Absorption in Online Book Reviews
English Annotation Guidelines

Moniek M. Kuijpers, Massimo Lusetti, Lina Ruh, Jonathan Tadres, Johanna Vogelsanger, Simone Rebora & Piroska Lendvai

Digital Humanities Lab, University of Basel, Switzerland

Last updated: June 2023

This research was funded by the Swiss National Science Foundation's Digital Lives Scheme (Grant project 10DL15_183194) and the Swiss National Science Foundation's Eccellenza scheme (Grant project PCEFP1_203293).

For more information, contact the first author: moniek.kuijpers@unibas.ch

Introduction

This document defines guidelines for annotating descriptions of narrative absorption in online book reviews. It was developed in the "Mining Goodreads" project, run at the Digital Humanities Lab of the University of Basel and funded by the Swiss National Science Foundation (10DL15_183194 and PCEFP1_203293), in which a corpus of English language online book reviews scraped from the Goodreads website was used. However, the annotation guidelines could also be used on corpora of book reviews from different websites or on corpora of other reader testimonials, like diary studies, or interview transcripts.We are currently developing German guidelines for annotating descriptions of absorption in German language online book reviews.

First, we will give some background information by introducing the "Mining Goodreads" project and describing the motivation for developing this annotation scheme. Second, we will set out the general guidelines, discussing matters like unit of tagging and workflow. Third, we will describe the main

categories used in the annotation scheme, the concepts of presence and negation, and finally a more detailed description of all of the specific tags that fall under these main categories. For each tag we will include both positive and negative example statements from our corpus, insofar as they were available in our corpus.

For more information on this project please refer to our papers on identification of absorption in user-generated book reviews, on resources creation and evaluation, and on annotation (Lendvai, Rebora & Kuijpers, 2019; Lendvai et al., 2020a; Lendvai et al., 2020b; Rebora, Kuijpers & Lendvai, 2020; Rebora, Lendvai & Kuijpers, 2020). We have made the curated meta-data corpus available on Open Science Framework, alongside a pre-print version of a paper on the annotation process and the development of these guidelines, and a pre-print version of another paper introducing the corpus. The corpus consists of one datafile which includes the cleaned and anonymised review texts, the annotations, curations, and relevant metadata for all review texts. Because of the sensitive nature of the data in the corpus, the full corpus is only accessible for researchers who submit a research plan to the authors, and thereafter receive a password to access the full corpus only for research purposes. The metadata files and tag set are openly available for anyone who is interested.

Background and aims

The "Mining Goodreads" project focused on the growing phenomenon of digital social reading. The book review website *Goodreads* holds a wealth of qualitative data about reading experience, text evaluation, and social interactions about reading. So far, this treasure trove of data has not been empirically investigated, and this is partly due to the fact that new methodologies have to be developed to extract the data from the website and to interpret it in a meaningful way, for example by focusing on a specific reading experience, such as absorption. This is exactly the gap that our project aimed to fill.

We developed computational linguistic methods to analyze an extracted corpus from Goodreads with a specific reader response in mind, namely absorption – the feeling of being lost in a book (Nell, 1988; Kuijpers, 2014). Absorption during *literary reading* involves sustained attention to, and open reflection on, the world of the story and the way in which the story is told (Kuijpers, 2014; Kuiken & Douglas, 2017). This focused form of reading leaves the reader unaware of their surroundings, their bodies, and the passage of time (Nell, 1988). Often, absorbing reading experiences involve a sense of being transported to the world of the story, accompanied by an intense emotional engagement with the characters inhabiting that world and vivid visual imagery of what that world looks like (Gerrig, 1993; Green & Brock, 2000; Kuijpers et al., 2014). The experience of absorption is a major component of reading pleasure and reading

motivation, and thus also of the field of empirical literary studies. Usually, within the field of empirical literary studies, this experience is captured using self-report instruments. One such instrument is the Story World Absorption Scale (SWAS; Kuijpers, Hakemulder, Tan & Doicaru, 2014). This scale consists of 18 statements divided into four main dimensions of absorption, namely attention, emotional engagement, mental imagery, and transportation. These 18 statements were taken as the basis for the tag set developed in this project.

The main aim of this project was to first convert *Goodreads* texts into a corpus that enables investigating Story World Absorption using machine learning techniques. To that end a group of five annotators manually annotated and matched the statements from the Story World Absorption Scale (Kuijpers et al., 2014) to the reader reviews posted on *Goodreads.* The annotation work enabled our additional aims to validate the SWAS using inter-annotator agreement scores across annotation rounds, and to enable cross-genre comparisons of readers' absorption using the curated corpus.

General guidelines

Before describing the tag set itself, we will first describe general guidelines that we established over the course of the Mining Goodreads project (For a full description of how this tag set was developed see Kuijpers et al., under review). Annotators were instructed to only assign labels when there is a semantic or conceptual match with a label from the tag set, meaning that a unit of text showed either similarities in terms of the language used to describe an absorbing experience or conceptual similarities between the unit of text and the abstract ideas behind the statements on the SWAS. As conceptual similarity especially can be complex and subjective, we employed a long practice phase during our annotation phase which included weekly discussions about these instances.

This also means that not all units of text receive labels. In fact, we have come to learn that most of the units of texts in online reader reviews do not receive labels when using this annotation scheme. However, in some cases, reviews are full of mentions of absorption and it might be necessary to assign multiple labels to one sentence, as seen in the example in Figure 1.

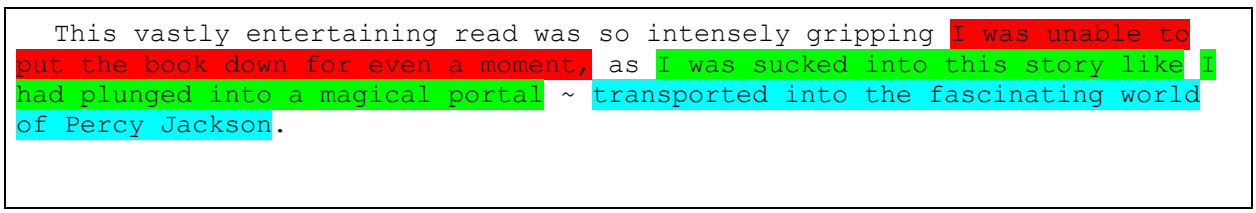

Figure 1. The green highlighted fragment is an instance of Attention, namely A3 (General sense of absorption); the red highlighted fragment is another instance of Attention, namely A7 (Inability to stop reading); and the blue highlighted fragment is an instance of Transportation, namely T4 (Deictic shift).

