# Digital humanities and digital social reading

**Simone Rebora** (iD)

Department of Foreign Languages and Literatures, University of Verona, Italy; Digital Humanities Lab, University of Basel, Switzerland

**Peter Boot** (iD) *

Huygens Institute for the History of the Netherlands, Amsterdam, Netherlands

**Federico Pianzola** (iD) *

Department of Human Sciences for Education, University of Milan-Bicocca, Italy; Sogang University, Seoul, South Korea

**Brigitte Gasser** (iD)

Digital Humanities Lab, University of Basel, Switzerland

**J. Berenike Herrmann** (iD)

Digital Humanities Lab, University of Basel, Switzerland; Faculty of Linguistics and Literary Studies, University of Bielefeld, Germany

**Maria Kraxenberger** (iD)

Digital Humanities Lab, University of Basel, Switzerland

**Moniek M. Kuijpers** (iD)

Digital Humanities Lab, University of Basel, Switzerland

**Gerhard Lauer** (iD)

Digital Humanities Lab, University of Basel, Switzerland

**Piroska Lendvai** (iD)

Digital Humanities Lab, University of Basel, Switzerland

**Thomas C. Messerli** (iD)

Digital Humanities Lab, University of Basel, Switzerland

**Pasqualina Sorrentino** (iD)

Department of German Philology, University of Göttingen, Germany

**Correspondence:**
Simone Rebora.
**E-mail:**
simone.rebora@univr.it
*These authors contributed equally to this work.

## Abstract

Prominent among the social developments that the web 2.0 has facilitated is digital social reading (DSR): on many platforms there are functionalities for creating book

reviews, 'inline' commenting on book texts, online story writing (often in the form of fanfiction), informal book discussions, book vlogs, and more. In this article, we argue that DSR offers unique possibilities for research into literature, reading, the impact of reading and literary communication. We also claim that in this context computational tools are especially relevant, making DSR a field particularly suitable for the application of Digital Humanities methods. We draw up an initial categorization of research aspects of DSR and briefly examine literature for each category. We distinguish between studies on DSR that use it as a lens to study wider processes of literary exchange as opposed to studies for which the DSR culture is a phenomenon interesting in its own right. Via seven examples of DSR research, we discuss the chosen approaches and their connection to research questions in literary studies.

## 1 Introduction

Over the last decades, with growing digitization and the fast-paced development of social media platforms, reading has become a more socially interactive experience than ever before, in which the Internet plays a key role. Platforms such as Goodreads, LovelyBooks, and Wattpad are online environments where millions of people from all over the world share their love for the written word. Members discuss what they read and what they judge as good or bad literature, they recommend books to one another, and try their hand at writing fiction. In the research community, this phenomenon has been labelled in many different ways (online book discussions, online reading and writing, (online) social reading, etc.). In our study, we propose the term digital social reading (DSR) for shared reading experiences which happen either online or offline but involve some use of digital technology and media, either for reading or for sharing experiences elicited by books. While this label disregards some key aspects of the phenomenon (e.g., the extensive writing activity in DSR communities), it still catches the determinant role of social interactions around the experience of reading, which are visible through DSR practices and platforms. Readers on the Web are increasingly becoming 'wreaders' (Landow, 2006), and scholars of literature are starting to recognize their centrality in the global system of literary production (Miall, 2018).

One of the first publications exploring the extent of DSR is by Leveratto and Leontsini (2008), who noted how the Internet has enabled a whole range of social interactions revolving around reading. After a series of articles that highlighted the relevance of DSR for literary studies (Schreier, 2010; Boot, 2011; Nakamura, 2013), the first extensive survey was accomplished by Cordón-García et al. (2013), who described 'social reading' by highlighting the increased relevance of readers and even proposing a connection with the 'Gutenberg parenthesis' theory, which sees print books as just a phase between ancient and modern (or digital) forms of orality. More recently, Murray (2018b) coined the term 'digital literary sphere', referring to Genette's concept of paratext (Genette, 1987) to locate its characteristic niche, generally 'in the (digital) margins' of books. In national contexts, beyond anglophone countries (e.g., Finn, 2013; Barnett, 2015; Thomas, 2020), DSR has received specific attention in a few other contexts, such as Italy (Faggiolani and Vivarelli, 2016), Germany (Bartl and Behmer, 2017), and Spanish speaking countries (Cruces Villalobos, 2017; Centro National de Innovación e Investigación Educativa, 2019; Cordón-García and Gómez-Díaz, 2019).

The importance of online book discussion for reception studies was argued by Montesi (2015), who discusses how social reading sites can show the impact of books – on readers individually as well as on society at large. Reading platforms and readers themselves are also relevant objects of study (i.e., their personal libraries and their social relations). Rehfeldt (2017) rejects the tendency of researchers to consider online book reviews a defective version of literary criticism (cf. Hugendick, 2008) and argues that lay reviews are better than professional reviews in showing the effect of the book on the reader, since users feel no need to be objective.

In the remainder of this article, we will discuss the state of the art of DSR research by referring to seven current case studies. In the first part, we propose a categorization of DSR research by identifying ten dominant categories; for each category, we discuss the disciplines or fields which are studying it (or may find useful to study it). In the second part, we present seven case studies conducted by our research team in correspondence with the categorization. Together, the case studies highlight the vital role that Digital Humanities can and should play in the study of DSR.

## 2 Categorizing DSR Research

Several taxonomies have been proposed for DSR. The first was by Stein (2010), who identifies four defining dichotomies in book discussions, which can be used to categorize types of DSR: online vs. offline; synchronous vs. asynchronous; formal vs. informal; and ephemeral vs. persistent. A more practical taxonomy, from the perspective of literary criticism, was given by Ernst (2015), who focuses on online literary criticism, distinguishing between online presence of print media, born online individual criticism (such as blogs), and social media-based criticism, which he, in turn, divides into multiple categories. Much more fine-grained is the taxonomy by Kutzner *et al.* (2019), which takes into consideration a total of fifteen dimensions, from the cultural artefact (print book, e-book, audiobook, etc.), to the presence/absence of off-topic communication, the type of author/reviewer gratification, and many others. These taxonomies can account for most of the practices and platforms that have emerged in recent years. However, one of their main limitations is that they use a purely descriptive approach that misses some important dimensions of social reading: first, the impact that DSR has on the wider cultural and social context; second, the disciplines involved in studying these aspects. To fill this gap, we propose a categorization that groups the studies on DSR into ten different categories that reflect the most relevant aspects of DSR.

For each category, we identify the scholarly discipline or field with which it is associated (Fig. 1). Research in each category can either be on DSR itself or use DSR as a lens to study wider reading practices.

Some categories will lend themselves more easily than others to this type of generalization: our estimate of this generalizability is represented by a position further away from the centre of the figure. As our main focus is on literary studies, we ignore research that uses or investigates DSR from the point of view of information technology (e.g. Tang *et al.*, 2014), or legal issues such as copyright or privacy (e.g. Shipman and Marshall, 2013). We also acknowledge that historiographic perspectives are frequently implied in the different categories. However, as DSR is a new and growing phenomenon, we do not yet see historiography as a category per se. Due to space constraints, our review is necessarily limited, but we also compiled a broader public Zotero bibliography (Pianzola *et al.*, 2019).

With reading-oriented research, we mean research that studies the process, experience, and impact of reading. The focus may be on the effects of the reading medium (paper, e-book); the research may use reviews and comments on texts to study reading processes, or it may differentiate among these processes by (genre of) book, time period, or author. However, in reading-oriented research, the focus is on the act of reading itself and not on the interaction among readers, wider social implications, or the digital reading platforms. The real strength of reading-oriented DSR research is in unprecedented access to the reader's experience. Driscoll and Rehberg Sedo (2018), for instance, investigate reading experiences in reviews on Goodreads, manually coding for experiential language as well as applying automated sentiment analysis. It is a good example of the multi-method approach that is often fruitfully applied in studying reading-oriented DSR. Manual coding brought out the different emotional registers that the reviews employ, whereas sentiment analysis was used for analysis at a larger scale. The authors conclude that the intimate experience of reading, formerly elusive to research, to some extent becomes visible on platforms such as Goodreads. While such statements do not acknowledge the achievements in historical reader response research (e.g., via the study of letters and diaries), they highlight how the wide availability of reading experience testimonies in DSR inevitably opens new perspectives for the research. Similar analyses have been done with respect to what readers value in a text (Milota, 2014), to metaphors for reading (Nuttall and
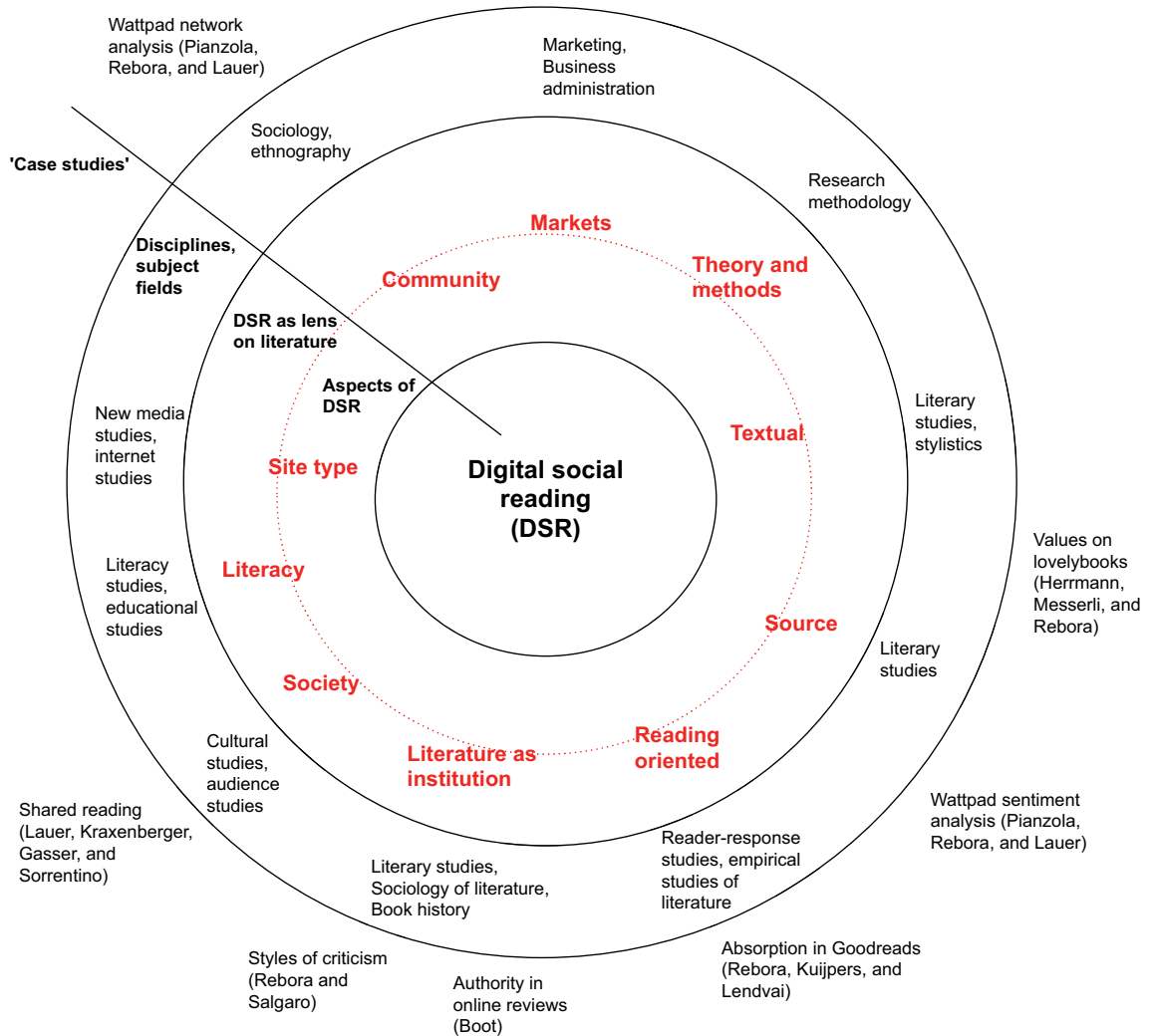
**Fig. 1** Aspects of DSR (in red) with the relevant disciplines, highlighting the interdisciplinary nature of DSR research.

