



University
of Basel

Online book reviews and other reader response corpora: A rich but overlooked data source in the study of literature's global dimensions?

Moniek M. Kuijpers, 20.11.2024

Contextualization

- Empirical literary studies
- Online book reviews
- Story World Absorption
- LOD is new to me
- Expand the usefulness of my corpora



My dream

Build a database for OA reader response data

- Searchable (e.g., type of data, sample, text read, methods used)
- Multi-lingual data to enable cross-cultural comparative research
- Linked to OA data repositories and repositories with literary text (and/or literary metadata) corpora

To expand and enrich the fields of empirical literary studies and computational literary studies

- Stimulate cross-cultural replication
- Stimulate meta-analysis research
- Offer a great teaching resource

The “Mining *goodreads*” project



Main aims:

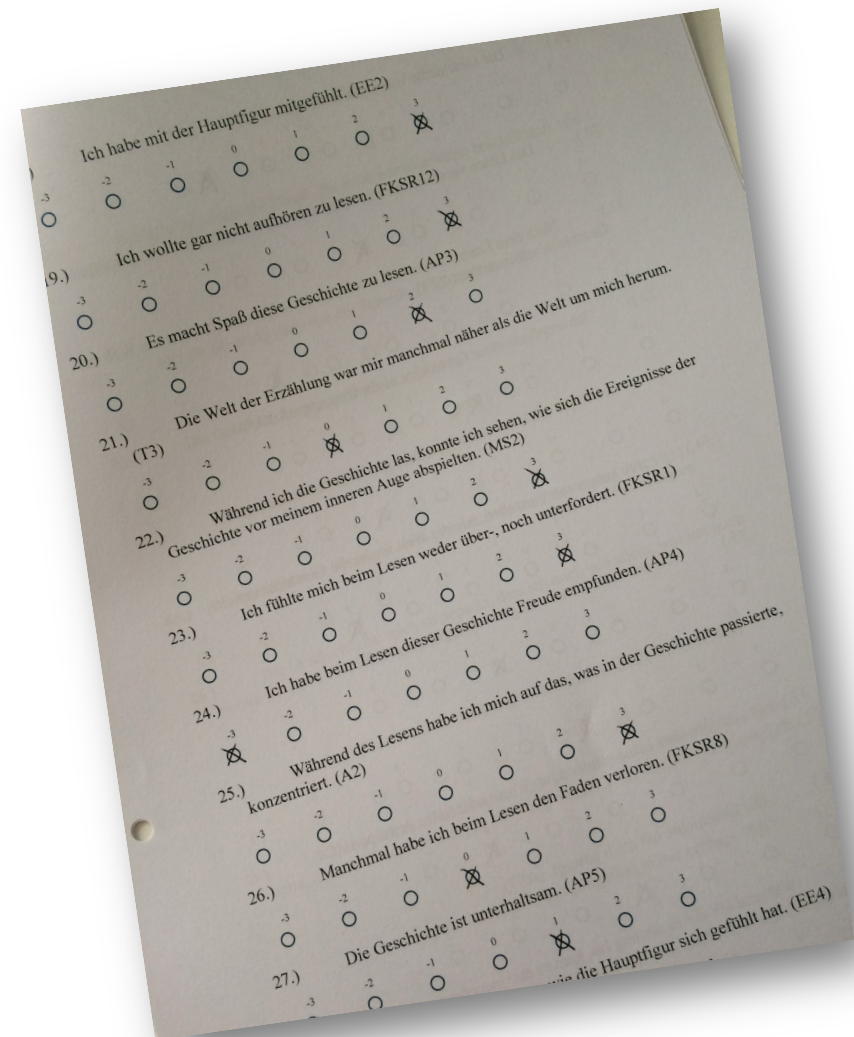
- (1) the validation of the SWAS through comparison with reviews on *Goodreads*
- (2) converting *Goodreads* into an extensive qualitative corpus for meaningful computational analysis of reader responses
- (3) the comparison of readers' absorption across different genres.

Introducing the Story World Absorption Scale

A subjective experience of **focused attention** on the story world, which leaves the reader **unaware of their surroundings, themselves and the elapse of time**, which is experienced as effortless and enjoyable.

Self-report instrument with 18 statements capturing 4 dimensions of story world absorption:

- Attention
- Mental imagery
- Emotional engagement
- Transportation



Kuijpers, Hakemulder, Tan & Doicaru, 2014

English and German SWAS items

Attention

E: I felt absorbed in the story

G: Ich war völlig in die Geschichte vertieft

Transportation

E: When I was reading the story it sometimes seemed as if I were in the story world too

G: Während ich die Erzählung las, kam es mir manchmal so vor, als befände ich mich auch in der Welt der Geschichte

Emotional Engagement

E: I felt connected to the main character in the story

G: Ich fühlte mich der Hauptfigur der Geschichte verbunden

Mental Imagery

E: I could imagine what the world in which the story took place looked like

G: Ich konnte mir vorstellen, wie die Welt, in der die Geschichte stattfand, ausgesehen hat