This is allowed, as the concept of absorption is multi-faceted and it is quite probable that when someone is discussing their state of absorption, they touch on multiple aspects of such an experience within one sentence. The same fragment could also have been tagged as shown in Figure 2, or to have "I had plunged into a magical portal" be tagged twice, once on its own as A3, and once as part of a larger tag for T4. Whether a sentence uses "I", "You", or "The Reader" as subject does not matter, as long as they are talking about an aspect of absorption.

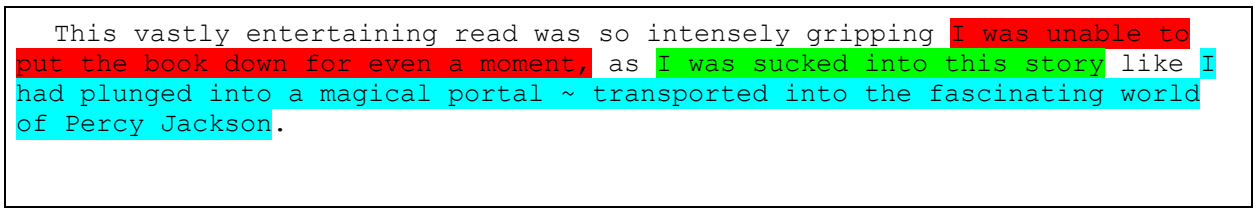

Figure 2. The green highlighted fragment is an instance of Attention, namely A3 (General sense of absorption); the red highlighted fragment is an instance of A7 (Inability to stop reading); and the blue highlighted fragment is an instance of Transportation, namely T4 (Deictic shift).

Workflow

The annotation task is quite involved and requires a lot of concentration. Thus, in terms of workflow, we recommend to spread out annotation work and not to annotate for more than an hour or so per day. This will make sure annotators are able to keep a fresh eye at all times. Additionally, we recommend annotators to go over their annotations after a few days to see if they have missed anything.

When using the INCEpTION tool (Klie et al., 2018, see Figure 3), which we used in the Mining Goodreads project, keep in mind to use either Google Chrome or Safari as a browser, as the program can be unstable in other browsers. When you have finished your annotation on a review (or other piece of text), always "lock" your review, so it is ready to be curated. Make sure you go through all of the dropdown menus. For example, if you want to assign the tag EE3 to a sentence in a review, as the sentence shows a semantic or conceptual overlap with that item on the SWAS, first select "absorption present", then select "Emotional Engagement", and then "EE3". If you do not use all of the drop-down layers, Inception will categorize your annotated phrase as separate from others that essentially tag for the exact same thing.

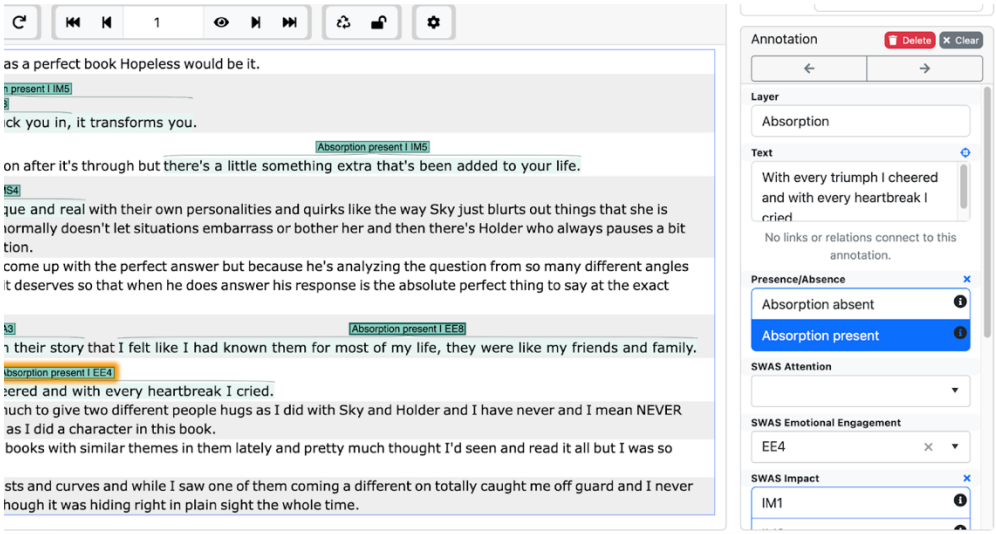

Figure 3. Screenshot of INCEpTION with the dropdown menus on the right side (the review itself is partially cut off for data protection reasons)

Unit of tagging

Annotators should only tag complete clauses (including subject, object and verb) (e.g., "I loved this book series", rather than "loved" or "I loved this book"). In cases where this was not possible, we aimed for complete noun phrases that were as long as possible. We recommend looking for the smallest unit to tag that still makes grammatical sense, unless just tagging this smallest unit changes the meaning of the unit as a whole, such as, for example, tagging "I loved this book", would change the meaning of the original "I loved this book series". We do advise a little leniency in some cases, as we are dealing with natural language, which can contain spelling mistakes or grammatical inaccuracies, keeping in mind that most of the users of Goodreads who write in English, are not actually native speakers of English. Annotators should also include punctuation marks (including periods, commas, colons). In cases where annotators select one sub-clause of a larger sentence, they should also include the comma or colon or semi-colon after the clause, such as shown in Figure 4.

This vastly entertaining read was so intensely gripping I was unable to put the book down for even a moment, as I was sucked into this story like I had plunged into a magical portal ~ transported into the fascinating world of Percy Jackson.

Figure 4. As this unit is just one part of a sub-clause of a larger sentence, the unit of tagging should also include the comma that comes after "moment".

What not to tag

Over the course of the Mining Goodreads project we have identified instances in online reviews that fall outside of the conceptualization of absorption. One such instance is author intentions or author expectations. Reviewers sometimes write about what they think an author intended to achieve in terms of reader emotions, or what they think an author expects them to feel when reading their novel. It may be that a reviewer writes about an author's intentions to draw the reader in. In such cases, the reviewer is not stating that they felt absorbed in the book, rather they are purely stressing that they think the author of the book would *want* them to be absorbed. As we are annotating for readers' sense of absorption, these instances where author intentions or expectations are discussed should not be tagged.

When reviewers talk about their intentions to pick up another book by the same author, they are not necessarily talking about "Anticipation (book series)", which is one of the tags in our annotation scheme. This tag is specifically meant for instances where reviewers talk about wanting to know how a specific story continues. As there are a lot of reviews on books that belong in a book series, and as the story in that case spans several books, we felt the need to add a tag that specifically addresses this type of anticipation. However, a mention by a reviewer that they are looking forward to reading something else by the same author does not refer to story anticipation, and thus does not belong under the conceptualization of narrative absorption. This difference is illustrated in Figure 5.

