 |< <   1 / 50   > >|

Figure 4. Keywords in the positive (focus) vs. the negative (reference) subcorpus

	Word	Frequency per million[?]				Word	Frequency per million[?]		
		Focus	Reference				Focus	Reference	
1	NP	38,870.14	30,929.74	•••	11	VVP	13,834.27	13,188.56	•••
2	NN	126,068.69	122,254.91	•••	12	VV	30,471.06	29,850.49	•••
3	PPZ	17,063.01	13,853.03	•••	13	WP	5,709.75	5,249.45	•••
4	VVZ	14,650.91	12,721.69	•••	14	VHP	3,162.47	2,722.85	•••
5	NNS	37,629.79	35,753.41	•••	15	MD	11,498.43	11,055.47	•••
6	SENT	53,773.14	52,197.49	•••	16	NPZ	3,224.66	2,810.69	•••
7	VBZ	17,864.39	16,592.40	•••	17	SYM	3,523.58	3,227.49	•••
8	CC	37,945.17	36,846.80	•••	18	TO	15,757.28	15,475.37	•••
9	VHZ	3,073.36	2,256.07	•••	19	IN	95,548.97	95,251.52	•••
10	JJ	67,739.84	66,910.65	•••	20	PDT	1,369.41	1,229.17	•••

Rows per page: 20 ▼   1–20 of 51 |< <   1 / 3   > >|

Figure 5. Key POS-tags in the positive vs. the negative subcorpus

more often than negative reviews recount part of the story. That explains the proper nouns, the use of 'he', the third person verbs, and also the present tense, as reviewers generally use the present tense to report on the story. A brief inspection shows that indeed most occurrences of NPs in the review texts are used to describe events from the story, with description or praise of authors as second most important category. But a systematic annotation effort would be necessary to validate this impression.

Table 8. Observed differences between positive and negative reviews

Key in positive reviews	Key in negative reviews
he, you	I
a	the
(third person) present tense	past tense
(proper) nouns	

This hypothesis is also consistent with the observed difference between the usage of the determiners 'a' (mostly with positive reviews) vs. 'the'. In general, 'a' is used to introduce new subjects into a text, while 'the' is used to refer to subjects already known. Table 9 shows nouns occurring after the definite or indefinite article, ordered by the fraction of occurrences after 'a(n)'. The table is split in two halves: Words above the bolded bar (under 'relationship') occur disproportionately more with 'the', the words below co-occur significantly more often with 'a(n)'. The words below the bar include the fixed expressions 'a bit' and 'a lot', but also general nouns that you would need to (begin to) recount a plot: 'a man', 'a girl', and many others. So, it makes sense that 'a(n)' should occur more in positive reviews. The words above the bar, especially the highly frequent ones, refer to objects that in the context of the review do not need to be introduced, and will therefore usually have 'the': 'the book', 'the story', 'the end', 'the author'.

If positive reviews more often recount events from the story, is there also something that negative reviews do more? We note that in our corpus, the negative reviews are on average longer than the positive reviews (mean of 123 compared to a mean of 106 words). What do they use these words for? One possibility is that negative reviewers feel more need to explain why they didn't like a book. An exploratory analysis using LIWC showed that negative reviews score significantly higher in all cognitive aspects (insight, cause, discrepancies, tentative language and differences), except certainty. These explanations may be regarded as a reviewer's description of their individual cognitive and emotional engagement in the reading experience, which may explain the higher use of 'I' in the negative reviews. The results also suggest that explanations predominantly use the past tense, which could explain the predominance of the past tense in negative reviews.

This is only the beginning of a more extensive analysis: there are many other aspects of these keyword lists that merit further investigation. Why does, for instance, 'and' occur more in positive reviews? An explanation could be that positive reviews use more groups of three words to describe (especially) characters: 'strong and determined and feisty', 'pretty and quirky and cool' or 'lost and lonely and unsure'. Another finding is that question marks are a very good predictor of negative reviews (and, confirming the findings of Taboada et al., these questions are indeed often sarcastic or rhetorical). We cannot go into these and other phenomena for reasons of space.

Table 9. Nouns occurring after 'a' or 'the', with the fraction of occurrences after 'a(n)' as part of the total number of occurrences.