English Corpus: Starting Tag Set

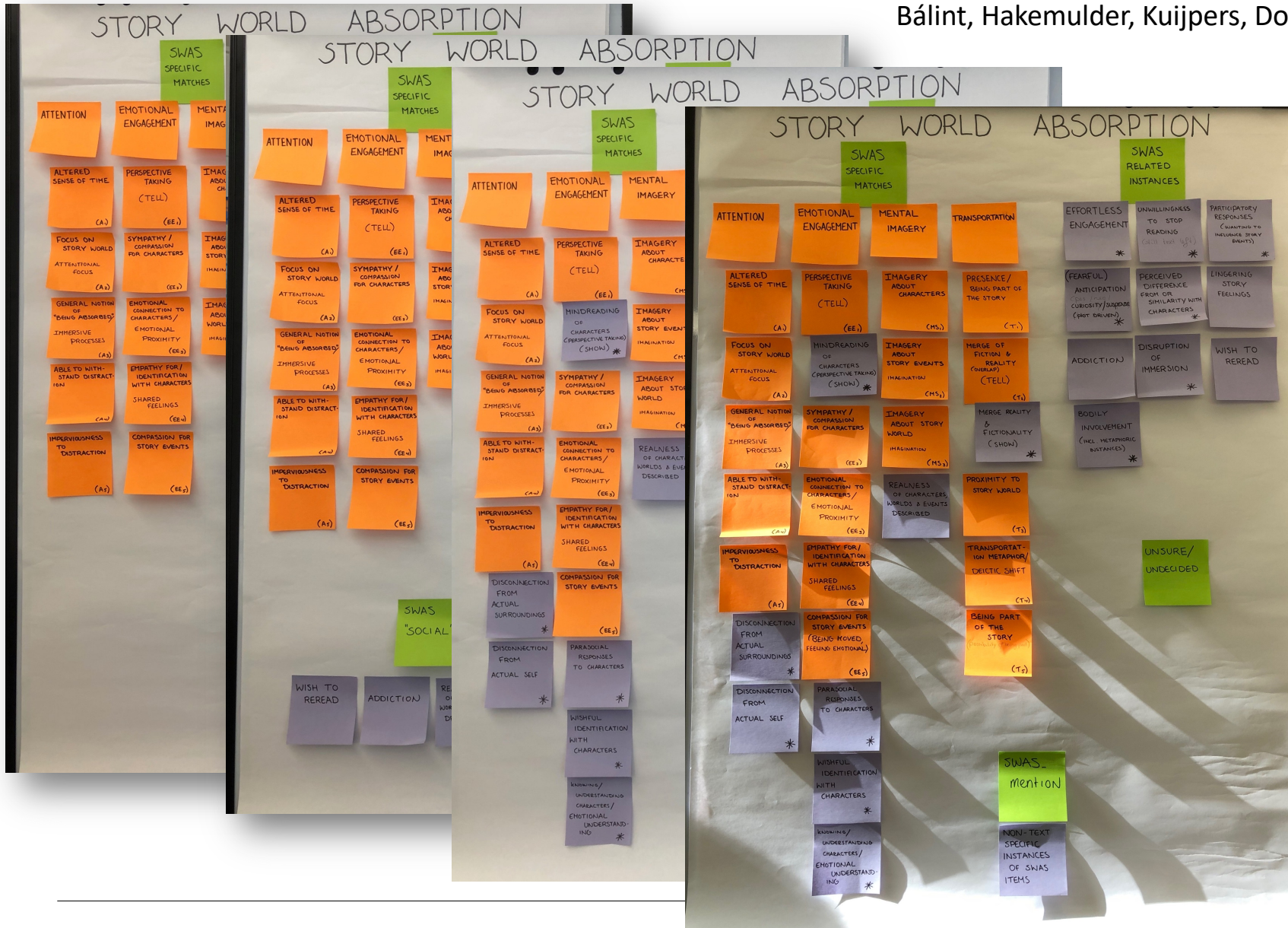
Story World Absorption	Attention	A1: When I finished the story I was surprised to see that time had gone by so fast
		A2: When I was reading the story I was focused on what happened in the story
		A3: I felt absorbed in the story
		A4: The story gripped me in such a way that I could close myself off for things that were happening around me
		A5: I was reading in such a concentrated manner that I had forgotten the world around me
	Emotional Engagement	EE1: When I read the story I could imagine what it must be like to be in the shoes of the main character
		EE2: I felt sympathy for the main character
		EE3: I felt connected to the main character
		EE4: I felt how the main character was feeling
		EE5: I felt for what happened in the story
	Mental Imagery	MS1: When I was reading the story I had an image of the main character in mind
		MS2: When I was reading the story I could see the situations happening in the story being played out before my eyes
		MS3: I could imagine what the world in which the story took place looked like
	Transportation	T1: When I was reading the story it sometimes seemed as if I were in the story world too
		T2: When reading the story there were moments in which I felt that the story world overlapped with my own world
		T3: The world of the story sometimes felt closer to me than the world around me
		T4: When I was finished with reading the story it felt like I had taken a trip to the world of the story
		T5: Because all of my attention went into the story, I sometimes felt as if I could not exist separate from the story

English Corpus: The annotation process

- We trained five annotators:
 - Prescribed literature on the development of the Story World Absorption Scale (Kuijpers et al., 2014)
 - Workshop on absorption with experts from the field at the start of the project
 - 2 practice rounds of annotating reviews
 - Discuss annotation disagreements every week, as well as "undecided candidates"
- Instructions:
 - main criterion for assigning a tag is semantic or conceptual similarity
 - freely establish the boundaries of a relevant text segment
 - allowed to assign more than one tag to the same text segment
 - also indicate when users explicitly mention or signal a *lack* of absorption
 - also indicate when users mention non-reviewed-text-specific instances of absorption (*"I usually like to read books that captivate me from the very first page"*)