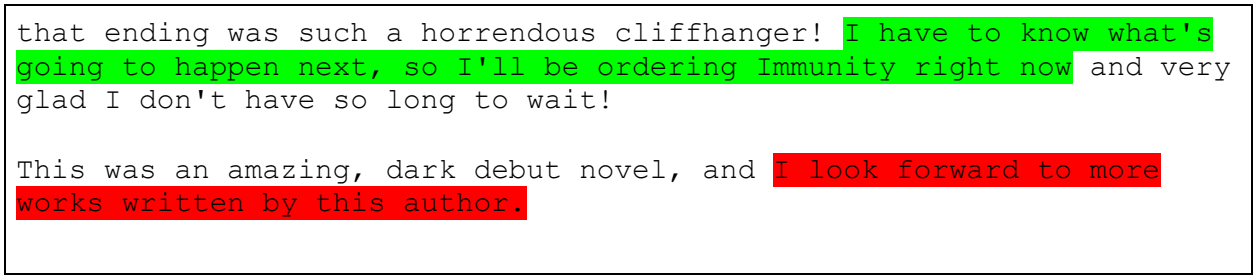

Figure 5. Example of the difference between anticipation of a book series and intention to pick up another book by the same author. The green highlighted fragment is an instance of Impact, namely IM3 (Anticipation book series), because it is about anticipation of story events particular to the book just read, whereas the red highlighted fragment should not be tagged as it is about anticipation unrelated to the specific story just read.

Sometimes the language used in reviews makes it seem like a unit should be tagged, as the wording used is very similar to that used in specific tags. Here follow some exceptions that should not be tagged, mainly because they either are too vague or could have too many potential meanings. In those cases, we can never be completely sure what the reviewer meant exactly.

The first of these is the word “relatable”. In the English language, this word has too many possible meanings. A reviewer could simply mean that they like the character, or they mean that they can identify with them, or that they relate to them. There is a tag that can be used when a reviewer stressed "I could relate to this character" (EE3), but this tag should not be used when a reviewer stresses "I found this character relatable".

There are some adjectives used to describe stories that we considered too ambiguous and therefore decided not to tag, which are: "riveting/riveted", "suspenseful", "poignant", "enchanting", and "invested". We also decided not to tag the phrase "rollercoaster of emotions". It is difficult to categorize any of these words in our annotation scheme, and as all of them are too ambiguous, we decided to specifically avoid tagging them.

Main categories

This tag set could be used at different levels. The top level includes five main categories: Attention, Emotional Engagement, Mental Imagery, Transportation, and Impact. The basis for this categorization was the original Story World Absorption Scale (SWAS; Kuijpers et al., 2014), which we validated and consequently expanded throughout the successive rounds of annotation in this project. The wording of some of the original SWAS statements was also changed to create more similarity to the language used on Goodreads, without losing the conceptual meaning and nuance of the original statements. This mostly meant simplifying the language used in the original SWAS statements.

The original SWAS contained 4 main categories (Attention, Emotional Engagement, Mental Imagery, and Transportation), to which we added the category of Impact. The original categories either had 3 or 5 lower-level tags (i.e., the statements from the Story World Absorption Scale), but each of them were expanded with a couple of additional tags. Based on an extensive literature review on the concept of narrative absorption since the development of the SWAS, we added seven tags based on an absorption inventory created by Bálint and colleagues (2016) in an interview study focused on absorption in reading literature and viewing films. We also added new tags ourselves throughout the annotation process. Whenever we encountered phrases we felt illustrated an aspect of an absorption experience that was not captured by any of the tags available and that was used often enough by different reviewers, we came up with an example statement to add to the tag set. The final tag set included 35 tags, divided up into Attention (7 tags), Emotional Engagement (12 tags), Mental Imagery (4 tags), Transportation (7 tags) and Impact (5 tags). The unequal distribution of number of tags per category is a consequence of our bottom-up approach where we let the data tell us what reviewers are talking about in their reviews.

Absorption-Mention

Throughout the annotation process we started to recognize that at times, reviewers would discuss their *general* reading experiences, habits, or preferences in their book reviews. Rather than talking about a specific experience with the book they were reviewing, they would comment on 'how they usually tend to read'. We realized that these instances could be useful for us as they sometimes pertained to absorbed reading, which is why we brought into life a higher-level category, namely "Absorption-Mention". This category is used to indicate that a non-review specific instance of absorption is mentioned, as shown in Figure 6. The annotators then need to specify in which of the five categories that mention falls, either Attention, Emotional Engagement, Mental Imagery, Transportation or Impact.

Sometimes, while reading sci-fi, it feels difficult to connect with the characters and truly care about their development

Figure 6. The yellow highlighted fragment is an example of Negative Absorption-Mention EE3 (Emotional connection) as it is a mention of negating Emotional connection with characters, but it is a general example ("while reading sci-fi") and not one specifically pertaining to the story reviewed.

Negation versus presence

The annotators could also indicate when users explicitly mentioned or signaled a *lack* or *negation* of absorption (e.g., "I struggled to get through a lot of the pages" (Negative Impact, namely IM1 (Effortless engagement) or "None of the characters really mattered to me″ (Negative Emotional Engagement, namely EE3 (Emotional connection)), to make these distinct from expressions indicating the *presence* of absorption. The tags that were used to make this distinction were "POS" for presence of absorption and "NEG" for negation of absorption, as shown in Figure 7.

The descriptions of people were fantastic; June and Day and all the other characters could not be any clearer in my head

I have no way to visualize the characters

Figure 7. These are two examples of the Mental Imagery MS1 category (I could imagine what the characters looked/smelled/sounded like); the one highlighted in yellow is a fragment in which the *presence* of MS1 was indicated (POS), whereas the one highlighted in red is a fragment in which the *negation* of MS1 was indicated (NEG).

Overview Complete Tag Set

The complete overview of the two lower-level categories of the tag set is depicted in Figure 8. However, before deciding on the proper tag, an annotator will first have to determine whether the segment they are annotating is an example of Absorption-Mention or whether it is a regular example of

Absorption. As a second step, they have to decide whether the segment denotes the Presence or the Negation of absorption. Only in a third step will they determine to which of the five main categories this segment belongs: Attention, Emotional Engagement, Mental Imagery, Transportation or Impact. As a final step they have to decide which of the 35 lower-level tags the segment describes.

