jcls. Journal of Computational Literary Studies

conference version

From Readers to Data

Uncertainty in Computational Literary Citizen Science

Gilad Aviel Jacobson[1]
Itay Marienberg-Milikowsky[2]
Yael Dekel[2]

1. Department of Humanistic Studies, Shalem College, Jerusalem, Israel.
2. Department of Hebrew Literature, Ben Gurion University of the Negev, Beer-Sheva, Israel.

OPEN ACCESS

Citation
Gilad Aviel Jacobson, Itay Marienberg-Milikowsky, and Yael Dekel (2025). "From Readers to Data. Uncertainty in Computational Literary Citizen Science". In: *CCLS2025 Conference Preprints* 4 (1). 10.26083/tuprints-00030142

Date published 2025-06-17
Date accepted 2025-04-17
Date received 2025-02-07

Keywords
Citizen Science, Phenomenology, Reader Response, Uncertainty, Hebrew, Novel

Note
This paper has been submitted to the conference track of JCLS. It has been peer reviewed and accepted for presentation and discussion at the 4th Annual Conference of Computational Literary Studies at Krakow, Poland, in July 2025. Please check jcls.io for the final journal version.

Abstract. We examine uncertainty in computational literary citizen science by analysing The Hebrew Novel Project, a large-scale initiative collecting reader interpretations of Hebrew novels. While citizen science projects typically treat uncertainty as noise, we demonstrate the value of treating it as meaningful data. Through statistical-phenomenological analysis of 1,026 questionnaire responses from 349 readers, we study how readers express uncertainty, from simple question-skipping to explicit rejection of interpretive frameworks. We uncover theoretically meaningful uncertainty patterns - certain literary concepts consistently elicit more uncertainty than others, and individual readers show varying but consistent levels of epistemic humility across different aspects of literary interpretation. We argue that this "productive uncertainty" provides insight into both the nature of literary texts and the process of reading, suggesting new directions for computational literary studies that embrace interpretive ambiguity. By taking uncertainty seriously, citizen science projects can address a wider scope of interpretive phenomena while maintaining methodological rigour.

1. Introduction

Citizen science has been increasingly used in research in recent years, primarily in the natural sciences but also in the social sciences and, more recently, in the humanities (Tauginienė et al. 2020). Despite the fundamental differences between these fields, most studies of citizen science share a common characteristic: they aim to address problems by acquiring data that is difficult to obtain otherwise. This is evident, for example, in projects that involve the public in documenting and counting of species in nature, in characterising galaxies (Galaxy Zoo), in river monitoring initiatives, and similar endeavors (Dickinson et al. 2010, Haklay 2012).

Nonetheless, citizen science is not limited exclusively to crowdsourcing. As researchers have demonstrated, public involvement often includes broad exposure to scientific findings, active contributions to additional stages of the research process, and sometimes even full participation in the entire research process (Wiggins and Crowston 2011). Moreover, even with minimal components of crowdsourcing, citizen science projects often demonstrate additional values that extend beyond the narrow academic scope: It empowers participants, individuals as well as communities, involving them in the

research and thereby serving the community in different ways (Bonney et al. 2016). A community participating in the process of monitoring a river, or in bird counting, may cultivate a closer connection to nature and assume greater responsibility for its interactions with the environment (Brossard et al. 2005). This often serves as one of the primary justifications for choosing this method, despite the various challenges it presents from a narrower traditional academic perspective, be they methodological, conceptual, or ethical. Taking this into account, from a scholar's perspective, citizen science may also foster a sense of contribution to the world beyond the purely scientific achievements.

While this notion appears fairly intuitive in environmental citizen science projects and similar initiatives, it is less apparent in the humanities, where the engagement with data, quantitative methods, as well as collaborative work has traditionally been limited. Digital Humanities - applying digital tools and computational methods to the study of the humanities - is opening new avenues in this regard as well (Tauginienė et al. 2020), especially in cases in which citizen science projects are related to the preservation of cultural heritage. In such cases, citizen science projects provide the public involved in it with symbolic rewards, similar to those associated with nature conservation. People who assist in deciphering medieval manuscripts or in tagging cultural objects are not merely contributing data; they are also helping to preserve culture.[1]