Harrison, 2020), or the ethical positions that readers take in processing a controversial book such as *We Need to Talk about Kevin* (Nuttall, 2017). We expect growth to occur especially in the fields of empirical literary studies, cognitive poetics, and reading research in general, when it comes to this area of DSR research.

Under literature as an institution we group research that considers online book discussion as a form of literary criticism or gatekeeping, looking at its role in the literary field and its relation to other actors. Underlying differences in literary values often

play a part in these investigations. Allington (2016), for instance, compares both types of review for Desai's *The Inheritance of Loss*, manually coding several aspects of evaluation and some political variables in the reviews, and then quantitatively analysing the results. He finds, among other things, that user reviews are much more negative than professional reviews, probably because the book is targeted at a literary rather than a popular audience.

Rather than comparing reviews, Verboord (2010) asks readers directly whether they trust what he calls

'expert' or 'internet' critics. He finds that people with an 'omnivorous' taste in books ('persons combining [...] "highbrow" tastes with "middlebrow" or "popular" tastes') have less confidence in expert critics. Stein (2015) discusses 'lay' literary criticism as a 'communicative practice in the literary system', and notes its tendency to be intellectually less demanding and therefore perhaps favouring less demanding literature. Johnson (2016) on the other hand studies US book blogs as what she calls 'the new gatekeepers', and positively appreciates book bloggers' attention towards more popular books. Part of this kind of research has a broader perspective including historical, societal, and economical reflections, mostly intersecting the field of book history (Murray, 2018b). Murray (2018a) has also suggested that book historians need to reinvent their discipline radically if it should be able to account for the current changes in reading habits.

With reading-oriented research, we mean research that studies the process, experience, and impact of reading. The focus may be on the effects of the reading medium (paper, e-book); the research may use reviews and comments on texts to study reading processes, or it may differentiate among these processes by (genre of) book, time period, or author. However, in reading-oriented research, the focus is on the act of reading itself and not on the interaction among readers, wider social implications, or the digital reading platforms. The real strength of reading-oriented DSR research is in unprecedented access to the reader's experience. Driscoll and Rehberg Sedo (2018), for instance, investigate reading experiences in reviews on Goodreads, manually coding for experiential language as well as applying automated sentiment analysis. It is a good example of the multi-method approach that is often fruitfully applied in studying reading-oriented DSR. Manual coding brought out the different emotional registers that the reviews employ, whereas sentiment analysis was used for analysis at a larger scale. The authors conclude that the intimate experience of reading, formerly elusive to research, to some extent becomes visible on platforms such as Goodreads. While such statements do not acknowledge the achievements in historical reader response research (e.g. via the study of letters and diaries), they highlight how the wide availability of reading experience testimonies in DSR inevitably opens new perspectives for the research. Similar analyses have been done with

respect to what readers value in a text (Milota, 2014), to metaphors for reading (Nuttall and Harrison, 2020), or the ethical positions that readers take in processing a controversial book such as *We Need to Talk about Kevin* (Nuttall, 2017). We expect growth to occur especially in the fields of empirical literary studies, cognitive poetics, and reading research in general, when it comes to this area of DSR research.

Under literature as an institution we group research that considers online book discussion as a form of literary criticism or gatekeeping, looking at its role in the literary field and its relation to other actors. Underlying differences in literary values often play a part in these investigations. Allington (2016), for instance, compares both types of review for Desai's *The Inheritance of Loss*, manually coding several aspects of evaluation and some political variables in the reviews, and then quantitatively analysing the results. He finds, among other things, that user reviews are much more negative than professional reviews, probably because the book is targeted at a literary rather than a popular audience.

Rather than comparing reviews, Verboord (2010) asks readers directly whether they trust what he calls 'expert' or 'internet' critics. He finds that people with an 'omnivorous' taste in books ('persons combining [...] "highbrow" tastes with "middlebrow" or "popular" tastes') have less confidence in expert critics. Stein (2015) discusses 'lay' literary criticism as a 'communicative practice in the literary system', and notes its tendency to be intellectually less demanding and therefore perhaps favouring less demanding literature. Johnson (2016) on the other hand studies US book blogs as what she calls 'the new gatekeepers', and positively appreciates book bloggers' attention towards more popular books. Part of this kind of research has a broader perspective including historical, societal, and economical reflections, mostly intersecting the field of book history (Murray, 2018b). Murray (2018a) has also suggested that book historians need to reinvent their discipline radically if it should be able to account for the current changes in reading habits.

Research focusing on society looks at larger social issues that DSR may exemplify or contest, such as (in)equality, participation, democracy, feminism, and inclusiveness. It also includes research that sees readers as an audience that may be passive, resistant, or that would rather request an active role, as in

discussions of reviewers as 'prosumers' (Toffler, 1980) or 'producers' (Bruns, 2008). Dörrich (2014) investigates audience rebellion in the realm of the book, by interviewing LovelyBooks users, noting that there will always be a tension between corporate control and consumer participation, expressed for instance in some users' concern for ownership of data and privacy (see also Albrechtslund, 2019). Steiner (2008), too, expresses doubts about the democratizing potential of the Internet, which could be 'the worst kind of fraud, since it makes people believe they have the power to influence the public sphere, when in reality the web is only another way for capital to profit.' In the context of fan fiction studies, researchers have generally taken a more positive approach towards users' active role in online writing. Pugh programmatically called her book on fan fiction *The democratic genre* (2005), mentioning among other things that readers become co-creators and consumers become less passive. Fan fiction scholars have also stressed the feminist character of much work in fan fiction (e.g. Leow, 2011). In general, however, it is fair to say that in other domains (such as news production) web 2.0 has had a heavier impact on society than in the domain of reading and writing.

With research focusing on literacy we move from a literary to an educational viewpoint. Literacy-oriented research considers DSR mostly as a tool for education in reading, writing, literature, and personal development, including uses of DSR in library and classroom environments (Blyth, 2014; Kalir *et al.*, 2020). Indeed, some of the earliest research came from digital library studies. Kaplan and Chisik (2005), for example, use a process of participatory design to create a digital book prototype in which young readers could interact through annotations. They motivate their research explicitly by the desire 'to preserve the values we perceive in the notions of reading for pleasure' (p. 8). Later attempts to get readers to discuss books moved online, into specifically created book clubs (AuYeung *et al.*, 2007) or existing platforms such as Goodreads (Thompson, 2010; Merga, 2015). For example, Miller's thesis (2011) investigates whether blogging about young adult literature influences adolescent literacy development. Moving from reading response to creative writing, Korobkova's thesis (2017) investigates affordances for literacy development built into Wattpad. One of the most important conclusions is

that on these sites users 'gain self-efficacy and a positive disposition toward literacy as a result' (p. 102). Affinity, authenticity, and affect are what motivate their involvement on these sites. Korobkova also notes that not all users are equal, they need 'differential routes to participation and success' (p. 152), a point echoed by Taddeo (2019).

Research in the community category looks at the interaction between users on DSR platforms and specific platform cultures, be it with ethnographic methods, network analysis tools, or other methods. For example, Rehberg Sedo (2011) uses participatory observation methods to study an online group of professionals (teachers, publishers) discussing young adult books. As in face to face book clubs, discussions are influenced by the authority recognized by members on the basis of their cultural capital. In these online affinity spaces, readers act not so much as independent agents but rather as members who learned strategies that allow them to be part of a community. The importance of community is also stressed by Lukoschek (2017). The need for exchange between like-minded readers often crosses the boundaries between individual communities: the same people who have book blogs also meet each other on Facebook, Twitter, LovelyBooks, and elsewhere. In a landmark study on Goodreads, Thelwall and Kousha (2017) investigated (among other things) the relative importance of the social and book-related features of the site. They concluded that 'Goodreads seems to be a book-based social navigation Social Networking Site (SNS) rather than being primarily either a book website or a general SNS' (p. 981). Book-based discussion sites can also be conceived as 'boundary objects' that enable the establishment of community and structure (Worrall, 2019), processes in which moderators often play a crucial role (Thomas and Round, 2016).

With the market label, we refer to studies that consider the relevance of DSR platforms, texts, and participants for commercial purposes. This is the focus of the work by Sutton and Paulfeuerborn (2017), who evaluate the impact of book blogs on the (German) market through an online survey, producing a purchase decision model that might be beneficial for both publishers and bloggers. Much more critical is the approach by Moody (2017), who emphasizes how market needs can support practices such as sabotaging and bullying, which are generally overlooked by

research focused solely on the positive aspects of participation. The complexity of this context is confirmed by Murray (2016), examining the ways in which readers' evaluations on Amazon and practices like book trailers and blog tours are drastically transforming marketing strategies. Traditional methodological frameworks might prove inadequate to study them, if not supported by an understanding of the algorithms that are used to filter and aggregate readers' evaluations, or of the digital environments where they flourish. This implicit call for DH methods finds only partial realization, such as the study by Faggiolani *et al.* (2018), who adopt network analysis to visualize the relationships between Italian publishers on the DSR platform aNobii. Much work still needs to be done on this aspect of DSR. A combination of marketing research and DH methods might throw new light on its internal dynamics.