Story World Absorption	Attention	A1 (Altered sense of time): While reading time moved differently
		A2 (Concentration): My attention was focused on the book
		A3 (General sense of absorption): I was absorbed in the book
		A4 (No distractions): I was not distracted while reading
		A5 (Forgetting surroundings): While reading I forgot the world around me
		A6 (Anticipation): I was on the edge of my seat / I wanted to know what would happen next
		A7 (Inability to stop reading): I did not want to put the book down / I could not put the book down
	Emotional Engagement	EE1 (Perspective taking): I could imagine what it must be like to be this character
		EE2 (Sympathy): I sympathized with this character
		EE3 (Emotional connection): I felt a connection to this character
		EE4 (Empathy): I felt how this character was feeling
		EE5 (Compassion for story events): I felt for what happened in this story
		EE6 (Anger): I felt angry at this character
		EE7 (Fear): I felt scared for this character
		EE8 (Emotional familiarity): I felt like I knew this character
		EE9 (Wishful identification): I wish I could be more like this character
		EE10 (Emotional understanding): I understood why this character did this
		EE11 (Parasocial response): I want to have some kind of relationship with this character
		EE12 (Participatory response): I wanted to involve myself in the story world events
	Mental Imagery	MS1 (Imagery of character): I could imagine what the characters looked/smelled/felt/sounded like
		MS2 (Imagery of story events): I could see/hear/feel/smell the story events clearly in my mind
		MS3 (Imagery of the story world): I could imagine what the story world looked/smelled/felt/sounded like
		MS4 (Realness): The character/story world felt real to me
	Transportation	T1 (Presence): While reading this I was in the story world
		T2 (Merge of fiction and reality): Elements from the story world came into my world
		T3 (Proximity of story world): The story world felt close to me
		T4 (Deictic shift): I felt transported to the story world
		T5 (Part of the story world): I felt part of the story world
		T6 (Return deictic shift): I returned from a trip to the story world
		T7 (Travel in story world): I lost myself in the story world / I traveled with the characters through the story world
	Impact	IM1 (Effortless engagement): It was an easy read / I devoured the book
		IM2 (Wish to reread): I will/have reread this book/parts of this book
		IM3 (Anticipation book series): I cannot wait to see how this unfolds in the next book
		IM4 (Addiction): I am addicted to this book / I cannot get enough of this book

		IM5 (Lingering story feelings): The book left me feeling … / This book stayed with me for a while

Figure 8. Complete overview of the two lower-level annotation categories

Specific tags

In the following section, we will describe each core tag individually in more detail. For each tag, we will give a conceptualization, an example statement, an annotation showing an example of presence of that tag, an annotation showing an example of negation of that tag (where applicable) and provide comments on how to recognize the tag or distinguish it from related categories.

In some cases, there are no example statements from reviews included, which means that we have found no match in our corpus for this particular tag. In those cases, the corresponding row in the table is left out.

Attention

Tag ID	**Conceptualization**	**Example statement**
A1	Altered sense of time	While reading time moved differently
Review example POS	"Finishing this book in two hours was a complete accident" "I just lost all sense of real time" "almost 800 pages that just fly by"	
Comments	We have found no matches for the negation of this tag in our corpus, which makes sense, as it is unlikely that reviewers would feel the need to express that their perception of the flow of time remained unaltered during their reading.	

Tag ID	**Conceptualization**	**Example statement**
A2	Concentration	My attention was focused on the book/I was concentrated while reading
Review example NEG	"Sometimes I was drifting away" "by the middle [of the book] my attention was lagging"	

Comments	This category is about being concentrated and focusing your attention, and it differs from A3 "General sense of absorption", in that this item is about concentration of attention specifically, whereas A3 includes more general metaphoric descriptions of being absorbed in a book. We did not find any positive matches that mention "attention" or "concentration" specifically, displaying an important difference between the measurement of absorption in lab settings, where this item is crucial to establish whether absorption did take place, and the measurement of absorption in unprompted reader testimonials. Unprompted readers do not tend to talk much about their levels of concentration during reading, unless the book failed to keep their attention. The agency is also more often given to the book (the book absorbed them, grabbed their attention).

Tag ID	**Conceptualization**	**Example statement**
A3	General sense of absorption	I was absorbed in the book/The book pulled me in
Review example POS	"I was so absorbed" "I am so engrossed in her stories" (*Absorption-Mention*) "I was hooked right from the first page and I hung on every sentence, every word, to the very end." "I became so immersed in their story" “This book has got me captivated” "I was sucked into this story" "This book captured me"	
Review example NEG	"The first half wasn't as gripping as the second half" "The plot was interesting and sometimes surprising but not gripping"	
Comments	This category encompasses all metaphorical expressions of generally feeling absorbed by the book (i.e., absorbed, immersed, captivated, grabbed, wrapped up in, submerged, sunk, engrossed, hooked, entranced, gripped, etc.). Even though the term “being hooked on something” could be used to describe addiction, we decided to tag mentions of the “being hooked by a book” under the A3 category rather than the IM4 category “Addiction”, as in the case of absorption “I was hooked by the book” can also refer to being captured by the book. Absorption generally seems to be expressed in terms where the book/text is the active agent and the reader is powerless to stop it. Whether this sentiment is expressed with the “I” or “The reader” as subject or the “book/text” as the grammatical subject of the sentence, does not matter. This metaphorical category differs from another metaphorical category T4 "I felt transported to the story world" in that T4 is specifically focused on the *transportation* metaphor, so on the trip to the story *world* and back (e.g., transported, stepped into the world of the story, returned from a journey/trip/vacation, etc.). **Metaphors that do not fall under either A3 or T4 are intrigued, interested, invested, riveted, and engaged,** as these terms are too ambiguous and could point to various experiences one could have while reading a book, whereby absorption would not be guaranteed.	

Tag ID	Conceptualization	Example statement
A4	Absence of distractions	I was not distracted while reading
Review example POS	"I don't even know how to explain the writing other than the fact that nothing distracted me" "I felt like the whole world was just me and the book"	
Review example NEG	"A sort of voice over is given by a dude called Johnny Truant and that distracted me from the main story." "I think this book is a few decades ahead technologically from where it should be. I try to force myself to forget it, but it is distracting."	
Comments	When using the SWAS in lab settings, usually this item is there to check whether participants were not distracted by anything external to the experiment, like noise or other participants. In these online reader reviews, we see very little to no positive mentions (i.e., absence of distractions). When we see negation of this item (i.e., distractions), the only type of distraction that reviewers tend to talk about are distractions internal to the text (e.g., typesetting or editing mistakes, historical inaccuracies, or style choices that the reviewer does not care for). There are some reviewers who talk about reading a book in one sitting, which should not be classified as "no distractions", as it is too vague (if combined with a description of time flying by, “reading a book in one sitting” could be tagged as A1). Another instance that should not be classified as A4 is reviewers who describe the effort they put into to building a distraction free environment to read in ("I made myself a little nest and didn't move until the final page was read."), as this is saying more about their desire to not be distracted, rather than whether or not they actually were distracted or not.	