Noun	Fraction of occurrences after 'a(n)'
rest	0,00
writing	0,01
beginning	0,01
middle	0,01
truth	0,02
end	0,02
fact	0,02
start	0,03
edge	0,04
cover	0,05
title	0,06
storyline	0,06
setting	0,06
author	0,07
past	0,07
ending	0,07
plot	0,08
action	0,09
future	0,10
idea	0,11
reader	0,12
story	0,12
heroine	0,14
world	0,16
case	0,17
point	0,18
book	0,19
series	0,19
hero	0,20
war	0,20
time	0,20
history	0,23
way	0,24
movie	0,25
novel	0,26
kind	0,29
reason	0,29
romance	0,30
moment	0,31
heart	0,31
mystery	0,34
relationship	0,36
life	0,38
love	0,38
family	0,38

Noun	Fraction of occurrences after 'a(n)'
day	0,39
read	0,41
problem	0,42
character	0,43
sequel	0,46
page	0,48
writer	0,54
place	0,55
girl	0,57
boy	0,57
year	0,59
person	0,60
man	0,62
sense	0,65
part	0,66
chance	0,69
woman	0,73
review	0,73
child	0,81
fun	0,88
couple	0,88
while	0,93
copy	0,97
fan	0,97
bit	0,99
lot	1,00

Conclusion

In this article, we looked at what an application of corpus-linguistic methods to online reviews can teach us about evaluation practices on online reviewing platforms. So far, there is no dominant research approach applied to online reviews, and up to now only a limited number of researchers have used corpus-based methods for their analysis. With some misgivings about its proprietary nature, we used Sketch Engine as the workbench with which to analyse them.

We looked at four questions about the online reviews. To answer the question whether online reviewers focus predominantly on their reading experience, bringing in their personal histories and other elements foreign to the book, we looked at a Word Sketch of 'plot'. We saw that online reviewers demonstrate a rich conceptualization of the dimensions of 'plot' and other key aspects of literary appreciation. In asking whether reviewers of literary books differ from reviewers of suspense books in what they consider beautiful, we used a word sketch difference. We found that suspense reviewers prefer character and plot (as well as the dimension of the books' materiality, the cover), literature reviewers prefer stylistic aspects and story. We also looked at the role of the

body in reading. From our analysis it emerged that positive reading experiences are often described in bodily terms, either literally or metaphorically. Finally, we investigated differences between positive and negative reviews. We concluded that positive reviews recount more events from the story, while negative reviews explain in cognitive terms, and from an individual perspective, why the book was a disappointment.

Taken together the findings show that online book reviews, when studied in large numbers, reveal patterns that do not appear at the level of the individual review. A corpus tool such as Sketch Engine provides a useful lens for studying these patterns; its useful tools include keyword computations and word sketches. For all of the questions that we discussed in this article, we have only been able to scratch the surface, but it is clear that online book reviews provide an important opportunity for research into the question how literature and fiction more widely can affect readers. To reuse the phrase that an anonymous Amazon reviewer used about Murakami, for the researcher these reviews can be 'a pretty sublime mix of WTF and OMG.'

WORKS CITED

Albrecht, Katharina. "Positioning BookTube in the Publishing World: An Examination of Online Book Reviewing through the Field Theory." *MSc Leiden University*, 2017, https://openaccess.leidenuniv.nl/handle/1887/52201.

Anthony, Laurence. "What Can Corpus Software Do?" *The Routledge Handbook of Corpus Linguistics*, Jan. 2022, pp. 103–25, https://doi.org/10.4324/9780367076399-9.

Archer, Dawn. *What's in a Word-List?: Investigating Word Frequency and Keyword Extraction*. Ashgate Publishing, Ltd, 2009.

Bachmann-Stein, Andrea. "Zur Praxis des Bewertens in Laienrezensionen." *Literaturkritik heute. Tendenzen–Traditionen–Vermittlung*, edited by Heinrich Kaulen and Christina Gansel, V&R Unipress, 2015, pp. 77–92, https://doi.org/10.14220/9783737002462.77.

Bachmann-Stein, Andrea, and Stephan Stein. "Demokratisierung der Literaturkritik im World Wide Web? Zum Wandel kommunikativer Praktiken am Beispiel von Laienrezensionen." *Musterwandel–Stilwandel. Aktuelle Tendenzen der diachronen Text (sorten) linguistik*, edited by Stefan Hauser et al., Sprache in Kommunikation und Medien, Peter Lang, 2014.