Textual-oriented DSR research is another category where DH can play a key role, in particular through computational linguistics and stylometry. This category of research is mostly interested in textual features characteristic of DSR platforms, such as style and wording. Inevitably, it has strong connections with the 'literature as institution' category, as the identification of a distinctive style generally derives from the confrontation with a model. Harada and Yamashita (2010) do precisely this, comparing online book reviews to reviews in newspapers. The focus here is not on the possible effects on traditional criticism, but rather on what distinguishes DSR *per se*. In Germany, Neuhaus (2017) points to distinguishing elements such as the lower quality of writing, the absence of specialized language, and frequent references to the I. Using also computational approaches, Mehling *et al.* (2018) identify the dominance of emotions, suspense, and enjoyment in the evaluation of books. In the English context, Hajibayova (2019) uses the LIWC software (Tausczik and Pennebaker, 2010) and manual annotation to devise a model for the language of Goodreads reviews.

With the source category, we refer to research that is most interested in what DSR activities say about the text that they comment on. Often this research uses reviews as a way to highlight a possible reception or interpretation of a work or genre, with a focus on the received work, not on the recipient. One clear example is the work of Gutjahr (2002), which focuses on the Christian book series *Left Behind*. Through analyses of Amazon reviews and interviews with readers, Gutjahr investigates the reasons for the success of the series, suggesting how it puts into question the very distinction between literary fiction and sacred texts. However, one of the most representative cases for this category is the research on Jane Austen's novels. While statistics on DSR platforms like Wattpad confirm that *Pride and Prejudice* is the most read (and most commented) classic among contemporary teenagers (Rebora and Pianzola, 2018), studies like that of Mirmohamadi (2014) investigate the 'digital afterlives' of the British author, focusing both on the reading/commenting activities and on the creative reinterpretations of fan fiction. Subjects can also be successful novels like *Gomorra* in Italy (Brugnatelli and Faggiolani, 2016) or more generally disregarded titles like the *Personal Recollections of Joan of Arc* by Marc Twain (Harris, 2019). In all cases, the main goal of these studies is to show how DSR can constructively contribute to literary criticism.

A large number of studies can be collected under the site type category. With this term we refer to research that describes the working logic and functionalities of one or more platforms, generally focusing on a single aspect (e.g. the reviews), without necessarily drawing conclusions about other aspects. One of the first examples is the work by Nakamura (2013), who provided a brief introduction to the then-understudied Goodreads platform. In a similar way, studies on platforms like LibraryThing (Pinder, 2012) and on phenomena like 'bookstagram' (book reviews on the Instagram platform, cf. Jaakkola, 2019) stimulated the interest of the research community towards DSR practices. The importance of such studies is undeniable, especially when they provide ample overviews (e.g. Cordón-García *et al.*, 2013, pp. 167–89; Cruces Villalobos, 2017).

To conclude our categorization, the studies grouped in the theory and method category focus both on the methodological needs of DSR research and on the theoretical impact it can have on disciplines such as book history and the history of reading. Among the first to highlight the possible relevance of the phenomenon, Maryl (2008) analysed reader responses on the Polish platform biblioNETka with the main goal of understanding if and how they can be useful for reading research. His conclusion was

mainly positive, but with awareness of all the risks and limitations that come with the analysis of such material (i.e. frequently noisy, unstructured, and unreliable). Bridle (2010), referring to Benjamin's philosophy, proposes a re-conceptualization of the 'aura' of books (shifting from paper's physicality to the text itself); Costa (2016) tries to re-define a phenomenology of reading, where the commenting and rewriting activities become an essential part of reading itself. While perspectives are generally positive and stimulating, Rowberry (2019) makes a relevant critical note about the future of theory and data-driven DSR research, adopting software criticism to highlight the inability of modern e-book technologies to provide relevant data for the study of reading.

## 3 Case studies

As shown by our overview, multimethodology is one of the main characteristics of DSR research. Methodological richness can sustain its development, but it might also hinder its coherent evolution, if no

disciplinary framework or central research field is identified. We propose that DH may provide this bond in at least two ways. First, like other subjects in the humanities, it can provide the tools for structuring and interconnecting the entire research field (cf. Es et al., 2018; Herrmann et al., 2019). Figure 2, for example, shows how a simple combination between a digital bibliography and network technologies can provide an efficient visualization of our categorization, highlighting the connections between categories. Second, and more importantly, its research interests and methodologies are epistemologically coherent with the goals of DSR research, as the seven case studies presented in this section will demonstrate.

In what follows, we discuss research conducted by our group and pertinent to the multi-methodological approaches to DSR that DH can provide. Inevitably, not all relevant aspects will be covered here. One important dimension is for example the historical perspective, working with retro-digitized materials and examining diachronic developments through time in terms of continuity and rupture (see, e.g., Chang et al.,
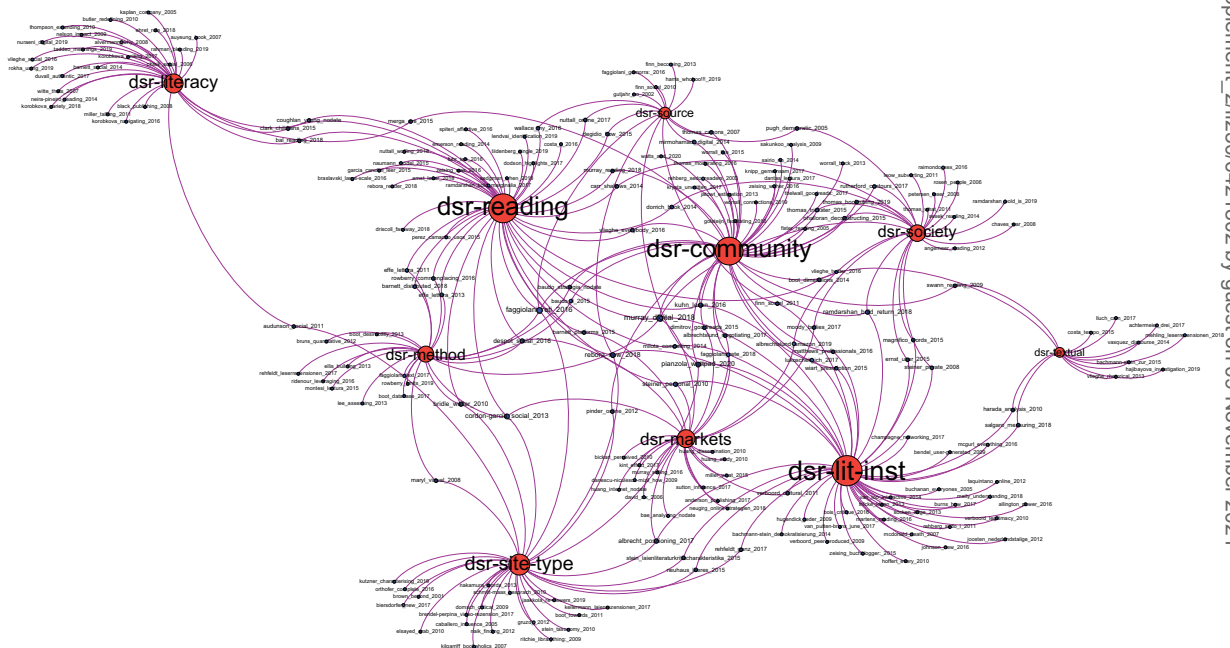


**Fig. 2** Network based on the DSR Zotero library (Pianzola et al., 2019). Edges represent connections between texts and categories in our categorization.

S. Rebora *et al.*

2020 on the relation between genre and book reviews). Each case study will be presented by following a tripartite structure: first, presentation of the research question that needs to be answered in DSR studies; second, introduction of the DH methodology that can be applied to it; third, discussion of the obtained results or of the encountered issues. Overall, the seven case studies will offer an exemplification of how fruitful the integration between DH methods and DSR research can be. Additionally, they will also provide a series of insights into the practical aspects of such a research.

## 3.1 Wattpad—network analysis (Pianzola, Rebora, and Lauer)

The first of our case studies explores Wattpad, the most popular platform for reading and commenting on fiction. It offers millions of stories written in more than thirty languages ranging over different genres, including literary classics, fan fiction, and original fiction. It is mostly accessed via smartphone and the average audience is between 12 and 25 years old. Previous research has mostly focused on the identity and activity of authors (Mirmohamadi, 2014; Ramdarshan Bold, 2018) and the possible educational applications of Wattpad (Korobkova and Rafalow, 2016; Taddeo, 2019). Using digital methods, namely network analysis, we were able to focus on readers and analyse the comments written by 300,000 users in the margins of twelve English novels. In this way, we reconstructed the network of social interactions related to reading Classics and Teen Fiction (Pianzola *et al.*, 2020). The goal was to see whether there is any difference in how teenagers read different genres socially on Wattpad. Given the huge quantity of data available—in the form of comments linked to the respective paragraphs/chapters/books and to the replies by other readers—visualizing the networks of interactions helped us to select users and comments that we wanted to observe more closely. Mixing distant reading with close reading of comments, we discovered that when the linguistic and cultural complexity of texts increases (Classics), readers tend to interact more, helping each other to understand the writing style and the historical context of the novel. However, with teen fiction stronger and more prolonged interactions between readers emerge, even

extending across different novels. Therefore, Wattpad can be considered both a community of peer learners and a social bonding tool, aspects that can be leveraged by educational projects that aim at promoting reading. In general, users comment much more actively on teen fiction novels than classics, confirming that Wattpad is a platform mainly used to read original stories written by teenagers (Contreras *et al.*, 2015; Taddeo, 2019).