Tag set development

Bálint, Hakemulder, Kuijpers, Doicaru & Tan, 2016



The corpus



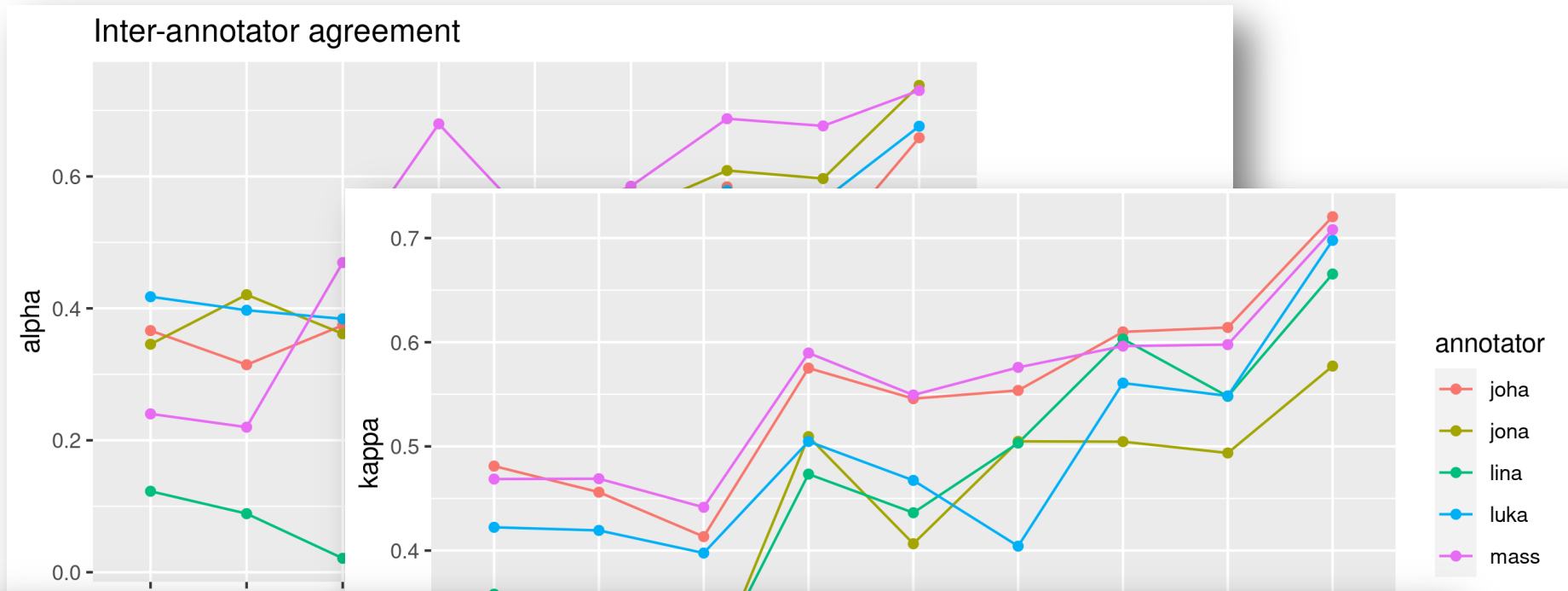
- Reviews from 6 different genres: fantasy, thriller, science fiction, romance, horror and mystery
- Pre-selected for “absorption-richness” (Rebora, Lendvai & Kuijpers, 2020; Lendvai et al., 2020)
- 5 stars
- Length
- keywords

Annotation tool: INCEpTION

The screenshot displays the INCEpTION web interface. At the top, the navigation bar includes 'INCEpTION', 'Projects', and 'Dashboard'. The main workspace is divided into three panels:

- Units Panel (Left):** A grid of unit numbers from 1 to 64. Units 16, 21, 26, 31, 46, and 61 are highlighted in green. Unit 14 is highlighted in purple. Units 22, 23, 24, 25, 27, 28, 29, 30, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 62, 63, and 64 are in white.
- Main Text Panel (Center):** Displays text from a document titled '12.02767052_the_hunger_games.00028840.txt'. The text includes sentences 21 through 28. Annotations are visible as colored boxes with labels like 'Absorption present | Lingering Story Feelings', 'Absorption present | Inability to Stop Reading', and 'Absorption present | A5'.
- Right Panel (Annotations):** Shows the configuration for the selected annotation layer 'Absorption'. It includes a 'Text' field with the snippet 'I read this series in a week-two or three days, to be exact-and I couldn't be bothered with life, my husband, or ANYONE.' and a 'Presence/Absence' dropdown menu currently set to 'Absorption present'.

Annotation Process: Inter-Annotator Agreement and Disagreement



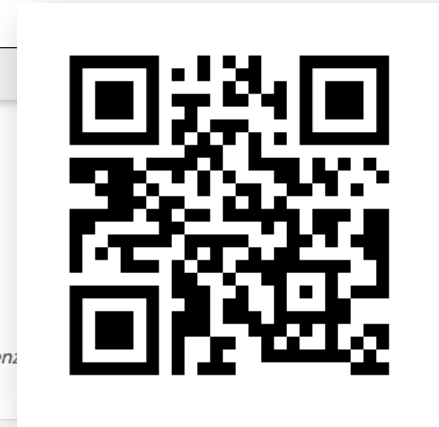
Couple: ('luka', 'mass')	Agreement:	0.6452718907107708
Couple: ('joha', 'mass')	Agreement:	0.415275720990939
Couple: ('joha', 'luka')	Agreement:	0.3986447824122873
Couple: ('joha', 'jona')	Agreement:	0.0
Couple: ('jona', 'lina')	Agreement:	0.0
Couple: ('jona', 'luka')	Agreement:	0.0
Couple: ('jona', 'mass')	Agreement:	0.0
Couple: ('joha', 'lina')	Agreement:	-0.0008592541673826393
Couple: ('lina', 'luka')	Agreement:	-0.0010616462594663645
Couple: ('lina', 'mass')	Agreement:	-0.0010886380564947284
Mean:		0.14561828556306536