Benne, Christian. "Tolle lege. Embodied reading and the 'scene of reading.'" *Language Sciences*, vol. 84, Mar. 2021, p. 101357, https://doi.org/10.1016/j.langsci.2021.101357.

Biber, Douglas. "Corpus-Based and Corpus-Driven Analyses of Language Variation and Use." *The Oxford Handbook of Linguistic Analysis*, Oxford University Press, 2015.

Blank, Grant, and Bianca C. Reisdorf. "The Participatory Web: A User Perspective on Web 2.0." *Information, Communication & Society*, vol. 15, no. 4, May 2012, pp. 537–54, https://doi.org/10.1080/1369118x.2012.665935.

Bondi, Marina, and Mike Scott, editors. *Keyness in Texts*. John Benjamins Publishing Company, 2010, https://doi.org/10.1075/scl.41.

Boot, Peter. "A Database of Online Book Response and the Nature of the Literary Thriller." *Digital Humanities 2017*, 2017, https://zenodo.org/record/6615131.

---. "Towards a Genre Analysis of Online Book Discussion: Socializing, Participation and Publication in the Dutch Booksphere." *Selected Papers of Internet Research IR 12.0*, 2011, http://spir.aoir.org/papers/boot.pdf.

Boyd, Brian. *On the Origin of Stories: Evolution, Cognition, and Fiction*. Harvard University Press, 2009, https://doi.org/10.4159/9780674053595.

Caracciolo, Marco. "Degrees of Embodiment in Literary Reading." *Expressive Minds and Artistic Creations: Studies in Cognitive Poetics*, Oxford University Press, 2018, https://doi.org/10.1093/oso/9780190457747.003.0002.

Catharin, Leonardo Gabiato, and Valéria Delisandra Feltrim. "Finding Opinion Targets in News Comments and Book Reviews." *Lecture Notes in Computer Science*, 2018, pp. 375–84, https://doi.org/10.1007/978-3-319-99722-3_38.

Crossley, Scott A., et al. "Sentiment Analysis and Social Cognition Engine (SEANCE): An Automatic Tool for Sentiment, Social Cognition, and Social-Order Analysis." *Behavior Research Methods*, vol. 49, no. 3, May 2016, pp. 803–21, https://doi.org/10.3758/s13428-016-0743-z.

Daniels, M. J. "Als ik realisme wil ga ik wel een uur uit het raam staan kijken." *Een kwalitatieve en kwantitatieve analyse van Nederlandse lezersrecensies op Bol.com en Goodreads." MA Radboud Universiteit*, 2016.

de Smedt, Tom, and Walter Daelemans. "'Vreselijk mooi!' (terribly beautiful): A Subjectivity Lexicon for Dutch Adjectives." *Proceedings of the 8th Language Resources and Evaluation Conference (LREC'12)*, 2012.

Dimitrov, Stefan, et al. "Goodreads versus Amazon: The Effect of Decoupling Book Reviewing and Book Selling." *Ninth International AAAI Conference on Web and Social Media*, 2015.

Dörrich, Matthea. *Book Consumption in Convergence Culture: An Exploratory Audience Study of Media Repertoires of Book Consumption in the Tension between Participation and Corporate Control*. Stockholm University, 2014, http://www.diva-portal.org/smash/get/diva2:744217/FULLTEXT01.pdf.

Driscoll, Beth. "Book Blogs as Tastemakers." *Participations*, vol. 16, no. 1, 2019.

Driscoll, Beth, and DeNel Rehberg Sedo. "Faraway, So Close: Seeing the Intimacy in Goodreads Reviews." *Qualitative Inquiry*, vol. 25, no. 3, Sept. 2018, pp. 248–59, https://doi.org/10.1177/1077800418801375.

Ernst, Thomas. "›User Generated Content‹ und der Leser-Autor als ›Prosumer‹. Potenziale und Probleme der Literaturkritik in Sozialen Medien." *Literaturkritik heute*, edited by Heinrich Kaulen and Christina Gansel, V&R Unipress, 2015, pp. 93–112, https://doi.org/10.14220/9783737002462.93.

Gee, James Paul. "Affinity Spaces: How Young People Live and Learn on Line and out of School." *Phi Delta Kappan*, vol. 99, no. 6, Feb. 2018, pp. 8–13, https://doi.org/10.1177/0031721718762416.