Although it is not a cultural heritage preservation project in the traditional sense of the term, this insight informed the development of the Hebrew Novel Project (Dekel and Marienberg-Milikowsky 2021), from which the data presented and analyzed in this article is taken. The Hebrew Novel Project aims to collect comprehensive literary data (poetic, thematic, and bibliographic) on Hebrew novels, from the first published novel in the mid-19th century to the present day. The project employs reader questionnaires relating to a corpus of approximately 8,500 novels spanning various genres, places of publication, and positions with regards to the literary canon. When launching the project in 2020, we sought not only to engage the public in data collection but also to foster a deeper understanding of the novel as a social phenomenon. We aimed at studying this using a methodology that actively engages with some of its social dimensions. This endeavor resonated with the public, as demonstrated by significant media and social media interest, surpassing that of the academic community.

However, we soon encountered a significant challenge, invoked by some of the responses to the questionnaire: we received complex contributions that not only provided valuable answers to our questions, but occasionally - in the designated slots in the questionnaire - also engaged with some of the items in a critical manner. Thus, reader responses reflected their knowledge of the novel they reported on, but also various forms of uncertainty, indeterminacy, epistemic humility and more. This is not surprising: In most citizen science projects, contributor uncertainty is usually reported (e.g. Tikkoun Sofrim project, Wecker et al. 2019), and used for discarding data points, or channeling ambiguous data to experts, with the general aim being to reach a resolution. But the unease that the contributor experiences is informative in and of itself, especially in such a complex questionnaire. Such unease may reflect lack of knowledge about the answer

1. A good example for this is found in the Tikkoun Sofrim Project, which uses crowdsourcing to train algorithms to recognize Hebrew handwriting in medieval manuscripts. See: https://tikkoun-sofrim.firebaseapp.com/en

or the terminology used in the question; it can also reflect an ambiguity in the novel that does not lend itself easily to the questions asked or the answers provided. In this paper, then, we shift our gaze from the typical questions and answers to the ones usually neglected, and ask: how does one derive meaningful research value from a systematic documentation of uncertainty?

The field of empirical literary studies (hereinafter: ELS) offers a different framework for thinking about our project. Dating back to the 19th century (Salgaro 2021), contemporary research has seen a growing interaction between ELS and various aspects of computational literary studies (hereinafter: CLS).[2] With the increasing integration of citizen science practices in CLS, this development is intensifying (Salgaro 2021, 538; Herrmann et al. 2021; Rebora et al. 2021): The combination of data-oriented research, quantitative analysis, and real readers' reactions to literature brings the two scholarly domains closer together, even if their objects of study, their methodologies, and their paradigmatic emphases may differ. This process is still in its early stages, and there is much to be done: We believe, for example, that scholars of CLS have a lot to learn from the well-established use of questionnaires in ELS, and the conceptual framework of cognitive poetics. At the same time, we are hesitant of the outright rejection of interpretive subjectivity that is sometimes advocated in ELS.[3] In this regard, the affiliation of CLS with non-computational and non-empirical approaches to literary study provides, for us, an important balancing anchor.

Indeed, while ELS engages with real readers and the ways they interact with literature, other approaches—particularly those widespread in the second half of the 20th century—tend to focus on abstract constructions of theoretical readers. For these approaches, uncertainty is viewed as a hypothetical reaction to the object of study. As is well known, some of the most influential schools of literary studies—from reader response criticism to post-structuralism—celebrate interpretative freedom, over-interpretation, ambivalence, and disagreement in different ways. Similar notions had already influenced literary studies earlier, notably in the work of Roman Ingarden and particularly his concept of indeterminacy, which he saw as inherent to literature due to its attempt to represent real objects.

Thus, the scope of our study, its grounding in the community of readers, its ambition of creating a 'democratic' database of novels and the reactions that they evoke, and, lastly, the inevitable computational analysis of the results, point to a complex negotiation between various interpretive traditions and data-driven approaches. On the one hand, it embraces the appeal of a plurality of interpretive voices; on the other hand, it imposes a normalizing framework on them. When it relates to readers as a resource for data collection, deliberately limiting their interpretive freedom by providing a structured mechanism for collecting the data, it faces a clear challenge vis-à-vis some of the traditional intellectual conventions. However, allowing space for uncertainty, and treating indeterminacy as valuable data – and not just as noise, as something to be regulated, validated or simply deleted – can bring the Hebrew Novel Project closer, in some senses,

2. The various activities of the International Society for the Empirical Study of Literature (IGEL) and its journal (Scientific Study of Literature SSOL), are all worthy of consideration, when wishing to integrate citizen science methods and goals within CLS.
3. For instance, Dixon and Bortolussi 2011 assert that "scientific methods require that observations be repeatable, and this requirement rules out subjective analyses that vary across individuals" (p. 65).

to traditional literary studies.