Story World Absorption	Attention	A1 (Altered sense of time): <i>While reading time moved differently</i>
		A2 (Concentration): <i>My attention was focused on the book</i>
		A3 (General sense of absorption): <i>I was absorbed in the book</i>
		A4 (No distractions): <i>I was not distracted while reading</i>
		A5 (Forgetting surroundings): <i>While reading I forgot the world around me</i>
		A6 (Attention was grabbed): <i>The book pulled me in/gripped me</i>
		A7 (Anticipation): <i>I was on the edge of my seat / I wanted to know what would happen next</i>
		A8 (Inability to stop reading): <i>I did not want to put the book down/ I could not put the book down</i>
	Emotional Engagement	EE1 (Perspective taking): <i>I could imagine what it must be like to be this character</i>
		EE2 (Sympathy): <i>I sympathized with this character</i>
		EE3 (Emotional connection): <i>I felt a connection to this character</i>
		EE4 (Empathy): <i>I felt how this character was feeling</i>
		EE5 (Compassion for story events): <i>I felt for what happened in the story</i>
		EE6 (Anger): <i>I felt angry at this character</i>
		EE7 (Fear): <i>I felt scared for this character</i>
		EE8 (Knowing characters): <i>I felt like I knew this character</i>
		EE9 (Wishful identification): <i>I wish I could be more like this character</i>
		EE10 (Emotional understanding): <i>I understood why this character did this</i>
		EE11 (Parasocial response): <i>I want to have some kind of relationship with this character</i>
		EE12 (Participatory response): <i>I wanted to involve myself in the story world events</i>
Mental Imagery	MS1 (Imagery of character): <i>I could imagine what the characters looked/smelled/felt/sounded like</i>	
	MS2 (Imagery of story events): <i>I could see/hear/feel/smell the story events clearly in my mind</i>	
	MS3 (Imagery of story world): <i>I could imagine what the story world looked/smelled/felt/sounded like</i>	
	MS4 (Realness): <i>The character/story world felt real to me</i>	
Transportation	T1 (Presence): <i>While reading this I was in the story world</i>	
	T2 (Merge of fiction in reality): <i>Elements from the story world came into my world</i>	
	T3 (Proximity of story world): <i>The story world felt close to me</i>	
	T4 (Deictic shift): <i>I felt transported to the story world</i>	
	T5 (Part of the story world): <i>I felt part of the story world</i>	
	T6 (Return deictic shift): <i>I returned from a trip to the story world</i>	
	T7 (Travel in story world): <i>I lost myself in the story world / I travelled with the characters through the story world</i>	
Impact	IM1 (Effortless engagement): <i>It was an easy read / I devoured this book</i>	
	IM2 (Wish to reread): <i>I will/have reread this book/parts of this book</i>	
	IM3 (Anticipation book series/oeuvre): <i>I cannot wait to see how this unfolds in the next book</i>	
	IM4 (Addiction): <i>I am addicted to this book / I cannot get enough of this book</i>	
	IM5 (Lingering story feelings): <i>The book left me feeling ... / This book stayed with me for a while</i>	

Example from the English Language Guidelines

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.	

AbsORB metadata corpus and annotation guidelines



Absorption in Online Book Reviews

Contributors: [Moniek Kuijpers](#), [Simone Rebor](#), [Piroska Lendvai](#), [Massimo Lusetti](#), [Lina Ruh](#), [Johanna Vogelsanger](#), [Tina Ternes](#)

Date created: 2023-04-27 02:46 PM | Last Updated: 2023-07-04 04:47 PM

Identifier: DOI 10.17605/OSF.IO/KR4V6

Category: Project

Description: *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 Excellence Program.*

License: CC-BY Attribution 4.0 International

Wiki

Online book reviews posted on social media platforms are a relatively new form of reader testimonials that can be of use to researchers from different disciplines to investigate reading experience and evaluation, as well as social discourse about reading. In this project we developed and approach for annotating online book reviews using the Story World Absorption Scale (SWAS, Kuijpers et al., 2014...)

[Read More](#)

Files

Click on a storage provider or drag and drop to upload

Name	Modified
Absorption in Online Book Reviews	
- OSF Storage (Germany - Frankfurt)	
- AbsORB Annotation guidelines	
- OSF Storage (Germany - Frankfurt)	
AbsORB Annotation Guidelines - English - June 2023.pdf	2023-06-27 12:16 PM
- AbsORB metadata corpus of annotated book reviews	
- OSF Storage (Germany - Frankfurt)	

Citation

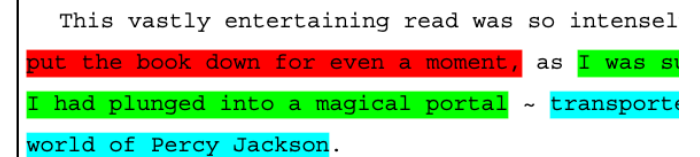
Components

- AbsORB Annotation guidelines**
Kuijpers, Rebor, Lendvai & 4 more
English language annotation guidelines to tag for mentions of absorption in reader response data.
- AbsORB metadata corpus of annotated book reviews**
Kuijpers, Rebor, Lendvai & 3 more
This is a corpus of English-language, annotated, online book reviews containing metadata only
- Preprint paper - Annotating for Absorption in Online Book Reviews**
Kuijpers, Rebor, Lendvai & 1 more
This is the preprint of a paper currently under review about the annotation process and the development of the annotation guidelines. It also includes...
- Preprint paper - Absorption in Online Reviews of Books. Presenting the English-Language AbsORB Metadata Corpus and Annotation Guidelines.**
Kuijpers, Rebor, Lendvai & 4 more