Tag ID	Conceptualization	Example statement
A5	Forgetting surroundings	While reading I forgot the world around me/I escaped my reality
Review example POS	"It is that kind of engaging story that makes you keep reading and forget the rest of the world for a few hours" "I love when the world can fade away" *(Absorption-Mention)*	
Review example NEG	"if I didn't have my phone screech at me at certain points during the day (e.g. make lunch or light a fire so that the house wouldn't be freezing), I don't think I would have even noticed anything around me."	
Comments	The main difference between A5 "Forgetting surroundings" and A4 "no distractions", is that classification under A4 requires mention of the word *distraction* specifically. In that sense, A4 names the something that distracted the reader (or specifically mentions there was nothing that was distracting), whereas A5 more generally describes forgetting the world around the reader (as a consequence of not being distracted). In a way "no distractions" facilitates the experience of forgetting the world around you.	

Tag ID	Conceptualization	Example statement
A6	Anticipation	I was on the edge of my seat/I wanted to know what would happen next
Review example POS	“I was afraid to flip every page” "I also wanted to find out what really became of Jane" "this one is a real page-turner, and one with enough twists and reveals that will keep the reader on her toes right up to the very last page"	
Review example NEG	"There are a couple of plot twists, and once I received the subtle hints - I figured out what was going to happen. That ruins the fun for me, I can see it coming"	
Comments	This category encompasses both anticipation in the broadest sense: wanting to know what will happen next (“Pageturner” and "being on the edge of my seat" are this type of anticipation) and fearful anticipation, meaning being anxious about what might happen next (these instances are closely related to suspense). This category is about anticipation felt while reading one book, about what might happen within that book. We have a different category for anticipation of sequels (IM3: Anticipation book series), to account for the kind of anticipation people feel at the end of a book which is part of a larger series (“I cannot wait for the next book to come out”).	

Tag ID	Conceptualization	Example statement
A7	Inability to Stop Reading/Unwillingness to stop reading	I did not want to put the book down/I could not put the book down
Review example POS	"from that point forward, I could not put the book down" "I felt the infamous "just-one-more-chapter" feeling"	
Review example NEG	"another dull one which I put aside about a third of the way through and didn't in fact end up picking back up again" "There were a couple points where I felt bored and took a break from reading for a couple days"	
Comments	This category includes both instances of not wanting to stop reading and instances of not being able to stop reading. In some cases, these types of fragments seem close to some of the Attention items (e.g., “I could not put the book down”), as we may assume that readers have to be focused on the book and not distracted in order to feel like they could not put the book down. However, these instances should be tagged as "Inability to stop reading", as the reasons why the reviewer felt incapable of stopping their reading may vary and shouldn’t be inferred by the annotator. “I did not want this story to end” is “inability to stop reading”, unless it goes together with “because I will miss the characters” (EE11: Parasocial response) or “because I cannot wait until the next book” (IM3: Anticipation book series).	

Emotional Engagement

Tag ID	Conceptualization	Example statement
EE1	Perspective taking	I could imagine what it was like to be this character
Review example POS	"I could easily imagine myself reacting in those exact same ways, thinking those exact same things, and feeling those exact same conflicts" "seeing it all through the eyes of someone […] allows me to feel everything."	
Comments	Perspective taking can be expressed by the reviewer through explicitly stating that they took the perspective of a character in the book they were reading, or it can be expressed by *showing* that they took the perspective of a character. Examples like the one above where the reviewer mentions "seeing it all through the eyes of someone" indicates that the reviewer took the perspective of the character without explicitly stating "I took the perspective of the character". We did not find any comments negating perspective taking, which makes sense, as a reader would more likely comment that they took the perspective of the character than commenting that they never did.	

Tag ID	Conceptualization	Example statement
EE2	Sympathy	I sympathized with this character/I felt for this character
Review example POS	"My stomach hurt as I ached for Iris and the position she was in"	
Review example NEG	"I found myself having a difficult time sympathizing with Rachel" "I didn't pity Adam at all"	
Comments	The sympathy category should not be used to tag for expressions of love for a character. Sympathy is a much more story-related emotion, either of pity, or compassion with the character because of what they are going through. Related to this, hating a character is not a negation of sympathy.	

Tag ID	Conceptualization	Example statement
EE3	Emotional connection	I felt a connection to this character/I related to this character
Review example POS	"which makes him a character you can connect with" "I really felt such a deep connection to certain characters" "I instantly connected with Annabel"	
Review example NEG	"I found it a bit hard to connect with her because of all the scientific jargon she spouted" "I can't say I really related to Anna's situation"	

Comments	“I relate to this character” and “I care for this character” and “I am attached to this character” are all examples of the EE3 category. Do not tag for "this character was relatable", as the word relatable is too ambiguous (i.e., it can mean likable, as well as understandable, or refer to identification or connection).

Tag ID	**Conceptualization**	**Example statement**
EE4	Empathy	I felt how this character was feeling
Review example POS	"I could feel the emotions of Darby" "I shivered, I feared and I suffered along with the characters" "I actually felt a bit scared like the characters"	
Review example NEG	"I found Scorcher as a narrator to be distant and difficult to empathize with"	
Comments	Under this category fall statements referring to empathy, which here is understood as Emotional Empathy (or Emotional Contagion), "meaning feeling similar emotions to someone else"	

Tag ID	**Conceptualization**	**Example statement**
EE5	Feeling for story events	I felt for what happened in the story/I sympathized with the story events
Review example POS	"I actually ache for what is going on between Ana & Christian" "I cried reading the last 30 pages or so, because I couldn't handle how everything turned out."	
Comments	The EE5 category is about feelings of compassion for the story events/situations that take place in a story world. It is the “story events-equivalent” of EE2, which is purely about sympathy for specific characters, but not for events. Mentions of crying can be tagged under EE5 if they are related to ‘for’ or ‘about’ (story event). Other than that, general descriptions of bodily feelings or emotions in response to reading a book, or general statements like "this book made me feel emotional" should not be tagged under this category. We did not find any comments negating compassion for story events, which makes sense, as a reader would more likely comment on the presence of compassion towards story events than commenting when they feel indifferent towards them.	