---. *Situated Language and Learning: A Critique of Traditional Schooling*. Routledge, 2004.

Gibbs, Raymond W. "Embodied Dynamics in Literary Experience." *Cognitive Literary Science: Dialogues between Literature and Cognition*, Feb. 2017, pp. 219–38, https://doi.org/10.1093/acprof:oso/9780190496869.003.0012.

Gomide, Andressa Rodrigues. *Corpus Linguistics Software: Understanding Their Usages and Delivering Two New Tools*. Lancaster University, United Kingdom, 2020.

Goźdź-Roszkowski, Stanisław, and Susan Hunston. "Corpora and beyond – Investigating Evaluation in Discourse: Introduction to the Special Issue on Corpus Approaches to Evaluation." *Corpora*, vol. 11, no. 2, Aug. 2016, pp. 131–41, https://doi.org/10.3366/cor.2016.0089.

Gruzd, Anatoliy, and DeNel Rehberg Sedo. "#1b1t: Investigating Reading Practices at the Turn of the Twenty-First Century." *Articles*, vol. 3, no. 2, June 2012, https://doi.org/10.7202/1009347ar.

Harris, Susan K. "Whohoo!!! Joan of Arc!!!!!" *American Literary Realism*, vol. 51, no. 2, Jan. 2019, pp. 136–53, https://doi.org/10.5406/amerlitereal.51.2.0136.

Hegel, Allison. *Social Reading in the Digital Age*. 2018. UCLA, PhD, https://escholarship.org/uc/item/193451g2.

Herrmann, Berenike, and Thomas C. Messerli. "hungere schon nach dem nächsten Band. Eine Untersuchung von Metaphern für Leseerfahrungen in Web 2.0 Literaturrezensionen." *DHd2020 Conference*, 2020.

Heydebrand, Renate Von, and Simone Winko. *Einführung in die Wertung von Literatur: Systematik, Geschichte, Legitimation*. Schöningh, 1996.

Hogan, Bernie. "The Presentation of Self in the Age of Social Media: Distinguishing Performances and Exhibitions Online." *Bulletin of Science, Technology & Society*, vol. 30, no. 6, Nov. 2010, pp. 377–86, https://doi.org/10.1177/0270467610385893.

Jaakkola, Maarit. "From Re-Viewers to Me-Viewers: The #Bookstagram Review Sphere on Instagram and the Uses of the Perceived Platform and Genre Affordances." *Interactions: Studies in Communication & Culture*, vol. 10, no. 1–2, 2019.

Jaworska, Sylvia. "Corpus Approaches." *The Routledge Handbook of Language and Media*, Aug. 2017, pp. 93–108, https://doi.org/10.4324/9781315673134-8.

Jerasa, Sarah, and Trevor Boffone. "BookTok 101: TikTok, Digital Literacies, and Out-of-School Reading Practices." *Journal of Adolescent & Adult Literacy*, vol. 65, no. 3, Oct. 2021, pp. 219–26, https://doi.org/10.1002/jaal.1199.

Jessen, A. F. N. *Lezen als sociale activiteit: van leesgezelschap tot online lezerscommunity*. Master Radboud University Nijmegen, 2016.

Kilgarriff, Adam. "Simple Maths for Keywords." *Corpus Linguistics Conference*, 2009.

---. "The Sketch Engine: Ten Years On." *Lexicography*, vol. 1, no. 1, July 2014, pp. 7–36, https://doi.org/10.1007/s40607-014-0009-9.

Kilgarriff, Adam, and David Tugwell. "Word Sketch: Extraction and Display of Significant Collocations for Lexicography." *Proc. ⬦ ⬦Collocations⬦ ⬦workshop ,⬦ ⬦ACL⬦ ⬦2001*, 2001, pp. 32–38.

Knoop, Christine A., et al. "Mapping the Aesthetic Space of Literature 'from Below.'" *Poetics*, vol. 56, June 2016, pp. 35–49, https://doi.org/10.1016/j.poetic.2016.02.001.

Koolen, Marijn, et al. "Online Book Reviews and the Computational Modelling of Reading Impact." *Computational Humanities Research*, 2020.

Koopman, Eva Maria (Emy), and Frank Hakemulder. "Effects of Literature on Empathy and Self-Reflection: A Theoretical-Empirical Framework." *Journal of Literary Theory*, vol. 9, no. 1, Jan. 2015, https://doi.org/10.1515/jlt-2015-0005.