There are more difficulties in the implementation of citizen science in literary studies, and specifically in CLS. First, in clear contrast to literary studies as shortly described above, computational research often treats data, at least in its processed form, in a robust manner, as if it were transparent and free of interpretive biases (Piper 2020). Second, in CLS research that relies on annotations (by expert researchers or trained assistants), the norm of an extensive work with annotation guidelines while striving for inter-annotator agreement has been justifiably established (Gius et al. 2021). Thus, in addition to the consideration of uncertainty as data, the introduction of a less-controlled project, driven by amateur contributions, seems to undermine the very foundations of the field's (traditional as well as computational) interpretative concepts; it resonates with past schools of literary theory and criticism (formalism, structuralism) as well as with the concept of indeterminacy as suggested by Ingarden (Ingarden 1973).[4]

But if answers to a highly detailed questionnaire dedicated to the characterization of complicated literary phenomena reflect, to some extent, indeterminacy, what should one do with such data, often considered as noisy or messy? A widespread tendency is to focus on agreed, validated information, to adjust and normalize disagreement, or to ignore uncertainties in different ways (e.g., using reports of uncertainty to redirect data to experts, enlarging the number of reports for those data to allow estimation of some underlying "consensus"). In some of the outcomes of the Hebrew Novel Project, we, too, strive for the agreed. However, in the present article, we choose to celebrate indeterminacy, treating it not as a potential source of noise in the data, but rather as a source of knowledge. Based on this, we seek to conceptualize indeterminacy in a way that will show its benefits to our project as well as other studies.

The next part of the article will be devoted to a brief review of the use of citizen science in CLS. We will then provide a detailed description of the approach we developed in The Hebrew Novel Project. Following that, the article will delve into a few specific findings, highlighting indeterminacy in response to a variety of items in the Hebrew Novel project's questionnaire. Lastly, we will turn to discuss the findings, using a statistical-phenomenological approach.

2. Computational Literary Studies and Citizen Science

The integration of citizen science into the humanities is still in its infancy, and, as noted earlier, is used primarily in digital humanities and more specifically in contexts of cultural digital preservation. Its presence in the subfield of CLS is still scarce, found only in a handful of innovative projects. These projects — some of which we will present here — can be seen as the beginnings of a new scholarly direction, which we propose to call *Computational Literary Citizen Science* (hereinafter: CLCS), linked also to the well-established tradition of ELS. Most of these projects draw on a relatively wide community of non-professional readers, keeping the task simple, sometimes referring to sociological and demographic aspects of the project participants, and usually also combining the crowdsourced findings with various automated techniques. Yet, in many ways, these

4. A different issue that will not be discussed here is disagreement between different readers of the same novel. We reserve this discussion for further accounts.

projects also differ from one another, and examining these differences will help us better situate our own work.

A recent example, "The DisKo project" (Diversitäts-Korpus [diversity corpus]), led by Marie Flüh, Mareike Schumacher and Peter Leinen, involves the use of citizen science to collect titles of novels that feature various non-binary gender representations.[5] This is achieved through a short questionnaire that includes some demographic questions, a request to list relevant titles, and an option to provide comments. The goal of this ongoing project is to compile a sufficiently large list of books – one that could not be compiled without the assistance of many readers – for future annotation by a professional team that will explore methods for automatic identification of non-binary gender characters in literature.

While the DisKo project collects titles, *Project Endings*, led by Helena Michie, Robyn Warhol and Huw Edwards-Evans, asks readers to delve into books and collect structural elements.[6] This recent literary citizen science initiative invites the readers to choose a serial Victorian novel from a predefined list and mark, using a Google Forms questionnaire, the narrative's strategies for the ending and the beginning of each part of the serial novel. *Project Endings* is rooted in literary studies more than in digital humanities, and is described by the leading researchers as "a 'medium data' study [...] because no computer application could do the required analysis".