Developing German Language Annotation Guidelines

- Same group of annotators, native and non-native (Swiss-) German speakers
- Translation of all annotation categories (with reference to German language research on absorption, e.g., Appel & Richter)
- Throughout the annotation process we adjusted translations, based on examples from reviews

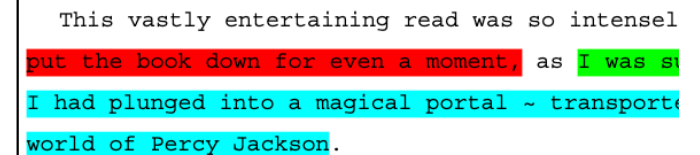
This also means that not all units of text receive labels. In fact, the units of texts in online reader reviews do not receive labels. However, in some cases, reviews are full of mentions of absorption. Annotators can assign multiple labels to one sentence, as seen in the example in Figure 1.



This vastly entertaining read was so intense I
put the book down for even a moment, as I was so
I had plunged into a magical portal ~ transported
world of Percy Jackson.

Figure 1. The green highlighted fragment is an instance of Attention (indicating absorption); the red highlighted fragment is another instance of Attention (indicating stop reading); and the blue highlighted fragment is an instance of Attention (indicating a shift).

This is allowed, as the concept of absorption is multi-faceted. When someone is discussing their state of absorption, they touch on multiple aspects within one sentence. The same fragment could also have been tagged with multiple labels. For example, “I had plunged into a magical portal” be tagged twice, once on its own and once as part of a larger tag for T4. Whether a sentence uses “I”, “You”, or “The Reader”, as long as they are talking about an aspect of absorption.



This vastly entertaining read was so intense I
put the book down for even a moment, as I was so
I had plunged into a magical portal ~ transported
world of Percy Jackson.

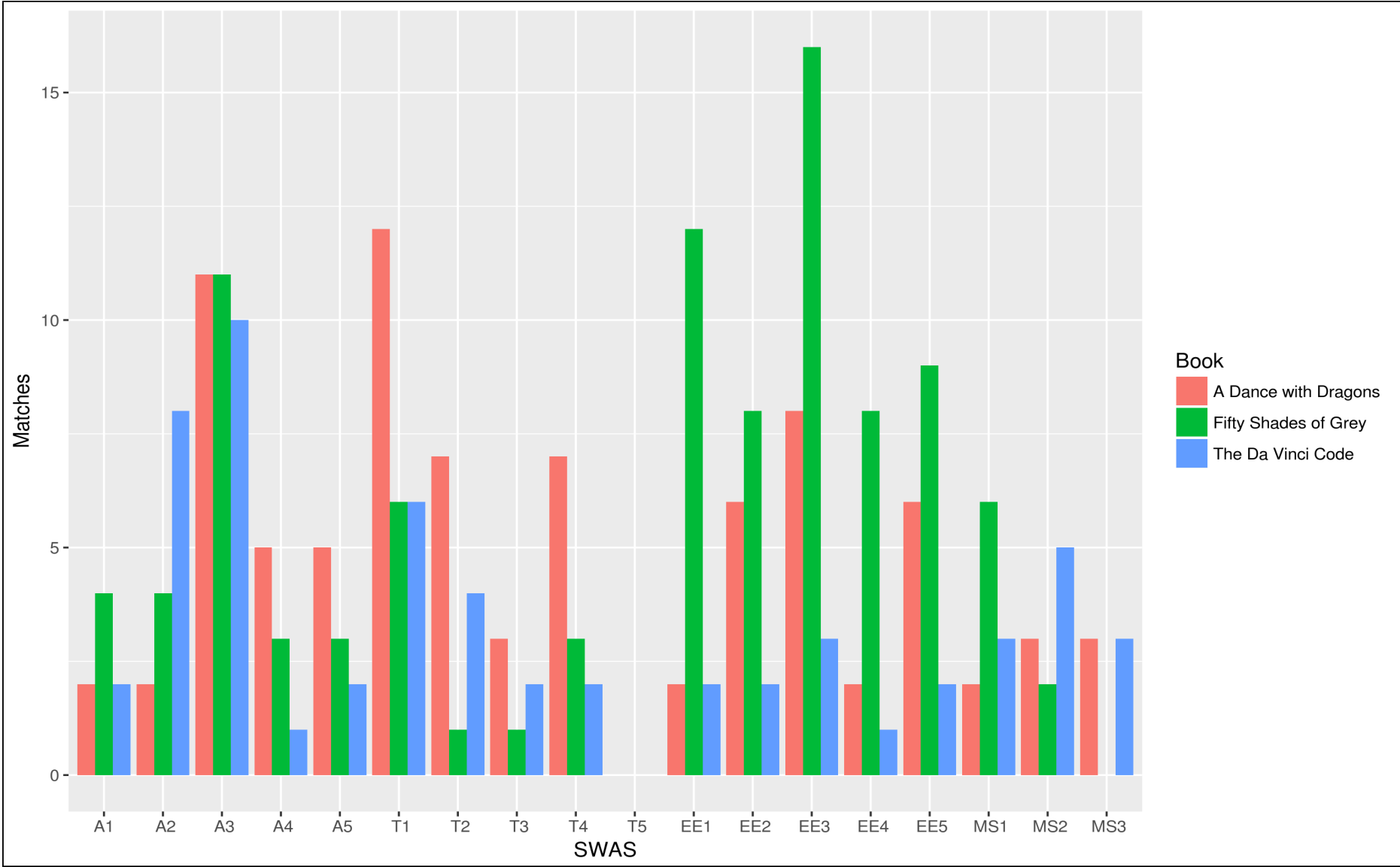
Example from the German Language Guidelines