## 3.2 Wattpad—sentiment analysis (Pianzola, Rebora, and Lauer)

In a second case study on Wattpad commenting practices, we looked at how comments in the margins of paragraphs enable us to investigate the progression of readers' response to a story, linking the verbalization of aesthetic, cognitive, and emotional reactions to specific text passages (Rebora and Pianzola, 2018; Pianzola *et al.*, 2020). Main goal of our project was that of testing if there is a match between the emotions represented in the story and those perceived (and verbalized) by readers. The method we have employed to explore the relationship between text and comments is that of sentiment analysis for the creation of the emotional arcs of stories (Reagan *et al.*, 2016; Jockers, 2017). Besides the text of the novels, we also applied this technique to the dataset of comments, creating a plot of the emotional valence of readers' response along the progression of the story (Fig. 3). By comparing the two plots we discovered a statistically significant positive effect of the story sentiment on the comments' sentiment, meaning that positive emotions in the story elicit readers' positive utterances. This effect is weaker with classics, probably because there are more user-user interactions that tend to have neutral values, since they are a more cognitive-oriented kind of activity aimed at understanding the text rather than expressing emotions. Moreover, looking at the intervals where the two sentiment values have extreme peaks or diverge the most allows to identify text parts that trigger stronger emotions, or reactions contrasting with the story events. This technique allowed us to semi-automatically select which text parts to perform close reading and further explore what elicited a certain reader response. For instance, we found that teenage readers love witty characters, conflicts of affects and values, and cultural references
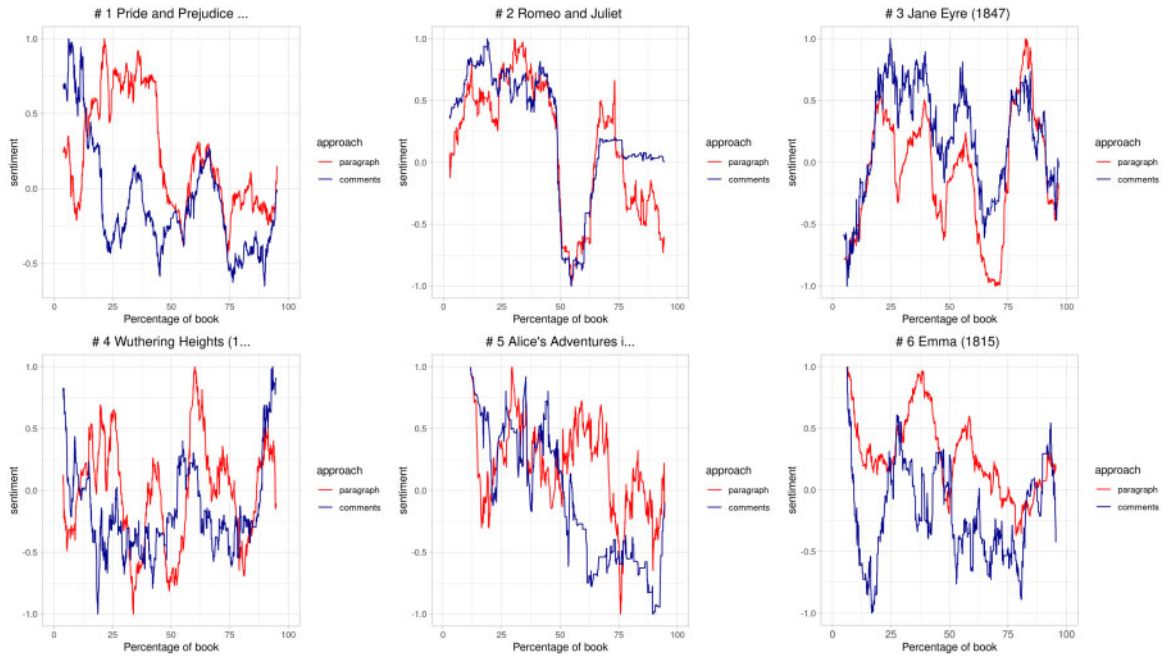
**Fig. 3** Graphs with the emotional arcs of story and comments for six classic novels.

that are familiar to them (Pianzola *et al.*, 2020). Overall, this kind of data and computational analyses can provide large-scale empirical evidence about the link between textual features and readers' emotional response to stories, thus offering a new resource to literary theory, cognitive stylistics, and reading research.

## 3.3 Shared reading (Lauer, Kraxenberger, Gasser, and Sorrentino)

In a follow-up study, we explore by questionnaires the reading and writing behaviour of young people on online literature-platforms, such as Wattpad, Archive of Our Own, or Fanfiktion.de. Starting point of this project is the assumption that reading literature will not decrease because of digitization. Rather, the way(s) literature is dealt with is changing and new, digitally coined practices arise (Lauer, 2020). In particular, this applies to the social aspects of reading and writing, since online literary platforms, just as other social media, promote active participation and interactive exchange among their users. Accordingly, the focus of this still on-going project lies in identifying the practices of online reading and writing, as well as

its (social) functions. Following a multi-method approach, the project pursues both descriptive and quantifiable approaches.

In a first step, the content and formal characteristics of the literature platforms Wattpad and Fanfiktion.de and their usage were described, using anonymized German user-content as examples. This description provides a general overview of the practices of these platforms and allowed for an identification of their functional characteristics. Based on a framework from social psychology (Kietzmann *et al.*, 2011, see also Glüer, 2018), a location of primary, tertiary and secondary functions of the investigated literature platforms was derived. Among others it could be shown that, although interactivity is of central importance for both platforms, they differ in particular with regard to the importance of the function of self-presentation as fan, reader, and/or writer. This function is of special relevance for the 'wreaders' (Landow, 2006) of Wattpad and the predefined publication channels of successful texts within this media machinery (Kraxenberger and Lauer, 2021).

In a second step, an explorative, qualitative study was conducted to find out why young people use

online literature platforms. Young people between the ages of 12 and 17 years from the German-speaking part of Switzerland were to be asked in problem-focused, guideline-based interviews. Participants were sought through libraries, teachers, social media, and posts on literature platforms. For data protection reasons, parents/guardians had to agree to the interviews. This was probably the reason why most interviews were eventually cancelled by the interviewees. Apparently, the young people consider these platforms as something private or part of their youth culture separated from the adult's world. Nevertheless, in terms of research methodology, these obstacles give already an insight into the social role of online literature platforms.

In a third step, this case study includes a larger quantitative survey to better understand the demographics of these users, their practices, motivations, and social interactions with each other. Data were collected in an online survey conducted in German and English, focusing on rather young users (13 years and above; $N = 315$). Participants were recruited through postings on various social media sites (Reddit, Instagram, Facebook, Twitter). The underlying rationale was that media and art exposure is typically self-sought, and that self-motivated users of literature platforms are also more likely to visit groups, sites, and fora specifically targeting such users (cf. Sarkhosh and Menninghaus, 2016). The aim of this survey is to test the validity of the initial description of DSR platforms and to gain a better understanding of users' practices on such platforms, the communicative behaviour between users, potential socio-cognitive benefits that users experience when reading and writing on literature platforms, as well as the underlying motivations to use them. Preliminary results indicate that literature platforms are not exclusively used by teenagers, but are also frequently visited by older users. Different age groups exhibit only minor differences in terms of their practices and motivations. Rather, the individual preference for either reading or writing on literature platforms appears to be the determining factor for the latter.

## 3.4 Evaluation on LovelyBooks (Herrmann, Messerli, and Rebora)

Whereas Wattpad affords its users the option to edit the epitext of the literary texts they are reading,

LovelyBooks and Goodreads are examples of social reading platforms that invite their users to review and rate literature on a separate platform. Another of our case studies assesses LovelyBooks as the most prolific German-language online platform (with currently more than 350,000 registered members, cf. LovelyBooks, 2020) to describe literary evaluation at a large, collective, scale. While it has been shown that metric literary evaluation dates as far back as the 18th century (Spoerhase, 2014), and the underlying values of literary evaluation have been linked to reviewers' linguistic practice (Heydebrand and Winko, 1996), this study is the first to bring both dimensions together, assessing a large-scale review platform for the complex relation between evaluation as represented in the written reviews and the users' ordinal evaluation in the 'star-ratings'.

To answer the question of how the ordinal scale rating maps onto lay reviewers' communicative practices of evaluation, a corpus of approx. 1.3 million lay book reviews by more than 54,000 users was harvested from the LovelyBooks platform. To describe an overall statistical association between the reviews' evaluative diction and the ordinal scale of the star ratings, we applied sentiment analysis to the reviews, rendering mean sentiment values for each rating category (one through five stars). We found that the variables rating and sentiment (SentiWS; Remus *et al.*, 2010) are significantly associated ($\chi2(4) = 227,469$, $p < 0.001$; Cramer's $V = 0.08$). While there is a predictable overall association of low sentiment and low ratings (and high sentiment and high ratings), our detailed analyses per rating category reveal that sentiment and users' quantitative ratings are associated in a non-intuitive way. Pearson residuals in Fig. 4 show that positive sentiment is only overrepresented in the highest category, five stars (indicated by the blue boxes), while clearly underrepresented in Categories 1–3 (red), which simultaneously overuse negative sentiment (blue).

We interpret this finding to question the typical positivity bias for online reviews (Hu *et al.*, 2009). While a four-star rating may intuitively seem 'positive', or a three-star rating 'neutral', the voices of the platform members themselves tell a different story: They overuse negative sentiment for anything other than the full five-star rating.
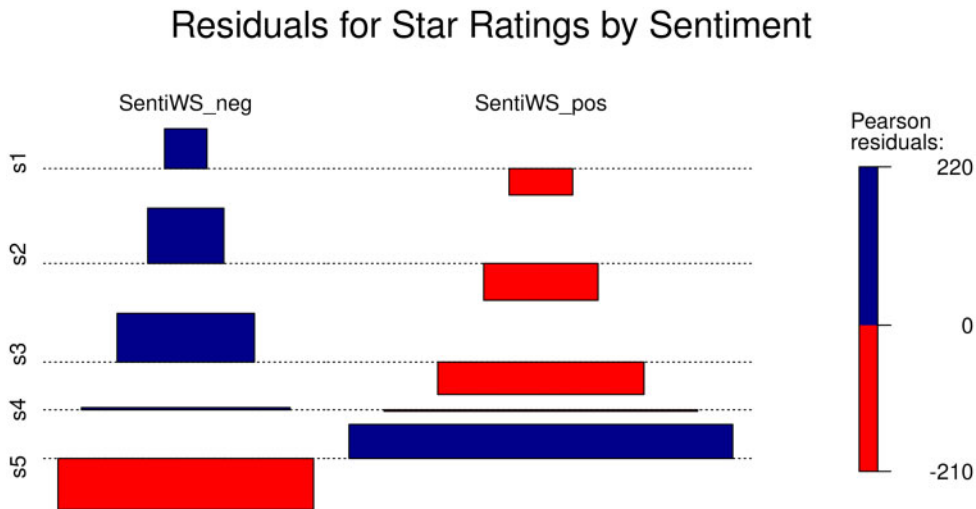
## Residuals for Star Ratings by Sentiment



**Fig. 4** Association plot of polar sentiment ('SentiWS_neg'; 'SentiWS_pos') across rating categories ('one to five stars'; s1–s5) in the LOBO corpus. The colour codes correspond to the sign of the residuals (the differences between observed and expected frequencies), with blue indicating an over-representation, and red an under-representation.