Kutzner, Kristin et al. "Exploring the content composition of online book reviews." *Workshop InfDH Methoden und Anwendungen der Computational Humanities*, 2020, p. 3.

Lexical Computing Ltd. *Statistics Used in Sketch Engine*. Lexical Computing Ltd, 2015, https://www.sketchengine.eu/wp-content/uploads/ske-statistics.pdf.

Lukoschek, Katharina. "'Ich liebe den Austausch mit euch'. Austausch über und anhand von Literatur in Social Reading-Communities und auf Bücherblogs." *Die Rezension. Aktuelle Tendenzen der Literaturkritik*, edited by Andrea Bartl and Markus Behmer, Köningshausen und Neumann, 2017.

McDonald, Rónán. *The Death of the Critic*. Continuum International Publishing Group, 2007.

McEnery, Tony, and Andrew Hardie. *Corpus Linguistics: Method, Theory and Practice*. Cambridge University Press, 2011, https://doi.org/10.1017/cbo9780511981395.

Menninghaus, Winfried, et al. "The Emotional and Aesthetic Powers of Parallelistic Diction." *Poetics*, vol. 63, Aug. 2017, pp. 47–59, https://doi.org/10.1016/j.poetic.2016.12.001.

Miller, Christopher R. "Being and Feeling: The Surprise Attacks of Paradise Lost." *Surprise. The Poetics of the Unexpected from Milton to Austen*, Cornell University Press, 2015, https://doi.org/10.7591/cornell/9780801453694.001.0001.

Miller, Donna L. *Talking with Our Fingertips: An Analysis for Habits of Mind in Blogs about Young Adult Books*. 2011. Arizona State University, PhD, http://repository.asu.edu/attachments/56903/content/Miller_asu_0010E_10883.pdf.

Moreton, Emma. "'I Never Could Forget My Darling Mother': The Language of Recollection in a Corpus of Female Irish Emigrant Correspondence." *The History of the Family*, vol. 21, no. 3, Mar. 2016, pp. 315–36, https://doi.org/10.1080/1081602x.2016.1155469.

Morgan, Ben. "Embodied Cognition and the Project of the Bildungsroman: *Wilhelm Meister's Apprenticeship* and *Daniel Deronda*." *Poetics Today*, vol. 38, no. 2, June 2017, pp. 341–62, https://doi.org/10.1215/03335372-3869287.

Mpouli, Suzanne, and Jean-Gabriel Ganascia. "Towards the Automatic Mining of Similes in Literary Texts." *Digital Humanities 2016. From Digitization to Knowledge 2016: Resources and Methods for Semantic Processing of Digital Works/Texts, Proceedings of the Workshop*, 2016.

Murray, Simone. "Secret Agents: Algorithmic Culture, Goodreads and Datafication of the Contemporary Book World." *European Journal of Cultural Studies*, vol. 24, no. 4, Dec. 2019, pp. 970–89, https://doi.org/10.1177/1367549419886026.

Nakamura, Lisa. "'Words with Friends': Socially Networked Reading on *Goodreads*." *PMLA*, vol. 128, no. 1, Jan. 2013, pp. 238–43, https://doi.org/10.1632/pmla.2013.128.1.238.

Neuhaus, Stefan. "»Leeres, auf Intellektualität zielendes Abrakadabra«. Veränderungen von Literaturkritik und Literaturrezeption im 21. Jahrhundert." *Literaturkritik Heute*, edited by Heinrich Kaulen and Christina Gansel, V&R Unipres, 2015, pp. 43–58, https://doi.org/10.14220/9783737002462.43.

Newell, Edward Daniel, et al. "To Buy or to Read: How a Platform Shapes Reviewing Behavior." *Tenth International AAAI Conference on Web and Social Media*, 2016.

Ni, Jianmo, et al. "Justifying Recommendations Using Distantly-Labeled Reviews and Fine-Grained Aspects." *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 2019, https://doi.org/10.18653/v1/d19-1018.

Nuttall, Louise, and Chloe Harrison. "Wolfing down the Twilight Series: Metaphors for Reading in Online Reviews." *Contemporary Media Stylistics*, Bloomsbury Academic, 2020, https://doi.org/10.5040/9781350064119.0007.