Tag ID	Conceptualization	Statement
EE2	Mitgefühl (Sympathie)	Ich habe Mitgefühl/Mitleid für diesen Charakter
Review example POS	"wie nahe mir die Protagonisten gekommen sind und wie betroffen mich ihr Schicksal machte."	
Review example NEG	"weshalb ich nicht mitgefiebert habe <u>und mir letztendlich die Schicksale irgendwie egal waren.</u> "	
Comments	<p>Tag EE2 (Mitgefühl/Sympathie), ist für Beschreibungen von Mitgefühl oder Mitleid. Dabei ist es wichtig, dass man solche Ausdrücke ("Ich hatte großes Mitleid mit diesem Charakter", "er tat mir richtig leid" etc.) klar trennt von EE4 (Mitfühlen/Empathie), wo oft die Formulierung "ich konnte mit den Charakteren mitfühlen/mitlachen/mitleiden" vorkommt. Bei EE4 geht es darum, dass der*die Leser*in die gleichen Gefühle wie die Charaktere fühlt, bei EE2 hingegen darum, dass der*die Leser*in dem Charakter gegenüber Mitleid (Mitgefühl, etc.) o.Ä. empfindet.</p> <p>"Mitleid haben" ist eine versteinerte Phrase, "mitleiden" wird als EE4 getaggt.</p>	

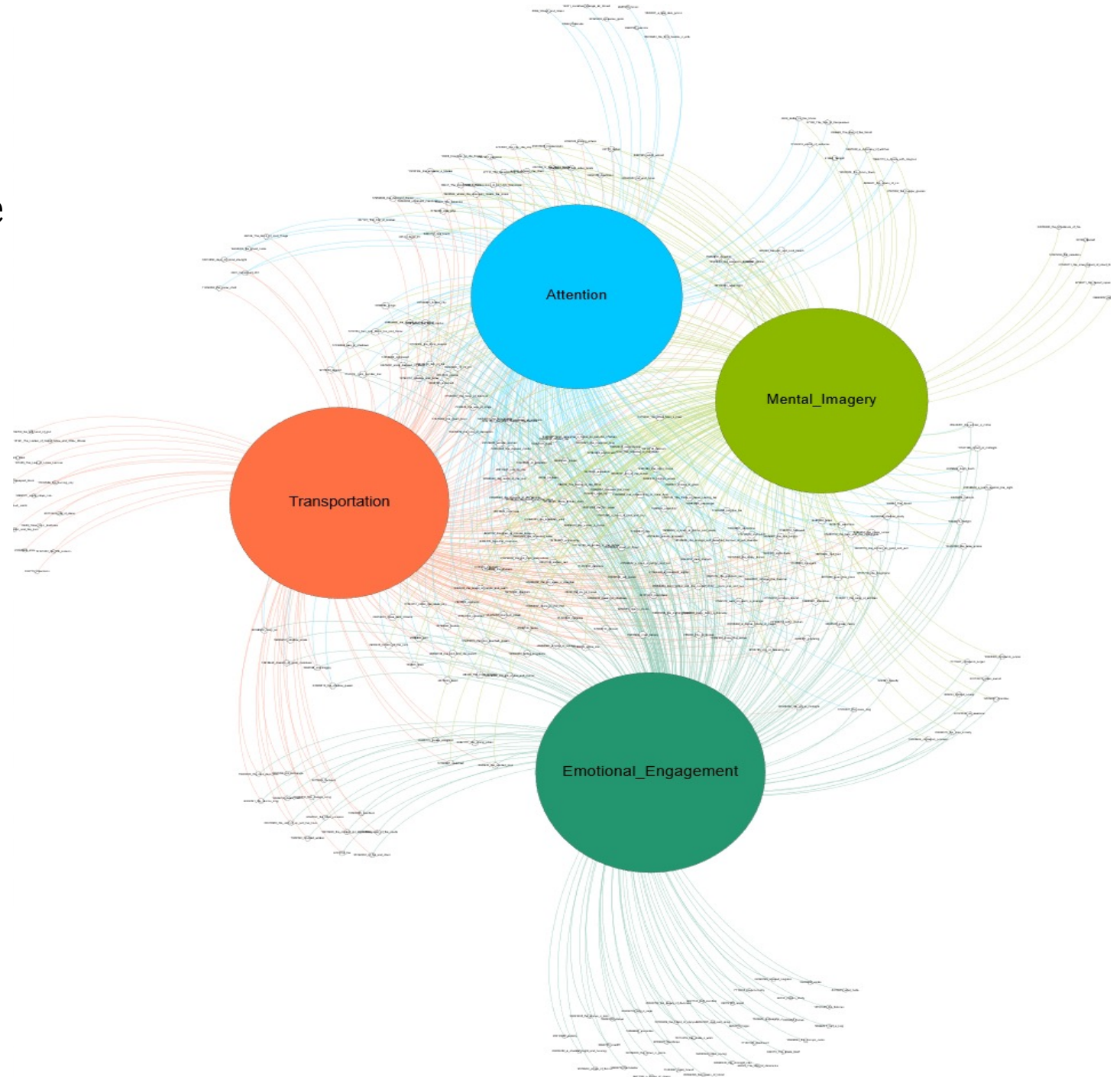
Examples of further research: Cross cultural comparison