Our post hoc analysis of features (based on a log-likelihood 'keyness' analysis of word-tokens per rating-category; using R-package 'polmineR' version 0.7.11) indicated that five-star reviews comparatively prefer intensifying expressions (including exclamation marks and lexical intensifiers such as *utter(ly)* and *marvelous(ly)*), are more likely to make general statements (*every*, *full(y)*), and more often use the first-person plural subject (*we*). Furthermore, they more often suggest effect-oriented underlying values (*beautiful(ly)*, *captivating*) and refer to body parts (*heart*, *hand*), potentially indicating a close figurative and material relation to the books as artefacts (supported by overuse of verbs such as set/put).

By contrast, already four-star ratings exhibit a more differentiated stance, expressing degree (*somewhat*, *some*, *little*), concession and limitation (*however*, *nevertheless*), as well as references to the 'star rating system' (*deduction*, *four*). Also, non-five-star reviews more often refer to acts of criticism and deliberation (*weakness*, *point of criticism*), and, while friendly, appear more distanced (*interestingly*).

Our analyses present a more nuanced view on evaluative practices in the DSR context, profiting from a combination of DH methods such as sentiment analysis and keyness analysis, and allowing first informed inferences about the relation between diction and the 'metric' evaluation of online reviews. Further research is needed to flesh out these observations and link them to theories of valuation in literary criticism as well as community- and self-related dimensions of DSR.

### 3.5 Sources of authority in online reviews (Boot)

With the advent of sites such as Wattpad and LovelyBooks, mentioned in the previous case studies, traditional authorities in the literary field are said to have become less important (McDonald, 2007). The research question in this case study asks which persons or institutions are considered authoritative by online reviewers. This is a question that takes an institutional approach to the study of literature; we look at which institutions are trusted by readers. Our interest is in today's reader in general and so uses DSR as a lens to study wider reading practice.

The methodology that we applied was to count references to possible authorities, such as traditional critics, newspapers, prizes, television programs, the book trade (publishers, booksellers, libraries), authors, teachers, websites, and private contacts. For a pilot investigation, reviews were downloaded from Dutch weblogs, mass review sites such as Crimezone and watleesjij.nu (What are you reading now?), an

online magazine (*8weekly*), and the *NRC Handelsblad* newspaper (Boot, 2013). We investigated which authorities were mentioned, what their role was, and whether the reviewer agreed with this. We used collections of search terms and regular expressions to search the downloaded reviews. Irrelevant hits were removed, 1,500 relevant hits were annotated.

Because of many limitations (for instance with respect to representativeness), the results are still tentative. The main findings however are summarized in Fig. 5. The four most frequently mentioned authorities are authors, companies and institutions, online critics, and prizes. This is noteworthy for a number of reasons: first, in the view of readers, the author is certainly not dead; second, commercial institutions are frequently mentioned, which is not quite the democratizing influence that is often expected; third, online critics, by and large peer critics, do play an important role; but fourth, the importance attached to prizes may be the revenge of the traditional critics, because they are often members of the juries that award these prizes. Finally, the question of whether our culture in this domain is switching from a vertical, hierarchical orientation into a more horizontal, and peer-oriented

orientation is important. It deserves fundamental investigation and online book discussion offers important insights into the question. It can only be studied in conjunction with the style that different groups of reviewers use, which is the subject of our next case study.

## 3.6 Styles of criticism (Rebora, in collaboration with Massimo Salgaro)

In this case study, we use a corpus of Italian book reviews (Salgaro and Rebora, 2018; Salgaro and Rebora, 2019) to understand how professional critics, journalists, and passionate readers differ in writing reviews and what features can be used to identify them.

The corpus is divided into three subsets: reviews published on DSR platforms (source: aNobii), in paper magazines (*Il Sole 24 Ore*), and in scientific journals (*Between*, *Osservatorio critico della germanistica*, and *OBLIO*). All sub-corpora have an approximate size of 650,000 tokens. Considering the high variance of text length (mean = 259 words; SD = 363 words), the reviews in the three sub-corpora
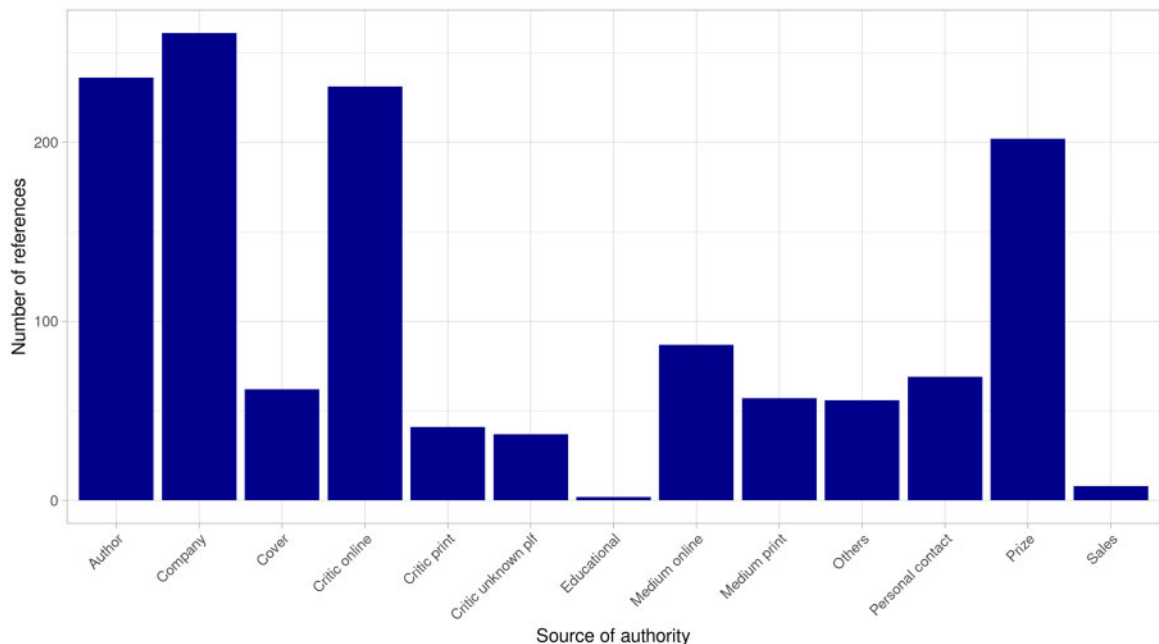


**Fig. 5** Mentioned authorities in the collection of downloaded online reviews.

were split and/or concatenated, generating a series of artificial text chunks of the same length. In this setup, we ran a series of experiments in machine learning combining a total of nine features. The first three were the results of a stylometric analysis (divided per category), using Cosine Delta distance and 2,000 MFW (Evert *et al.*, 2017). The remaining six were based on simple word counts, using as resources:

- an extensive lexicon of literary criticism (Beck *et al.*, 2007);
- selections of terms related to mental imagery and emotional aesthetic response, derived from tools in empirical aesthetics (e.g., Knoop *et al.*, 2016);
- the 'social', 'emotion', and 'body' dimensions in the LIWC Italian dictionary (Agosti and Rellini, 2007).

First, we tested the efficiency of machine learning methods in assigning the reviews to the three categories. Notwithstanding the limited number of features, results are promising. See Table 1 for an overview.

Second, we evaluated the relevance of features in the classification (using Logistic Regression for 250 word-long chunks). Figure 6 shows how stylometric distances (represented by the reddest cells for each category) are the most effective features. Interesting outcomes are also the ineffectiveness of the lexicon of criticism for scientific journals and the effectiveness of mental imagery for both DSR and paper magazines. These results suggest that the shared opinion according to which professionalism in book reviews is a matter of content more than a matter of form might be wrong, as stylistic features (at least those measured by stylometry) prove more efficient in the classification.

Such conclusions will need to be verified via close reading and via a more thorough confrontation with the theories of literary criticism. However, the outcomes of this case study confirm the relevance of corpus-based machine learning approaches for the textual study of DSR and for its comparison with more institutionalized forms of criticism.

## 3.7 Absorption in Goodreads (Kuijpers, Rebora, and Lendvai)

In the last case study, an instrument developed in empirical literary studies to capture the experience of story world absorption was used as an annotation tool to investigate Goodreads reviews (Rebora *et al.*, 2018). Story world absorption is a multi-faceted experience, comprised of deep focused attention that results in loss of awareness of self and surroundings and the track of time; emotional engagement with characters, vivid mental imagery of what the characters and the story world look like; and the experience of deictic shift of the reader from the real world to the story world (Kuijpers *et al.*, 2014). As the experience of absorption is hard to simulate in a lab and instruments like the Story World Absorption Scale (SWAS) are used to retrospectively assess reading experience based on an experimenter-selected story, we focus on developing ways to study absorbing experiences in 'the wild' in a data-driven way, i.e., comparing the statements in the SWAS to unprompted reviews on Goodreads.

Apart from the benefits for instrument validation in empirical literary studies (i.e., do readers use similar language to describe their absorbing reading experiences as researchers do when they are trying to capture these experiences in experimental settings?), one avenue that this type of data-driven research allows us to explore is large-scale genre comparative studies on absorption as it naturally occurs during reading. To study the issue of genre differences, we adopted manual annotation as a comparison tool.