Tag ID	**Conceptualization**	**Example statement**
EE6	Anger	I felt angry at this character

Review example POS	"I was so personally affronted by the actions of certain characters" "Jake is a total idiot sometimes though, when his doubts of being good enough and fears of commitment rise. I got super mad at him at those times."
Comments	An expression of anger should only be tagged if it is specifically related to something happening in the story, as in that case it would be a sign of being absorbed enough in the story that the reader cares enough to be angry towards a character. Expressions of anger towards the author, for not living up to a readers' expectations, for example, should not be tagged. We did not find any comments negating anger, which makes sense, as a reader would more likely comment on the presence of anger towards a character than commenting when they feel indifferent towards them.

Tag ID	**Conceptualization**	**Example statement**
EE7	Fear	I felt scared for this character
Review example POS	"During her chase, I became afraid for her"	
Review example NEG	"The danger didn't really worry me in any way"	
Comments	This category is closely related to A6 (Anticipation) in a way, but differs in that the fear in this category is focused on the character and being fearful for what will happen to them. In the case of anticipation, potential feelings of fear are aimed at the story events, the progression of the plot.	

Tag ID	**Conceptualization**	**Example statement**
EE8	Emotional familiarity	I felt like I knew this character
Review example POS	"you feel like you have known the characters for your whole life" "There are very few people in real life that I am closer to, and know more about, than the characters in these books."	

Comments	Under the category of emotional familiarity fall statements of "feeling like you know the character", but also “Getting back to characters that you know” or “meeting characters again” in those cases where people are talking about sequels in book series. We did not find any negation of this category in any of the reviews, but a negation of this category could be a statement like "I felt like I didn't really get to know the characters". This category is closely related to EE11 (Parasocial response), the main difference between the two being that In the EE11 category reviewers talk about characters as having relationships with them, and naming those relationships (e.g., friends, siblings, “book boyfriend”), whereas in this category the descriptions are about a more general sense of familiarity with the characters.

Tag ID	**Conceptualization**	**Example statement**
EE9	Wishful Identification	I wish I could be more like this character
Review example POS	"wish I could be more like her" "I want to live adventures like them. I want to be as ambitious as Adam and live a life like Gansey with the happiness of Blue. I want to do something dangerous like Ronan, buying a really awesome car and just drive a race. I want to experience a friendship as good as theirs"	
Comments	We didn’t find any comments negating this category, which may indicate that reviewers either do not experience "fear of identification" (i.e., wishing to not be or become like a character) or do not tend to talk about these kinds of experiences in their reviews.	

Tag ID	**Conceptualization**	**Example statement**
EE10	Emotional Understanding	I understood why this character did this
Review example POS	"But then I understand his struggles more" "I've never gone through anything like what they do, but I felt like I could understand what they were feeling"	
Review example NEG	"As much as I try, I do not understand her" "I just couldn't get how she kept making excuses for him and let herself be emotionally and physically abused" "And I just don't understand why he couldn't accept some help from his friends"	
Comments	This is a broad category, it covers understanding of actions, desires, thoughts, feelings, emotions and attitudes of characters and in that sense, it is similar to Cognitive Empathy, meaning the ability to understand someone else’s perspectives and feelings. This category is different from EE4 (Empathy) and EE3 (Emotional connection), as it is concerned with (cognitive) understanding of why characters behave in a certain way, their motivations and reactions, rather than sharing in their feelings, or feeling similar to them.	

Tag ID	Conceptualization	Example statement
EE11	Parasocial Response	I (want to) have some kind of relationship with this character
Review example POS	"I usually feel sad to have to part from them by the end of the book" "I want to talk to the characters, eat some bread and cheese with Kvothe, Sim, and Wil" "I think about Lennie surrounded by the orange walls of The Sanctum feeling this pain vaster than the vastest desert and it hurts my heart and makes me wish I was her best friend"	
Review example NEG	"I was less keen to spend much time with the main and supporting characters"	
Comments	Only use the parasocial response tag, if there is an 'I' statement. This will hopefully rule out instances where parasocial responses need to be inferred (too vague). "I was rooting for this character" and "I feel sad to have to part from characters" are both instances of the parasocial response category. "I did not want this story to end" is "inability to stop reading", unless it goes together with "because I will miss the characters" (parasocial response) or "because I cannot wait until the next book (anticipation). It is very unlikely that you will find a semantic match for this category, as reviewers tend to express their wish about having relationships with characters by explicitly using the names of the characters.	

Tag ID	Conceptualization	Example statement
EE12	Participatory Response	I wanted to involve myself in the story events
Review example POS	"In my mind I was screaming 'WATCH OUT' and 'Don't walk down this street by yourself in the middle of the night'" "There were several times that I wanted to reach into my book and give her a hug and lend her my shoulder to cry until all her tears ran out" "I had moments where I yelled for characters"	
Comments	In the broadest theoretical sense, participatory responses represent a commentary of the reader on the story events. In our tag set, the participatory response category is used to annotate mentions of the reviewer wanting to involve themselves somehow in the story events. It is closely related to parasocial response, but the crucial difference between the two lies in the focus of the response: in the case of parasocial response there is a desire to have a relationship with a character (perhaps even outside of the confines of the story or story world), whereas with participatory response there is a desire to involve oneself into the story events particular to the book they read or exert some influence over those specific story events.	

Mental Imagery

Tag ID	Conceptualization	Example statement
MS1	Imagery of characters	I could imagine what the characters looked/smelled/sounded like
Review example POS	"you could actually see where they were, *what they were wearing*, and how they reacted" "I could vividly picture them all" "The descriptions of people were fantastic; June and Day and all the other characters could not be any clearer in my head"	
Review example NEG	"I have no idea what they look like or how far they deviate from humans" "I have no way to visualize the characters"	
Comments	This category is only for imagining what characters are like in terms of the senses (how they look, how they smell, what they sound like). In terms of imagining what a person is feeling or thinking, this should be tagged as one of the emotional engagement categories.	

Tag ID	Conceptualization	Example statement
MS2	Imagery of story events	I could see/hear/smell the story events clearly in my mind
Review example POS	"you could actually see where they were, what they were wearing, and *how they reacted*" "their mannerisms and how they talked was richly detailed and made for a story that was so easy to visualize for me" "some pretty bad-ass battle scenes were as vivid as a movie"	
Review example NEG	"It's really choppy and hard to visualize"	
Comments	This category is only for imagining what events and situations in the story are like in terms of the senses (how they look, how they smell, what they sound like). In terms of imagining what it must be like to be in that situation or feel for the situation, you should use either items from the Transportation category or the EE5 category respectively.	