Focusing on an even smaller literary element, Andrew Piper and colleagues explicitly integrate citizen science and academic research (Piper et al. 2024). In this computationally ambitious project, participants are asked to identify predefined types of character interactions within specific sentences from contemporary literature. This task focuses on supporting and refining natural language processing (NLP) methodologies and on validating automated practices. The goal is to acquire accurate and objective information, with low-agreement findings used to improve model training. The tagging process requires minimal interpretation (only one sentence is annotated at a time), and the emphasis is on achieving high levels of agreement. A similar approach is used in another ongoing project by Piper, which focuses on annotating character emotions.[7] Both projects are disseminated through the Zooniverse platform, with the tagline: "Help us annotate literary characters to build AI that can better understand human storytelling." Thus, Piper's projects clearly demonstrate what appears to be a typical human-machine interrelationship: the primary goal of the human contribution is to improve the algorithm, and not necessarily explore the different human perspectives. In the end, the purpose of human annotation is to serve the machine, even if eventually, the computational results will serve the human. The results are noteworthy: "With respect to Citizen Science as a mechanism of crowd-sourced text annotation, we find annotation quality on par with trained student annotators. As prior work has suggested, Citizen Science projects achieve the same quality standards as other approaches and bring with them the affordances of a volunteer, community-based approach to scientific discovery" (Piper et al. 2024, 479). Following this success – in terms of data accuracy – the authors voice the hope that "more projects in NLP and DH will utilize this significant resource".

5. https://msternchenw.de/disko-das-diversitaets-korpus/
6. This is an ongoing part of a larger project on the Victorian novel, whose details are found here: https://readinglikeavictorian.osu.edu/
7. https://txtlab.org/2024/09/new-citizen-science-project-reading-emotions/

Although their focus differs, citizen science was employed in the three studies reviewed so far to obtain unambiguous data: to expand the corpus of literature featuring non-binary characters in the first case, to characterise beginning and ending strategies in the second, and to improve the accuracy of automated literature analysis models in the third.

The following study, which is actually the earliest, takes a different direction, one closer to that of the empirical study of literature. Karina van Dalen-Oskam's *The Riddle of Literary Quality* is an extensive two-stage citizen science project (Dalen-Oskam 2023). In the first stage, almost 14,000 readers filled out a survey about the subjective literary quality of contemporary Dutch and translated novels, from a list of best-selling novels. The second stage consisted of computational text analysis of the same novels. The survey (titled The National Reader Survey) was opened for seven months in 2013 and included sixteen questions, both demographic and pertaining to the participants' opinion on the literary quality of the novels they have read (Koolen et al. 2020). Interestingly, *The Riddle* did not use the term Citizen Science or similar terms. Moreover, it dealt with agreement and disagreement (notions that can be seen as related to some extent also to indeterminacy) as part of what can be described as the sociology of literature, actively creating a more diverse profile of respondents based on their gender and geographic location.

The focus of The Hebrew Novel Project is neither the reader, nor sociology of literature. The subjective perspective of its participants (whose demographic and sociological backgrounds are not made explicit in the questionnaire) is apparent in the data through its interpretive literary as well as thematic questions. The data arising from the project suggests a novel question: how does indeterminacy contribute to the research of literature itself?

3. The Hebrew Novel Project

The Hebrew Novel Project was born out of two seemingly contradictory intellectual passions: on the one hand, the urge to organize, to systematically map the entire large-enough yet not-too-large corpus of the Hebrew novel, and on the other, an impulse to disrupt, shown in the enthusiasm for the noise that arises from as many human thorough readings as possible. Interestingly, the tension has been particularly significant in the development of CLS, especially in light of the implicit dialogue between Franco Moretti's "Conjectures on World Literature" (Moretti 2000) and Erich Auerbach's "Philology of World Literature" (Auerbach 2012 [1952]). In short, while Auerbach was criticizing the very idea of a research based on collective work, Moretti proposed a research method based on second-order reading that therefore relies on more than one reader. In the Hebrew Novel Project we took this intention a step further, as both these scholars certainly did not consider literary research based on a *non-scholarly* community, a community of 'ordinary' readers whose variety of *different* readings include uncertainties – rather than a unifying synthesis that adjusts them. Our interest in these different readings is phenomenological, as we want to better understand what can be learned from indeterminacy as such. This phenomenological subjectivity resonates with Wolfgang Iser's understanding of the role of the reader in filling *gaps* in the text: Indeterminacies engage

the readers and require them to participate in the meaning-making of the text, a process that is highly subjective (Iser 1980).