This work raised the issue of shared interpretation (i.e., between researchers and readers), in particular concerning the extent to which it is possible to identify and adjudicate text spans that reference absorption experiences in unstructured natural language input. Five annotators were trained to annotate a corpus of pre-selected reader reviews from the website Goodreads with an extended absorption tag set (counting 145 labels) based on the eighteen statements that compose the SWAS. Each annotator freely established the boundaries of a relevant text segment and

**Table 1.** Accuracy values for the classification of book reviews in the corpus (using a leave-one-out strategy for training and testing)

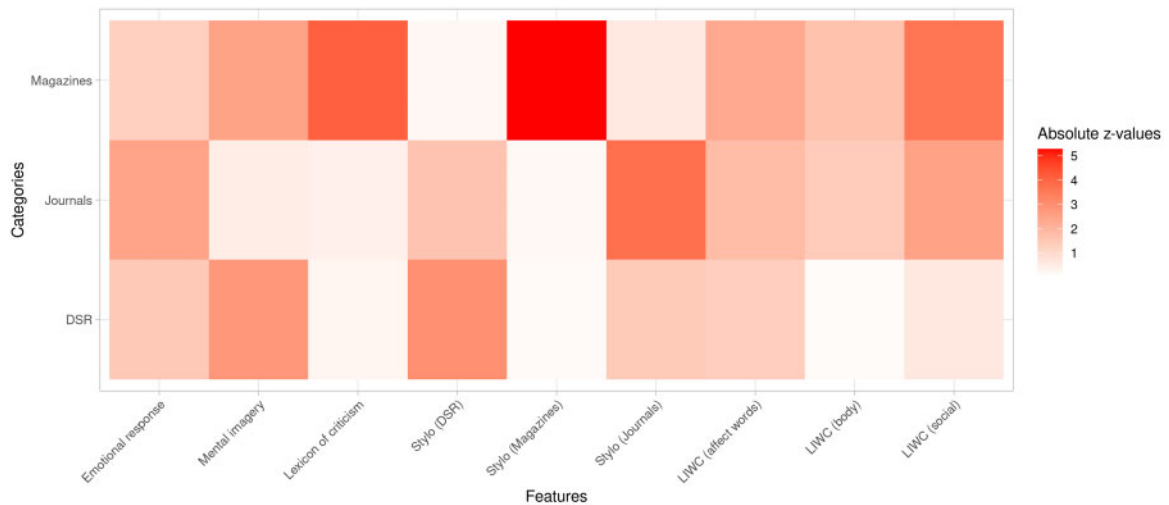|  | SVM | Logistic regression |
| --- | --- | --- |
| 250-word-long chunks | 0.94 | 0.938 |
| 500-word-long chunks | 0.96 | 0.964 |
| 1,000-word-long chunks | 0.976 | 0.978 |

**Fig. 6** Importance of features (absolute *z*-values) for the classification of book reviews, using Logistic Regression on 250 word-long chunks.

was allowed to assign more than one tag to the same text segment. The main criterion for assigning a tag was semantic or conceptual similarity between the statements in the tag set and a text segment. The annotation work was divided into a total of ten rounds of 60–200 reviews (about thirty per week) and after each round the tag set was further specified and the guidelines sharpened. At the end of the annotation process, a total of 1,025 reviews were annotated, with inter-annotator agreement increasing from fair to substantial (mean Krippendorff's alpha = 0.73 in Round 9; for all details, see Rebora *et al.*, 2020a,b). Annotated reviews were also used to train a machine learning classifier, with promising results that suggest a possible automatization of the task (see Lendvai *et al.*, 2019, 2020).

Aggregating the annotations across Rounds 2–9, we were able to group 204 reviews of fantasy books, 324 reviews of romance novels, and 170 reviews of thrillers. Preliminary analyses of these annotations per genre (see Fig. 7) showed that the dimension of Emotional Engagement is used most often by people reviewing romance novels, whereas Attention is used mostly by reviewers of thrillers. These are also the two dimensions of absorption that are used most often in general to describe absorbing reading experiences on Goodreads. Thrillers slightly dominate also for Mental Imagery, while no significant differences (p-values

always > 0.05) were found for the dimension of Transportation. These findings show the usefulness of combining methods from natural language processing with those from empirical literary studies to extend the research on a particular topic like absorption.

# 4 Conclusion

In this paper, we presented a categorization of the research on DSR and discussed seven case studies that show the key role of DH in the study of the phenomenon: while a unique, shared approach cannot be identified, it is indeed the multi-methodology brought by DH that made the advancement of our case studies possible. DH is notoriously a hard subject to define (Terras *et al.*, 2013)—if it is a subject, and not a practice, field, or discipline. What many definitions have in common, however, is that they define DH research as the union of digitally-supported research into traditional humanities subjects with research into digital culture or artefacts (Gibbs, 2013). Research into DSR fits into both aspects of this definition. For example, data collection and database structuring are at the basis of almost all case studies: advanced knowledge of markup languages, web technologies like APIs, and computational techniques like web crawling are
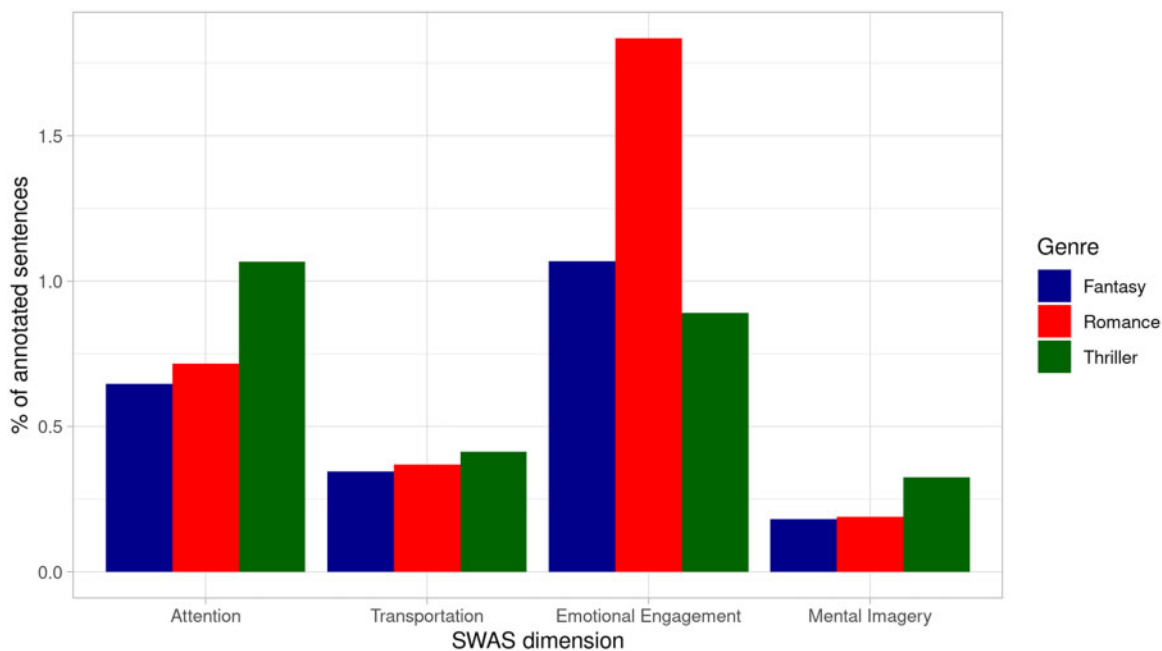
**Fig. 7** Percentages of annotated sentences in fantasy, romance, and thriller reviews.

fundamental for the very feasibility of the projects. Expertise should not be limited to technical aspects, though, as approaches like stylometry, sentiment analysis, and semantic annotation require the discussion of theoretical frameworks like stylistics (Herrmann *et al.*, 2015), theory of emotions (Hogan, 2016), and conceptual modelling (Flanders and Jannidis, 2019). In addition, as our categorization has shown, connections can also be opened to disciplines like sociology, new media studies, educational studies, and many others. Advanced methodologies such as network analysis and machine learning can be involved only after having defined these frameworks.

It is especially the accessibility to research of DSR which can be a game-changing factor for reading research. The digital, online, textual, and massive nature of DSR allows researchers access to evidence of reading on a scale that was unimaginable twenty years ago. As is apparent from our case studies, this scale will require us to use all the technologies that in DH we have come to associate with 'distant reading': with respect to our material, in fact, we face the same situation of abundance that led Moretti (2005) to coin that phrase in the context of researching world literature. The steering of our focus towards the sociological aspects of reading research, then, can be seen as a confirmation of the strict connection between the concept of distant reading and sociology of literature, as already highlighted by Ted Underwood (2017). We are just starting to understand what this will mean for DSR's potential for literary studies. The only way to increase this understanding is through analysis and exploration, through theorizing and testing, being aware of all the limitations of both the study subject and current methodologies. With our work, we hope to have cast the groundwork for all this, by indicating a study area where DH can find new stimuli, new challenges, and new opportunities to grow further.

# Funding

# References

**Agosti, A. and Rellini, A.** (2007). *The Italian LIWC Dictionary*. Austin, TX: LIWC.Net.

**Albrechtslund, A.-M. B.** (2019). Amazon, kindle, and goodreads: implications for literary consumption in the digital age. *Consumption Markets & Culture*, 1–16. doi:10.1080/10253866.2019.1640216.

**Allington, D.** (2016). 'Power to the Reader' or 'Degradation of Literary Taste'? Professional critics and Amazon customers as reviewers of the inheritance of loss. *Language and Literature*, **25**(3): 254–78.

**AuYeung, C., Dalton, S. and Gornall, S.** (2007). Book Buzz: Online 24/7: virtual reading clubs and what we've learned about them. *Partnership: The Canadian Journal of Library and Information Practice and Research*, **2**(2): 1–19.

**Barnett, T.** (2015). Platforms for social reading: the material book's return. *Scholarly and Research Communication*, **6**: 1–23.

**Bartl, A. and Behmer, H.** (eds.) (2017). *Die Rezension: Aktuelle Tendenzen Der Literaturkritik*. Würzburg: Königshausen & Neumann.

**Beck, R., Kuester, H. and Kuester, M.** (2007). *Basislexikon anglistische Literaturwissenschaft*. Paderborn: Fink.

**Blyth, C. S.** (2014). Exploring the affordances of digital social reading for L2 literacy: the case of ecomma. In *Digital Literacies in Foreign and Second Language Education. San Marcos: Calico*, pp. 201–26.

**Boot, P.** (2013). The Desirability of a Corpus of Online Book Responses. In *Proceedings of the Second Workshop on Computational Linguistics for Literature*, pp. 32–40.

**Boot, P.** (2011). Towards a genre analysis of online book discussion: socializing, participation and publication in the Dutch Booksphere. *Selected Papers of Internet Research*, **12**(0), pp. 1–16.

**Bridle, J.** (2010). *Walter Benjamin's Aura: Open Bookmarks and the Future EBook | Booktwo.Org*. http://booktwo.org/notebook/openbookmarks/ (accessed 28 September 2020).

**Brugnatelli, E. and Faggiolani, C.** (2016). Gomorra: 10 Anni Di Conversazioni Su ANobii. In Faggiolani, C. and Vivarelli, M. (eds.), *Le Reti Della Lettura: Tracce, Modelli, Pratiche Del Social Reading*. Milano: Editrice Bibliografica, pp. 261–303.

**Bruns, A.** (2008). *Blogs, Wikipedia, Second Life, and Beyond: From Production to Produsage*. New York: Peter Lang.

**Centro National de Innovación e Investigación Educativa**. (ed.) (2019). *Lectoescritura Digital*. Madrid: Ministerio De Educación Y Formación Profesional.