Tag ID	Conceptualization	Example statement
MS3	Imagery of story world	I could imagine what the story world looked/smelled/sounded like
Review example POS	"I felt like I could actually see the hot dusty streets of Mereen and the cool palatial pyramids" "you could actually see *where they were*, what they were wearing, and how they reacted" "I was able to very vividly imagine the magnificent land of Trollus" "I could smell the coffee. See the streets alive with movement and energy. Hear the music in the air."	

Review example NEG	"I felt that there could have been a little more insight into what the world around these characters looked like"
Comments	This category is only for imagining what the world in the story is like in terms of the senses (how it looks, how it smells, what it sounds like). In terms of imagining what it must be like to be *in* that world, you should use items from the Transportation category. We also use this category for mentions of “atmosphere”.

Tag ID	**Conceptualization**	**Example statement**
MS4	Realness	The character/story world felt real to me
Review example POS	"The characters grow on you, you'll feel like they're real." "These characters are just as real to me as individuals I know in real life."	
Review example NEG	"Tatiana does not ring at all true on any level, nor Karla's devotion to her, just don't think real zealots would behave that way."	
Comments	This category pertains to expressions of the story world or the characters in it feeling like they are real, like they could happen in real life. These statements usually have to do with believability and are included in the tag set because a suspension of disbelief is a vital component of feeling absorbed. Readers often express that realism or feeling as if something could happen in real life are important parts of being able to suspend disbelief long enough to feel absorbed. Care has to be taken though with tagging "realism" or "realistic"; depending on the context this can also refer to a literary period/genre/style, rather than the feeling of a character or world being real. Additionally, keep in mind that mentions of characters or worlds can appear real to reviewers, even in the case of reading fantasy or science fiction novels.	

Transportation

Tag ID	**Conceptualization**	**Example statement**
T1	Presence	While reading this I was in the story world
Review example POS	“I felt as though I was actually inside of the story” "I felt like I was right there in the crowd watching him"	
Comments	This tag is closely related to T5 (being a part of the story world). The difference between T1 and T5 lies in the fact that T1 describes the reader more as a spectator, whereas T5 describes the reader more as an active participant.	

Tag ID	Conceptualization	Example statement
T2	Merge of fiction and reality	Elements of the story world came into my world
Review example POS	"We had a thunderstorm the other day and I just thought about Zeus having another one of his fits!" "I almost feel private about it - as if it [the story] has become a part of me and talking about is like letting others peek into my soul" "I will always be on the lookout for a note stuck to a fence or tucked in a tree branch telling me about her sleepless nights talking to her sister"	
Comments	The merging of fiction and reality can be expressed by the reviewer through explicitly stating that some elements of the story world came into their own world during or after reading a book, or it can be expressed by *showing* that elements of the story world came into their own world. Examples like the one above where the reviewer mentions "being on the lookout for a note stuck to a fence or tucked in a tree branch" indicates that the reviewer is expecting events that took place in the story world to happen in the real world. It is very unlikely that you will encounter semantic matches for this statement.	

Tag ID	Conceptualization	Example statement
T3	Proximity of story world	The story world felt close to me
Review example POS	"I read it well inland, but the feel of place was strong enough to completely involve me." "It [the story world] has become a part of me"	
Comments	This category refers to a general sense of closeness of the story world, which can be meant in two different metaphorical ways: emotional closeness such as in the second example "the story world has become a part of me" or a sense of physical closeness like in the first example where the reviewer is stating that even though they were not in a similar place to the one in the story, the place felt close enough ("strong enough") to evoke a strong involvement.	

Tag ID	Conceptualization	Example statement
T4	Deictic shift	I felt transported to the story world
Review example POS	"I'm instantly transported once I open the pages" "You are dropped in the middle of a different world" "I'm pulled into another time and place"	

Comments	Both item A3 and T4 use metaphorical language to describe an experience of absorption. The main difference is that A3 is more general (e.g., absorbed, immersed, captivated, grabbed, wrapped up in, submerged, sunk, engrossed, hooked, entranced, gripped, pulled in the book, sucked in the book, drawn into the book, etc.), whereas T4 is specifically focused on the transportation metaphor, so on the "trip to the story world and back" (e.g., transported, stepped into the world of the story, returned from a journey/trip/vacation, etc.). **Metaphors that do not fall under either A3 or T4 are intrigued, interested, invested, riveted, engaged.** When a reviewer states "it was not easy to get into", this is NOT indicative of transportation, because that 'technically' means that in the end they did get into it. This is rather a Negation of IM1 (Effortless engagement).

Tag ID	**Conceptualization**	**Example statement**
T5	Part of the story world	I felt part of the story world
Review example POS	"You felt like you were sitting by the campfire at camp Half Blood while he shared his first adventure there" "I find myself right there alongside Meghan, Puck and Ash" "I lived with Kristen from the time she was three until she died."	
Comments	This tag is closely related to T1 (Presence). The difference between T1 and T5 lies in the fact that T1 describes the reader more as a spectator, whereas T5 describes the reader more as an active participant.	

Tag ID	**Conceptualization**	**Example statement**
T6	Return deictic shift	I returned from a trip to the story world
Review example POS	"I feel like I've just returned from a long vacation in Martin's fantasy kingdom" "I finished this book feeling like I had just come home from a trip away, and like a long trip, I felt out of place coming out of my own bedroom" "Going to live in the Russell universe for a few hours is a vacation, and it is by far the cheapest one I know of"	
Comments	This category is in some ways the counterpart to T4, which is a direct description of "being transported to a story", of making a movement into the text world. In this category, comments can be subsumed in which reviewers make clear that they have felt transported while reading, because now that reading is over they feel like they have "returned" from another place.	

Tag ID	**Conceptualization**	**Example statement**
T7	Travel within story world	I lost myself in the story world/I traveled with the characters through the story world

Review example POS	"I would be this unnamed character that followed the main character around in his adventures" (*Absorption-Mention*) "My second stroll around Paris with Anna and Etienne was just as breathtaking and lovely as my first time. I loved exploring the city again." "I just sunk into this book and disappeared into the 1300s in Norway"
Comments	When trips within the story world are mentioned (e.g., "I walked along with the character on their journey") use T7, and not T4, as it is not about a trip from the actual world to the story world, it is about being and moving within the story world.