Finally, in order to better understand the essence of the Hebrew Novel Project, we will describe its similarities and differences with other literary projects, traditional as well as computer assisted. First, the Hebrew novel project is not a close reading project. While in traditional literary studies the most widely accepted approach is that of close reading of individual texts, here we tackle a different problem – the Hebrew novel in general – by gathering data on as many texts as possible. Despite this, the Hebrew Novel project *is* actually based on close readings: the readers who participate in the project fill out an exhaustive questionnaire about a Hebrew novel they have recently read, and are advised to hold the book near them while answering the questionnaire. Most of the questionnaire items require participants to reflect on the novel, delving into some of its stylistic and thematic features. This is a form of second-order distant reading which we named elsewhere *distant public reading* (Dekel and Marienberg-Milikowsky 2021).

Second, as a whole, it is not a typical computational text analysis project. While computing power takes place in different stages of our project – from data gathering (with Google forms) to its statistical analysis (with Excel, R and MatLab) – it has no role in the reading itself. The reading is done by humans, without any algorithmic element, and part of our focus in analysing the reports is to highlight the individual readings that are attested to by the different contributions. It should be noted that while we have digital access to many of the novels, for the current article which focuses on the readers and their uncertainties, we are refraining from processing them with text analysis tools. It should also be noted that some of the other parts of the project rely more than the one presented here on text analysis techniques.

Third, in contrast to another common approach in computational literary studies, the Hebrew Novel Project is also not an annotation project in the usual sense of the term. Typical annotation projects aim both to enable distant reading and to document close reading. We, however, do not use in-line annotations at all, as the comments of those who participate in our project are not attached to specific textual segments; rather, the readers provide their structured feedback at the level of the entire novel (genre, plot, characters, time, space, etc.), and, to some extent, to its external circumstances (e.g., in questions of reception and importance). However, as we have argued elsewhere (Münz-Manor and Marienberg-Milikowsky 2023), the tension between describing a work as a whole and a detailed tagging of its text is a fertile tension for a more sophisticated annotation theory and practice.

As argued by Gius and Jacke, not all disagreements should be processed equally; some can (or should) be resolved but others not: "literary analysis should more often be inspired by the shared effort of agreed disagreement" (Gius and Jacke 2017, 251). The same can be said about uncertainties. Yet, within the framework of our project, we cannot judge the veracity of readers' claims, except in cases of a clear mistake (about some of the non-interpretative bibliographical data). Since the focus of the current paper is phenomenological, we are not concerned with the veracity of readers' responses. The question of errors, agreement and consensus may be dealt with in future papers, which will approach the same data through a different prism.

4. Findings

Our questionnaire was designed to collect data about several categories (bibliography, narratology, time and space, themes, language) using multiple-choice items, linearly scaled items, and a few short-answer questions that allow for more personal and interpretive free text responses. And yet, although the readers mostly choose the best option (or multiple options) out of a few given answers, many of these choices (or, all of them, except the bibliographic ones) depend on interpretation. While most of the questions are required and non-skippable, in a few cases, pertaining to complex literary concepts which nevertheless were explicated in the questionnaire, we allowed the readers to skip a question in cases of uncertainty. Thus, this structured questionnaire calls for interpretation, disagreement, ambivalence and indeterminacy.

It is important to note that the Hebrew Novel project was constructed as a Citizen Science project, and our sensitivity to reader uncertainty and ambiguity grew from studying the corpus of filled questionnaires. Therefore, the data analysed is uneven, in the sense that items provided heterogeneous opportunities for expressing uncertainty and ambiguity. The analysis should therefore be assessed for what it is: a demonstration of possible modes of expressing uncertainty and ambiguity, and the kinds of insights we may glean from them, while not providing an exhaustive exploration of all aspects of uncertainty and ambiguity relevant for each item.