**Chang, K.** et al. (2020). Book reviews and the consolidation of genre. doi:http://dx.doi.org/10.17613/02q2-1v27.

**Contreras, D. J. A. J.** et al. (2015). The 'Wattyfever': constructs of Wattpad readers on Wattpad's role in their lives. *Laguna Journal of Arts and Sciences Communication Research*, **2**(1): 308–27.

**Cordón-García, J. A. and Gómez-Díaz, R.** (eds.) (2019). *Lectura, Sociedad y Redes. Colaboración, Visibilidad y Recomendación En El Ecosistema Del Libro*. Madrid: Marcial Pons.

**Cordón-García, J.-A.** et al. (2013). *Social Reading*. Oxford: Chandos.

**Costa, P.** (2016). *Il Futuro Della Lettura. L'esprienza Del Testo Nell'era Postmediale*. Milano: Egea.

**Cruces Villalobos, F.** (ed.) (2017). *¿Cómo leemos en la sociedad digital?: lectores, booktubers y prosumidores*. Madrid: Fundación Telefónica.

**Dörrich, M.** (2014). *Book Consumption in Convergence Culture: An Exploratory Audience Study of Media Repertoires of Book Consumption in the Tension between Participation and Corporate Control*. MA thesis, Stockholm University.

**Driscoll, B. and Rehberg Sedo, D.** (2018). Faraway, so close: seeing the intimacy in Goodreads reviews. *Qualitative Inquiry*, 1–12. doi:10.1177/1077800418801375.

**Ernst, T.** (2015). <User Generated Content> Und Der Leser-Autor Als< Prosumer>. Potentiale Und Probleme Der Literaturkritik in Sozialen Medien. In Kaulen, H. and Gansel, C. (eds.), *Literaturkritik Heute*. Göttingen: V&R Unipress, pp. 93–112.

**Es van, K., Wieringa, M. and Schäfer, M. T.** (2018). Tool criticism: from digital methods to digital methodology. In *Proceedings of the 2nd International Conference on Web Studies. WS.2 2018*. New York: Association for Computing Machinery, pp. 24–7. doi:10.1145/3240431.3240436.

**Evert, S.** et al. (2017). Understanding and explaining delta measures for authorship attribution. *Digital Scholarship in the Humanities.*, **36**(suppl.2): ii4–ii16.

**Faggiolani, C., Verna, L. and Vivarelli, M.** (2018). La rete degli editori. Modelli di text-mining e network analysis a partire dai dati di aNobii. *DigitCult*, **3**(2): 1–18.

**Faggiolani, C. and Vivarelli, M.** (eds.) (2016). *Le Reti Della Lettura: Tracce, Modelli, Pratiche Del Social Reading*. Milano: Editrice Bibliografica.

**Finn, E.** (2013). Becoming yourself: the afterlife of reception. *Pamphlets of the Stanford Literary Lab*, **3**: 1–25.

**Flanders, J. and Jannidis, F.** (eds.) (2019). *The Shape of Data in the Digital Humanities: Modeling Texts and Text-Based Resources*. London/New York: Routledge.

**Genette, G.** (1987). *Seuils.* Paris: Editions Du Seuil.

**Gibbs, F.** (2013). Digital humanities definitions by type. In *Defining Digital Humanities. A Reader.* Ashgate, pp. 289–97. http://ccftp.scu.edu.cn/Download/505594d1-330f-4831-94a1-ec87d218345a.pdf#page=306 (accessed 28 September 2020).

**Glüer, M.** (2018). Digitaler Medienkonsum. In Lohaus, A. (ed.), *Entwicklungspsychologie des Jugendalters.* Berlin, Heidelberg: Springer, pp. 197–222.

**Gutjahr, P. C.** (2002). No longer left behind: Amazon.Com, reader-response, and the changing fortunes of the Christian novel in America. *Book History,* **5**: 209–36.

**Hajibayova, L.** (2019). Investigation of Goodreads' reviews: Kakutanied, deceived or simply honest? *Journal of Documentation,* **75**(3): 612–26.

**Harada, T. and Yamashita, S.** (2010). The analysis of differences between online book reviews and those in newspapers. *Joho Chishiki Gakkaishi,* **20**(2): 65–72.

**Harris, S. K.** (2019). *Whohoo!!! Joan of Arc!!!!! American Literary Realism,* **51**(2): 136–53.

**Herrmann, J. B., Frontini, F., Rebora, S., Rybicki, J. and Bories, A.-S.** (2019). DLS tool criticism. An anatomy of use cases. Pre-Conference Workshop, In *DH2019 Book of Abstracts.* Utrecht: ADHO. https://dev.clariah.nl/files/dh2019/boa/1071.html (accessed 28 September 2020).

**Herrmann, J. B., Schöch, C. and van Dalen-Oskam, K.** (2015). Revisiting style, a key concept in literary studies. *Journal of Literary Theory,* **9**(1): 25–52.

**Heydebrand, R. V. and Winko, S.** (1996). *Einführung in die Wertung von Literatur: Systematik - Geschichte - Legitimation.* Paderborn, Zürich [Etc.]: Schöningh.

**Hogan, P. C.** (2016). Affect studies and literary criticism. *Oxford Research Encyclopedia of Literature. doi:* 10.1093/acrefore/9780190201098.013.105.

**Hu, N., Pavlou, P. A. and Zhang, J.** (2009). Why do online product reviews have a J-shaped distribution? Overcoming biases in online word-of-mouth communication. *Communications of the ACM,* **52**(10): 144–7.

**Hugendick, D.** (2008). Jeder Spielt Reich-Ranicki. Die Zeit-online. https://www.zeit.de/2008/17/KA-Laienliteraturkritik (accessed 28 September 2020)

**Jaakkola, M.** (2019). 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,* **10**(1–2): 91–110.

**Jockers, M. L.** (2017). *Introduction to the Syuzhet Package. The Comprehensive R Archive Network.* https://cran.r-project.org/web/packages/syuzhet/vignettes/syuzhet-vignette.html (accessed 28 September 2020).

**Johnson, R. E.** (2016). *The New Gatekeepers: How Blogs Subverted Mainstream Book Reviews.* Richmond, VA: Virginia Commonwealth University.

**Kalir, J. H.** et al. (2020). 'When I Saw My Peers Annotating': student perceptions of social annotation for learning in multiple courses. *Information and Learning Sciences.* doi: 10.1108/ILS-12-2019-0128.

**Kaplan, N. and Chisik, Y.** (2005). In the company of readers: the digital library book as 'Practiced Place'. In *Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries.* pp. 235–43.

**Kietzmann, J. H., Hermkens, K., McCarthy, I. P. and Silvestre, B. S.** (2011). Social media? Get serious! Understanding the functional building blocks of social media. *Business Horizons,* **54**(3): 241–51.

**Knoop, C. A.** et al. (2016). Mapping the aesthetic space of literature 'from below'. *Poetics,* **56**: 35–49. doi: 10.1016/j.poetic.2016.02.001.

**Korobkova, K. A.** (2017). Writing Media: Mobile Story-Sharing Apps as New Learning Ecologies. Ph.D. thesis, UC Irvine.

**Korobkova, K. A. and Rafalow, M.** (2016). Navigating digital publics for playful production: a cross-case analysis of two interest-driven online communities. *Digital Culture & Education,* **8**(1): 77–89.

**Kraxenberger, M. and Lauer, G.** (forthcoming, 2021). Die Plattform als Bühne: Zur Inszenierung von wreaders. In *Wolfenbütteler Arbeitskreis für Bibliotheks-, Buch- und Mediengeschichte, Inszenierung des Buchs im Internet (accepted).*

**Kuijpers, M. M.** et al. (2014). Exploring absorbing reading experiences. Developing and validating a self-report scale to measure story world absorption. *Scientific Study of Literature,* **4**(1): 89–122.

**Kutzner, K., Petzold, K. and Knackstedt, R.** (2019). Characterising social reading platforms. A taxonomy-based approach to structure the field. In *Proceedings of the 14th International Conference on Wirtschaftsinformatik.*

**Landow, G. P.** (2006). *Hypertext 3.0: Critical Theory and New Media in an Era of Globalization,* 3rd edn. Baltimore, MD: Johns Hopkins University Press.

**Lauer, G.** (2020). *Lesen im digitalen Zeitalter.* Darmstadt: wbg.

**Lendvai, P., Rebora, S. and Kuijpers, M. M.** (2019). Identification of reading absorption in user-generated

book reviews. In *Preliminary Proceedings of the 15th Conference on Natural Language Processing (KONVENS 2019): Kaleidoscope Abstracts*. Erlangen: German Society for Computational Linguistics & Language Technology, pp. 271–2.

**Lendvai, P.** et al. (2020). Detection of reading absorption in user-generated book reviews: resources creation and evaluation. In *Proceedings of the 12th Language Resources and Evaluation Conference. Marseille.* France: European Language Resources Association, pp. 4835–41. https://www.aclweb.org/anthology/2020.lrec-1.595 (accessed 28 September 2020).

**Leow, H. M. A.** (2011). Subverting the canon in feminist fan fiction. *Transformative Works and Cultures*, 7. https://journal.transformativeworks.org/index.php/twc/article/view/286/236 (accessed 28 September 2020).

**Leveratto, J.-M. and Leontsini, M.** (2008). *Internet et la sociabilité littéraire.* Éditions de la Bibliothèque publique d'information. doi:10.4000/books.bibpompidou.197.

**LovelyBooks** (2020). Das Unternehmen - Daten und Fakten. https://s3-eu-west-1.amazonaws.com/media.lovelybooks.de/LovelyBooks-Daten_und_Fakten.pdf (accessed 28 September 2020).

**Lukoschek, K.** (2017). 'Ich liebe den Austausch mit Euch'. Austausch über und anhand von Literatur in Social Reading-Communities und auf Bücherblogs. In Bartl, A., Behmer, M. (eds.), *Die Rezension. Aktuelle Tendenzen Der Literaturkritik*. Würzburg: Königshausen & Neumann, pp. 225–52.

**Maryl, M.** (2008). Virtual communities – real readers: new data in empirical studies of literature. In Auracher, J. and van Peer, W. (eds.), *New Beginnings in Literary Studies*. Cambridge: Cambridge Scholars Publishing, pp. 390–406.

**McDonald, R.** (2007). *The Death of the Critic*. London, New York: Continuum International Publishing Group.

**Mehling, G.** et al. (2018). *Leserrezensionen auf Amazon. de: Eine teilautomatisierte inhaltsanalytische Studie*. Bamberg: University of Bamberg Press.