	Statement	English Absorption Present	German Absorption Present	English Absorption Negated	German Absorption Negated
ATTENTION	A1 While reading time moved differently	3	6	0	0
	A2 My attention was focused on the book	10	4	2	1
	A3 I was absorbed by the book	186	283	8	33
	A4 I was not distracted while reading	3	?	2	?
	A5 While reading I forgot the world around me	16	2	4	1
	A6 I wanted to know what would happen next	108	152	3	24
	A7 I did not want to put the book down	145	136	5	23
EMOTIONAL ENGAGEMENT	EE1 I could imagine what it must be like to be this character	35	45	0	13
	EE2 I sympathized with this character	53	23	4	4
	EE3 I felt a connection to this character	69	115	10	15
	EE4 I felt how this character was feeling	72	67	1	3
	EE5 I felt for what happened in the story	79	84	0	4
	EE6 I felt angry at the character	19	11	1	0
	EE7 I felt scared for the character	5	6	0	0
	EE8 I felt like I knew this character	11	22	0	1
	EE9 I wish I could be more like this character	8	5	0	0
	EE10 I understood why this character did this	26	48	5	22
	EE11 I want to have some kind of relationship with this character	79	22	0	0
	EE12 I wanted to involve myself in the story events	42	10	0	0
MENTAL IMAGERY	MS1 I could imagine what the characters looked like	15	7	3	0
	MS2 I could see the story events clearly in my mind	20	8	0	1
	MS3 I could imagine what the story world looked like	25	51	2	0
	MS4 The character/story world felt real to me	73	61	0	8
TRANSPORTATION	T1 While reading this I was in the story world	13	18	0	0
	T2 Elements from the story world came into my world	14	8	0	0
	T3 The story world felt close to me	4	15	0	0
	T4 I felt transported to the story world	19	54	0	0
	T5 I felt part of the story world	34	23	0	1
	T6 I returned from a trip to the story world	3	1	0	0
	T7 I traveled with the characters through the story world	26	30	0	0
IMPACT	IM1 It was an easy read/I devoured this book	68	100	40	79
	IM2 I will reread this book/I have reread this book	112	26	4	2
	IM3 I cannot wait to see how this story unfolds in the next book	167	176	3	4
	IM4 I am addicted to this book/I cannot get enough	89	21	2	1
	IM5 This book stayed with me	192	83	2	2

Examples of further research: Genre Differences



Examples of further research: Core experiential responses to corpora of literary texts



Linking reader responses to corpora of literary texts?



- Use sentiment analysis to analyze whether sentiments expressed in novels are echoed in what readers talk about (e.g., are themes picked up on)
- Investigate the relationship between the use of literary features in books and emotional responses from readers (e.g., absorption, empathy)
- Investigate whether changes in an author's literary style over time, are reflected in readers responses to their novels (e.g., popularity)



Limitations

- Data needs to be annotated to be useful or be able to answer interesting questions
 - Labor-intensive and expensive
- Corpora consist of personal data, used without consent, but anonymized
 - Unclear as of yet, what data protection and copyright laws are to be heeded when using this type of data for research purposes
 - Reviews are on books that are mostly still copyrighted
- In general, how to make such an endeavour sustainable
 - Funding for long-term maintenance is needed
 - Long-term peer review is required

CORE-LITES Network (COST-Action proposal)

Phase 1: Theoretical articulation

- Deciding on concepts to study in different modalities
- Deciding on study protocols

Phase 2: implementation

- Setting up repositories for protocols, measures, materials
- Data collection
- Setting up educational program for undergraduate students and training for researchers

Phase 3: Assessment

- Evaluation of the efficacy of the research, repositories and education/training
- Making plans for sustainability