We first demonstrate the simplest form of reader uncertainty manifested by skipping an item, as items in the questionnaire were occasionally skipped. The questionnaire, which contained 77 items in total of varying types, contained 9 scaled items (see appendix A). Of these 9 items, readers were allowed to skip 4 (see Fig. 1a):

- "How would you estimate the typical linguistic register of the novel (1: very low - 5: very high)?" (register l-h; n=13/987 skipped). Readers were requested to skip this item if the answer to the previous item ("was Hebrew a spoken language at the time the novel was written?") was negative. We therefore excluded such skips in our analysis, and only included skips in this item if the previous item was answered in the affirmative (n=987/1026).
- "To what extent does the plot leave gaps that the reader must fill using their knowledge or imagination"? (gaps; n=26/1026 skipped).
- "Where along the conventional-experimental axis would you locate the novel?" (conv.-exp.; n=53/1026 skipped).
- "To what extent, in your opinion, does the novel employ intertextuality?" (intertext.; n=106/1026 skipped).

These items elicited different degrees of skipping (1.3%-10.3%), which we interpret as expressing varying degrees of uncertainty or ambivalence. The uncertainty may result from unfamiliarity with the term (such as intertextuality, that while explained briefly in the questionnaire, is not necessarily familiar to the non-professional reader), a property of the novel, or its perception by the reader, that defies an easy response. In these scaled items, it is impossible to disentangle these disparate explanations, as the readers had no means of providing a more detailed account of the type of difficulty they encountered.

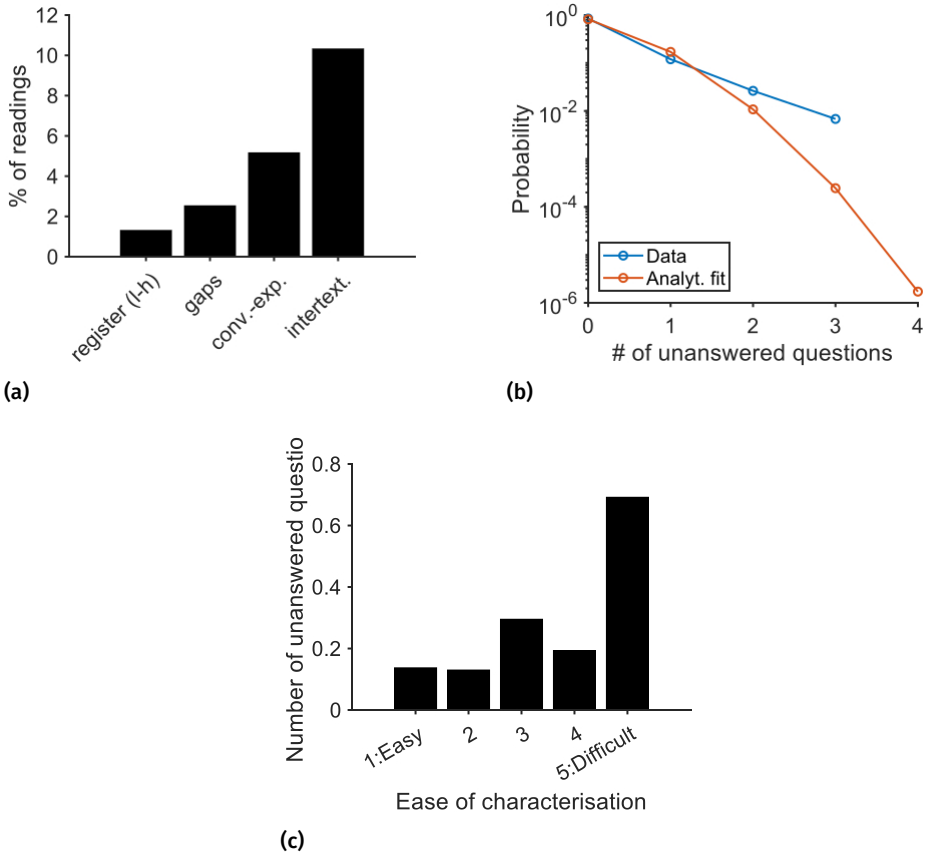


Figure 1: Reader ambivalence in scaled items. Four of the nine scaled items in the questionnaire allowed readers to express ambivalence by skipping the item. **(a)** Percent unscored questions across all questionnaires (n=1026), sorted from least to most skipped. **(b)** Probability of observing questionnaires with 0-4 skipped items. Blue line: real data. Red line: analytic fit, using marginals and assuming items are skipped independently. **(c)** Readers who found it most difficult to characterise the book using the questionnaire (5 on x-axis), skipped on average more scaled items (y-axis) (1-way ANOVA: $F=8.91$, $p<10^{-6}$).