**Merga, M. K.** (2015). Are avid adolescent readers social networking about books? *New Review of Children's Literature and Librarianship*, **21**(1): 1–16.

**Miall, D. S.** (2018). Reader-response theory. In Richter, D.H. (ed.), *A Companion to Literary Theory*. Chichester: John Wiley & Sons, Ltd, pp. 114–25.

**Miller, D. L.** (2011). Talking with Our Fingertips: An Analysis for Habits of Mind in Blogs about Young Adult Books. Ph.D. thesis, Arizona State University.

**Milota, M.** (2014). From 'compelling and mystical' to 'makes you want to commit suicide': quantifying the spectrum of online reader responses. *Scientific Study of Literature*, **4**(2): 178–95.

**Mirmohamadi, K.** (2014). *The Digital Afterlives of Jane Austen: Janeites at the Keyboard*. Basingstoke: Palgrave Macmillan.

**Montesi, M.** (2015). La lectura en las redes sociales: posibilidades de análisis para el investigador. *Revista Ibero-Americana de ciencia da informacao*, **8**(1): 67–81.

**Moody, S.** (2017). Bullies and blackouts: examining the participatory culture of online book reviewing. *Convergence, 25*(5–6): 1063–76.

**Moretti, F.** (2005). *Graphs, Maps, Trees: Abstract Models for a Literary Theory*. London/New York: Verso.

**Murray, S.** (2016). 'Selling' literature: the cultivation of book buzz in the digital literary sphere. *Logos*, **27**(1): 11–21.

**Murray, S.** (2018a). Reading online: updating the state of the discipline. *Book History*, **21**(1): 370–96.

**Murray, S.** (2018b). *The Digital Literary Sphere: Reading, Writing, and Selling Books in the Internet Era*. Baltimore, MD: Johns Hopkins University Press.

**Nakamura, L.** (2013). 'Words with friends': socially networked reading on goodreads. *PMLA*, **128**(1): 238–43.

**Neuhaus, S.** (2017). Vom Anfang und Ende Der Literaturkritik. Das Literarische Feld zwischen Autonomie und Kommerz. In Bartl, A., Behmer, M., Hielscher, M. (eds.), *Die Rezension: Aktuelle Tendenzen der Literaturkritik. Konnex*. Würzburg: Königshausen & Neumann, pp. 33–59.

**Nuttall, L.** (2017). Online readers between the camps: a text world theory analysis of ethical positioning in *We Need to Talk About Kevin*. *Language and Literature*, **26**(2): 153–71.

**Nuttall, L. and Harrison, C.** (2020). Wolfing down the Twilight series: metaphors for reading in online reviews. *Contemporary Media Stylistics*. https://pure.hud.ac.uk/en/publications/wolfing-down-the-twilight-series-metaphors-for-reading-in-online- (accessed 28 September 2020).

**Pianzola, F.** et al. (2019). *Digital Social Reading Public Bibliography*. doi:10.5281/zenodo.3525468.

**Pianzola, F., Rebora, S. and Lauer, G.** (2020). Wattpad as a resource for literary studies. Quantitative and qualitative examples of the importance of digital social reading and readers' comments in the margins. *PLoS One*, **15**(1). doi: 10.1371/journal.pone.0226708

**Pinder, J.** (2012). Online literary communities: a case study of LibraryThing. In Lang, A. (ed.), *From Codex to Hypertext: Reading at the Turn of the Twenty-First Century*. Massachusetts: UP, pp. 68–87.

**Pugh, S.** (2005). *The Democratic Genre: Fan Fiction in a Literary Context*. Bridgend: Seren.

**Ramdarshan Bold, M.** (2018). The return of the social author: negotiating authority and influence on Wattpad. *Convergence: The International Journal of Research into New Media Technologies*, **24**(2): 117–36. doi: 10.1177/1354856516654459.

**Reagan, A. J.** et al. (2016). The emotional arcs of stories are dominated by six basic shapes. *EPJ Data Scienve*, **5**: 5–31. doi:10.1140/epjds/s13688-016-0093-1.

**Rebora, S., Lendvai, P. and Kuijpers, M. M.** (2018). Reader experience labeling automatized: text similarity classification of user-generated book reviews. In *EADH2018*. Galway: EADH. https://eadh2018.exordo.com/programme/presentation/90 (accessed 28 September 2020).

**Rebora, S., Kuijpers, M. M. and Lendvai, P.** (2020a). Mining Goodreads. A digital humanities project for the study of reading absorption. In *Sharing the Experience: Workflows for the Digital Humanities. Proceedings of the DARIAH-CH Workshop 2019*. Neuchâtel: DARIAH-CAMPUS. doi:10.5281/zenodo.3897251.

**Rebora, S., Lendvai, P. and Kuijpers, M. M.** (2020b). Annotating reader absorption. In *DH2020 Book of Abstracts*. Ottawa, ON: ADHO.https://dh2020.adho.org/wp-content/uploads/2020/07/358_AnnotatingReaderAbsorption.html (accessed 28 September 2020).

**Rebora, S. and Pianzola, F.** (2018). A new research programme for reading research: analysing comments in the margins on Wattpad. *DigitCult - Scientific Journal on Digital Cultures*, **3**(2): 19–36.

**Rehberg Sedo, D.** (2011). "I Used to Read Anything That Caught My Eye, But. . .": Cultural authority and intermediaries in a virtual young adult book club. In Rehberg Sedo, D. (ed.), *Reading Communities from Salons to Cyberspace*. New York: Palgrave Macmillan, pp. 101–22.

**Rehfeldt, M.** (2017). Leserrezensionen als Rezeptionsdokumente. Zum Nutzen nicht-professioneller Literaturkritiken Für Die Literaturwissenschaft. In Bartl, A. and Behmer, M. (eds.), *Die Rezension. Aktuelle Tendenzen Der Literaturkritik*. Würzburg: Königshausen & Neumann, pp. 275–89.

**Remus, R., Quasthoff, U. and Heyer, G.** (2010). SentiWS – a publicly available German-language resource for sentiment analysis. In *Proceedings of the International Conference on Language Resources and Evaluation, LREC 2010*, 17–23 May 2010, p. 4.

**Rowberry, S. P.** (2019). *The limits of big data for analyzing reading*. Participations, **16**(1): 237–57.

**Salgaro, M. and Rebora, S.** (2018). Measuring the 'Critical Distance'. A corpus-based analysis of Italian book reviews. In Spampinato, D. (ed.), *AIUCD2018 - Book of Abstracts*. Bari: AIUCD, pp. 161–3. doi:10.6092/unibo/amsacta/5997.

**Salgaro, M. and Rebora, S.** (2019). Towards a computational stylistics of book reviews. In Allegrezza, S. (ed.), *AIUCD 2019 - Book of Abstracts*. Udine: AIUCD, pp. 110–6. doi: 10.6092/unibo/amsacta/6361.

**Sarkhosh, K. and Menninghaus, W.** (2016). Enjoying trash films: underlying features, viewing stances, and experiential response dimensions. *Poetics*, **57**: 40–54.

**Schreier, M.** (2010). 'Is it possible to give a 6 out of 5 stars?': book selection and recommendation in the internet age. *Primerjalna Književnost*, **33**: 307–20.

**Shipman, F. M. and Marshall, C. C.** (2013). Are user-contributed reviews community property?: exploring the beliefs and practices of reviewers. In *Proceedings of the 5th Annual ACM Web Science Conference*. ACM, pp. 386–95.

**Spoerhase, C.** (2014). Das Maß der Potsdamer Garde Die ästhetische Vorgeschichte des Rankings in der europäischen Literatur- und Kunstkritik des 18. Jahrhunderts. In Barner, W., Lubkoll, C., Osterkamp, E., Raulff, U. (eds.), *Jahrbuch der Deutschen Schillergesellschaft*. Berlin, München, Boston: De Gruyer, pp. 90–126.

**Stein, B.** (2010). *A Taxonomy of Social Reading: A Proposal*. http://futureofthebook.org/social-reading/ (accessed 28 September 2020).

**Stein, S.** (2015). Laienliteraturkritik–Charakteristika und Funktionen von Laienrezensionen im Literaturbetrieb. In Kaulen, H. and Gansel, C. (eds.), *Literaturkritik heute*. Göttingen: V&R Unipress, pp. 59–76.

**Steiner, A.** (2008). Private criticism in the public space: personal writing on literature in readers' reviews on Amazon. *Participations*, **5**(2). https://www.participations.org/volume%205/Issue%202/5_02_steiner.htm (accessed 28 September 2020).

**Sutton, K. M. and Paulfeuerborn, I.** (2017). The influence of book blogs on the buying decisions of German readers. *Logos. Journal of the World Book Community*, **28**(1): 45–52.

**Taddeo, G.** (2019). Meanings of digital participation into the narrative online communities. *Italian Journal of Sociology of Education*, **11**(07): 331–50.

**Tang, M.-C., Sie, Y.-J. and Ting, P.-H.** (2014). *Evaluating Books Finding Tools on Social Media: A Case Study of*

*ANobii. Information Processing & Management*, **50**(1): 54–68.

**Tausczik, Y. R. and Pennebaker, J. W.** (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, **29**(1): 24–54.

**Terras, M., Nyhan, J. and Vanhoutte, E.** (2013). *Defining Digital Humanities: A Reader*. Farnham: Ashgate Publishing, Ltd. http://ccftp.scu.edu.cn/Download/505594d1-330f-4831-94a1-ec87d218345a.pdf (accessed 28 September 2020).

**Thelwall, M. and Kousha, K.** (2017). Goodreads: a social network site for book readers. *Journal of the Association for Information Science and Technology*, **68**(4): 972–83.

**Thomas, B.** (2020). *Literature and Social Media*. London/New York: Routledge.

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

**Thompson, P.** (2010). Extending the reading experience with Web 2.0 book groups. *Language Arts Journal of Michigan*, **25**(2): 9.

**Toffler, A.** (1980). *The Third Wave*. New York: Morrow.

**Underwood, T.** (2017). A genealogy of distant reading. *DHQ: Digital Humanities Quarterly*, **11**(2). http://www.digitalhumanities.org/dhq/vol/11/2/000317/000317.html (accessed 28 September 2020).

**Verboord, M.** (2010). The legitimacy of book critics in the age of the internet and omnivorousness: expert critics, internet critics and peer critics in Flanders and the Netherlands. *European Sociological Review*, **26**(6): 623–37.

**Worrall, A.** (2019). 'Connections above and beyond': information, translation, and community boundaries in LibraryThing and Goodreads. *Journal of the Association for Information Science and Technology*, **70**(7): 742–53.