Impact

Tag ID	**Conceptualization**	**Example statement**
IM1	Effortless Engagement	It was an easy read/I devoured this book
Review example POS	"The pages fly with ease" "I was glued to my kindle tapping through the pages as fast as my eyes would let me devour the text"	
Review example NEG	"It's not the easiest book to get into, especially because the beginning is hard to understand, at least it was for me until the third attempt to start it" "Initially, I found myself struggling with On the Jellicoe Road. I was lost, confused and felt like I was reading a book written in a foreign language"	
Comments	When a reviewer states "it was not easy to get into", this is NOT indicative of transportation, because it could 'technically' mean that in the end they did get into it. This is rather a Negation of IM1 (Effortless engagement), because even though in the end they did get into it, it was not effortless for them. "I devoured this book" or "I eat this kind of story up" is effortless engagement, as it expresses an ease of consumption.	

Tag ID	**Conceptualization**	**Example statement**
IM2	Wish to Reread	I will/have reread this book/parts of this book
Review example POS	"one that I'm definitely going to buy, read and re-read" "I [...] re-read paragraphs to let the enormity of it all sink-in"	
Review example NEG	"It was a good thing I just read the ebook copy of this book because I think I'm never gonna read it again. […] This book was too disturbing to be read again" "this isn't a book I feel the need to own for future re-reads"	
Comments	We included the "wish to reread" category in this tag set as mentions of rereading (either planned or already happened) seem to go hand in hand with high levels of enjoyment and engagement with a book.	

Tag ID	Conceptualization	Example statement
IM3	Anticipation Book Series	I cannot wait to see how this unfolds in the next book
Review example POS	"I am eager to see how he evolves over the course of the next few books" "that ending was such a horrendous cliffhanger! I have to know what's going to happen next, so I'll be ordering Immunity right now and very glad I don't have so long to wait!"	
Review example NEG	"I didn't feel the urge to continue with this epic trilogy, I thought this book was not going to exceed my expectations" "I don't want a spin-off (which has been hinted at, and seems supported by the way this one cut off)."	
Comments	Because we look at book series as one story, “looking forward to the next book” is a form of “anticipation”. “I did not want this story to end” should be tagged as A7 (Inability to stop reading), unless it goes together with “because I will miss the characters” (EE11: Parasocial response) or “because I cannot wait until the next book (IM3: Anticipation book series)	

Tag ID	Conceptualization	Example statement
IM4	Addiction	I am addicted to this book/I cannot get enough of this book
Review example POS	"It stayed with me, even when I wasn't reading it; making me think about the characters and storyline and would positively drive me crazy until I would pick it back up!" "I think I have a new obsession. […] and I now need more!" "I honestly couldn't get enough of them!"	
Review example NEG	"I can't say this was a satisfying wrap up to the series […] Nor did it leave me wanting more"	
Comments	Even though the term “being hooked on something” could be used to describe addiction, we decided to tag mentions of the “being hooked by a book” under the A3 category rather than the IM4 category “Addiction”, as in the case of absorption “I was hooked by the book” can also refer to being captured by the book.	

Tag ID	Conceptualization	Example statement
IM5	Lingering Story Feelings	The book left me feeling …/ This book stayed with me for a while
Review example POS	"It left me feeling drained and subdued" "it is one of those stories that just stick with you"	

Review example NEG	"…it has been a while since I have read it and I have probably read hundreds of books afterwards so I forgot all the details and even what exactly happened in the end"
Comments	This category is a rather broad one pertaining to instances in which the book has an impact on the reviewer even after reading was finished. We included this category, because we felt that while reading through the reviews, it seemed that absorption in a book does not necessarily end when the actual reading is finished; the lingering feelings are a part of why the reviewer felt absorbed.

References

Gerrig, R. J. (1993). *Experiencing narrative worlds: On the psychological activities of reading*. New Haven, Conn.: Yale University Press.

Green, M. C., & Brock, T. C. (2000). The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology, (79)*5, 701–721. doi: 10.1037/0022-3514.79.5.701.

Klie, J.-C., Bugert, M., Boullosa, B., de Castihlo, R. E., & Gurevych, I. (2018). The INCEpTION Platform: Machine-Assisted and Knowledge-Oriented Interactive Annotation. In *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations,* 5-9. Association for Computational Linguistics.

Kuijpers, M. M. (2014). *Absorbing stories. The effects of textual devices on absorption and evaluative responses* (Doctoral dissertation). Utrecht University, Utrecht. Available at: https://dspace.library.uu.nl/handle/1874/304849.

Kuijpers, M. M., Hakemulder, F., Tan, E. S. H., & Doicaru, M. M. (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. doi: 10.1075/ssol.4.1.05kui.

Kuiken, D., & Douglas, S. (2017). Identifying forms of absorption that facilitate aesthetic and explanatory outcomes. In F. Hakemulder, M. M. Kuijpers, E. S. H. Tan, K. Bálint & M. M. Doicaru (Eds.), *Narrative absorption* (pp. 177–198). Amsterdam: John Benjamins Publishing. doi: 10.1075/lal.27.12kui.

Lendvai, P., Darányi, S., Geng, C., Kuijpers, M. M., Lacalle, O. L. de, Mensonides, J.-C., Rebora, S., & Reichel, U. D. (2020). Detection of Reading Absorption in User-Generated Book Reviews: Resources Creation and Evaluation. In *Proceedings of the 12th Language Resources and Evaluation Conference,* 4835–4841. Marseille: France: European Language Resource Association.

Lendvai, P., Reichel, U., Kuijpers, M. M., & Rebora, S. (2020). Ranking of Social Reading Reviews Based on Richness in Narrative Absorption. In *Proceedings of the 5th Swiss Text Analytics Conference (SwissText) & 16th Conference on Natural Language Processing (KONVENS)*. Zurich: CEUR.

Lendvai, P., Rebora, S., & Kuijpers, M. M. (2019). Identification of Reading Absorption in User-Generated Book Reviews. In *Proceedings of the 15th Conference on Natural Language Processing (KONVENS 2019)* (pp. 271–272).

Nell, V. (1988). *Lost in a book: The psychology of reading for pleasure.* Yale University Press.

Rebora, S., Lendvai, P., & Kuijpers, M. (2018). Reader experience labeling automatized: Text similarity classification of user-generated book reviews. In *EADH 2018 Abstracts*. Galway: EADH.

Rebora, S., Kuijpers, M. M., & Lendvai, P. (2020). 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.

Rebora, S., Lendvai, P., & Kuijpers, M. M. (2020). Annotating Reader Absorption. In *DH2020 Book of Abstracts*. Ottawa: ADHO.