Pennebaker, J. W., et al. *The Development and Psychometric Properties of LIWC2015*. University of Texas at Austin, 2015, http://hdl.handle.net/2152/31333.

Pinch, Trevor, and Filip Kesler. "How Aunt Ammy Gets Her Free Lunch: A Study of the Top-Thousand Customer Reviewers at Amazon. Com." *Managing Overflow in Affluent Societies*, 2011, https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.474.3695&rep=rep1&type=pdf.

Ramsay, Stephen. "Who's in and Who's out." *Defining Digital Humanities. A Rfeader*, Ashgate, 2013.

Rebora, Simone, et al. "Digital Humanities and Digital Social Reading." *Digital Scholarship in the Humanities*, vol. 36, no. Supplement_2, Oct. 2021, pp. ii230–50, https://doi.org/10.1093/llc/fqab020.

Rehfeldt, Martin. "Leserrezensionen als Rezeptionsdokumente. Zum Nutzen nicht-professioneller Literaturkritiken für die Literaturwissenschaft." *Die Rezension. Aktuelle Tendenzen der Literaturkritik*, edited by Andrea Bartl and Markus Behmer, Köningshausen und Neumann, 2017.

Riddell, Allen, and Karina van Dalen-Oskam. "Readers and Their Roles: Evidence from Readers of Contemporary Fiction in the Netherlands." *PloS One*, vol. 13, no. 7, July 2018, p. e0201157, https://doi.org/10.1371/journal.pone.0201157.

Ross, Catherine Sheldrick. "Metaphors of Reading." *Journal of Library History*, 1987.

Salgaro, Massimo, and Simone Rebora. "Measuring the 'Critical Distance.'" *A Corpus-Based Analysis of Italian Book Reviews*, AIUCD, 2018.

Sutton, Kim Maya, and Ina Paulfeuerborn. "The Influence of Book Blogs on the Buying Decisions of German Readers." *Logos. Journal of the World Book Community*, vol. 28, no. 1, June 2017, pp. 45–52, https://doi.org/10.1163/1878-4712-11112124.

Taboada, Maite, et al. "On Being Negative." *Corpus Pragmatics*, vol. 1, no. 1, Mar. 2017, pp. 57–76, https://doi.org/10.1007/s41701-017-0006-y.

Thelwall, Mike, and Kayvan Kousha. "Goodreads: A Social Network Site for Book Readers." *Journal of the Association for Information Science and Technology*, vol. 68, no. 4, Dec. 2016, pp. 972–83, https://doi.org/10.1002/asi.23733.

Tian, Q. *Self-Presentation and Social Interaction on Blogs: A Structural Equation Modeling of the Uses and Gratifications of Blogging*. 2009. PhD, http://digitalarchive.gsu.edu/communication_diss/15/.

Tiberius, Caronline, and Adam Kilgarriff. "The Sketch Engine for Dutch with the ANW Corpus." *Fons Verborum, Leiden, Instituut Voor Nederlandse Lexicologie/Gopher*, 2009.

Tobin, Vera. *Elements of Surprise: Our Mental Limits and the Satisfactions of Plot*. Harvard University Press, 2018, https://doi.org/10.4159/9780674919570.

Tognini-Bonelli, Elena. *Corpus Linguistics at Work*. John Benjamins Publishing, 2001, https://doi.org/10.1075/scl.6.

Vlieghe, Joachim, et al. "Everybody Reads: Reader Engagement with Literature in Social Media Environments." *Poetics*, vol. 54, Feb. 2016, pp. 25–37, https://doi.org/10.1016/j.poetic.2015.09.001.

Walsh, Melanie, and Maria Antoniak. "The Goodreads 'Classics': A Computational Study of Readers, Amazon, and Crowdsourced Amateur Criticism." *Journal of Cultural Analytics*, vol. 6, no. 2, Apr. 2021, https://doi.org/10.22148/001c.22221.

Wan, Mengting, et al. "Fine-Grained Spoiler Detection from Large-Scale Review Corpora." *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2019, https://doi.org/10.18653/v1/p19-1248.

Zhang, Chengzhi, et al. "Examining Differences among Book Reviews from Various Online Platforms." *Online Information Review*, vol. 43, no. 7, Nov. 2019, pp. 1169–87, https://doi.org/10.1108/oir-01-2019-0037.