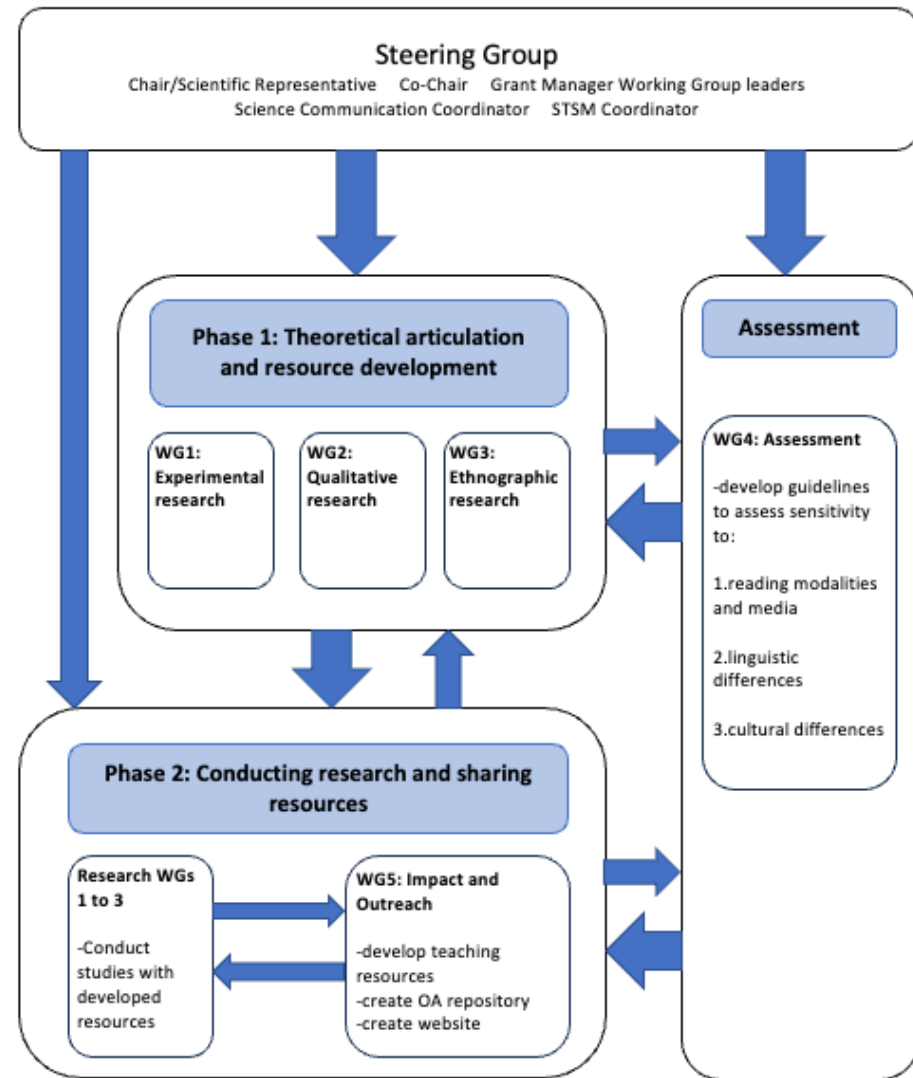


Figure 1. Organizational structure



University
of Basel

Thank you
for your attention.

Contact me:

Moniek.Kuijpers@unibas.ch

www.moniekkuijpers.com

And thanks to my collaborators:

Simone Rebora, Piroska Lendvai, Lina Ruh,
Johanna Vogelsanger, Massimo Lusetti,
Tina Ternes & Antonia Vogler

Output of “Mining Goodreads” and “SHARD” projects

Resource creation (scraping and pre-selection of absorption-rich reviews)

- 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)*, 271-272.
- Lendvai, P., Darányi, S., Geng, C., Kuijpers, M. M., Lopez de Lacalle, O., Mensonides, J.-C., Rebora, S., & Reichel, U. (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.
- 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*.

Tools creation (Annotation guidelines and metadata corpus)

- Kuijpers, M. M., Lendvai, P., Lusetti, M., Rebora, S., Ruh, L., Tadres, J., Ternes, T., & Vogelsanger, J. (2023). Absorption in Online Reviews of Books: Presenting the English-Language AbsORB Metadata Corpus and Annotation Guidelines. *Journal of Open Humanities Data*, 9(13), pp. 1–7.

Annotation (process and alternative discourse analysis methods):

- Kuijpers, M. M. (2022). Bodily involvement in readers’ online book reviews: Applying Text World Theory to examine absorption in unprompted reader response. *Journal of Literary Semantics*, 51(2), 111-129.
- Kuijpers, M. M., Lusetti, M., Rebora, S., & Lendvai, P. (2024). Validation of the Story World Absorption Scale through annotation of online book reviews. *Journal of Cultural Analytics*.

References

- Bálint, K., Hakemulder, F., Kuijpers, M., Doicaru, M., & Tan, E. S. (2016). Reconceptualizing foregrounding: Identifying response strategies to deviation in absorbing narratives. *Scientific Study of Literature*, 6(2), 176–207. <https://doi.org/10.1075/ssol.6.2.02bal>
 - Goodreads 2022. www.goodreads.com (Accessed on 28 March 2022).
 - Kuijpers, M. M., Hakemulder, F., Tan, E. S., & Doicaru, M. M. (2014). Exploring Absorbing Reading Experiences: Developing and Validating a Self-Report Measure of Story World Absorption. *Scientific Study of Literature* 4 (1), 89-122. doi: 10.1075/ssol.4.1.05kui.
-



University
of Basel

Additional slides

Validation of SWAS

- Does the SWAS reflect absorbing reading experience to the best of its abilities?
- Does it encompass all of the theoretical aspects of absorption?
- Has absorption (or the scholarship on absorption) changed since the scale was originally developed?
- Should it encompass aspects of absorbed reading inspired by current reading practices?
- Can the questions be phrased in a more understandable or relatable way?
- Is straight translation of measuring instruments feasible?

Goodreads examples in English and German

The Wise Man's Fear (The Kingkiller Chronicle, #2)

There is so much one can say when reviewing a book written by one's favorite author. I LOVE ROTHFUSS! He has created a world that will forever remain a place I want to go to. I want to talk to the characters, eat some bread and cheese with Kvothe, Sim, and Wil. Hear the beauty that is Kvothe's music and singing. While I can make these things appear and happen in my mind, there is a desire to make it real. When you get down to it...It's like Hogwarts. It's a world I can escape to. I keep it protected in the deepest part of my heart because it is so dear to me. It has become a part of me.

The Name of the Wind (The Kingkiller Chronicle, #1)

Ich hab so etwas noch nie gelesen und werde höchstwahrscheinlich sowas nie wieder lesen. Noch nie konnte mich ein Buch so dermaßen in den Bann ziehen und festhalten, dass ich nicht mehr davon loslassen konnte. Dieses Buch ist wie ein Droge - es macht unglaublich süchtig und man will immer mehr. Dieser Weltenaufbau ist die reinste Wucht und dennoch so verständlich und klar, als wäre ich selbst schon mal da gewesen.

„Der Name des Windes“ ist das beste Buch, was ich je gelesen habe. Und das erste Buch was ich re-readen werde und nicht nur einmal.

Discussion of differences between English and German absorption expressions and guidelines

- The German reviews use more synonyms than the English reviews
- Eintauchen is often used to denote general absorption in the overall book, but has a movement connotation that is often used in Transportation categories