Nevertheless, we were able to demonstrate that item skipping tends to cluster in certain questionnaires more than predicted by random distribution. To this end, we calculated the frequency of skipping each item across all questionnaires, and calculated the expected frequency of questionnaires with $0, \cdots, 4$ skips under the assumption that the skips are independent of each other. As shown in Fig. 1b, the data (blue line), when compared to the above calculation (orange line), shows an excess of questionnaires with 2 skips ($\times 2.4$) and 3 skips ($\times 27.8$). This indicates that item skipping within a questionnaire is correlated. Such correlations may arise either from a property of the readers (some readers exhibiting higher ambivalence, epistemic doubt, or lack of acquaintance with terminology, compared to others), or the novels (some novels eliciting more ambivalence in readers across different questions).

Finally, we asked whether item skipping exhibits a relationship to the last, reflective, question in the questionnaire, a scaled item in which readers were asked to report how easy it was for them to characterize the novel using the questionnaire. As seen in Fig. 1c, the mean number of skipped items tends to increase when the reported difficulty in novel characterisation increases. Thus, an explicit report of ambivalence was statistically linked to an implicit one-the number of skipped scaled items, with readers reporting the maximal difficulty (5 vs. $1-4$, post-hoc contrast after a one-way ANOVA test; $F = 8.91, p = 4.5 \cdot 10^{-7}$).

Next, we extended our characterisation of uncertain responses to a wider range of items, as readers were provided with different means of expressing uncertainty and ambiguity in different items. In some, there was no opportunity provided (e.g. multiple choice questions or scaled items that could not be skipped). In others, one or two of the answers that allowed readers to express their uncertainty or ambiguity (such as "unknown terminology", "hard to define") were provided. In items that contained the option for free text, readers could add other categories of uncertainty / ambiguity that were not offered to them.

To demonstrate the different kinds of ambiguity and uncertainty in the questionnaire, we analysed a subset of 23 items that represented the various item types: scaled items, numerical items, and various types of items providing multiple choice, free text, or combinations thereof. For the items with free text answers, we manually tagged all answers that reflected some degree of uncertainty or ambiguity. We then divided uncertain or ambiguous answers into nine categories, according to the common features they share:

1. No answer (the reader skipped answering this item).
2. "Term unknown" (uncertainty regarding question).
3. "It is unknown" (objective uncertainty regarding answer).
4. "Impossible to answer" (a more emphatic form of 3).
5. "Hard to define" (a less emphatic form of 3).
6. "I do not know" (subjective uncertainty regarding answer).
7. "I do not know" + informative answer.

8. "I do not remember accurately".

9. Rejection of question.

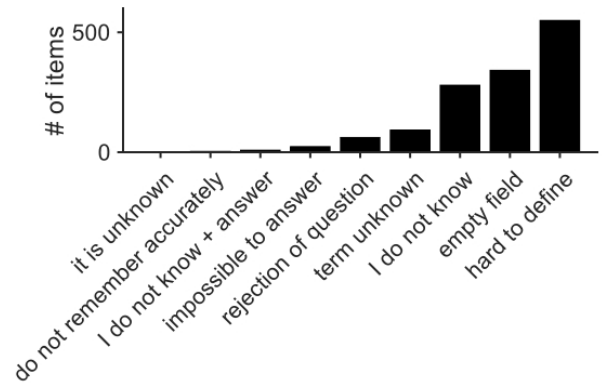
Figure 2a depicts the prevalence of these categories of uncertainty/ambiguity for the 23 selected items. Some categories were infrequent (category 3 (it's unknown): $n = 3$; category 8 (memory): $n = 4$), while others appeared with high frequency (category 5 (hard to define): $n = 550$; category 1 (no answer): $n = 342$). It is clear that expressions of uncertainty/ambiguity that were offered as options in the questionnaire, either implicitly (skipping) or explicitly (choosing an uncertain/ambiguous answer provided in a multiple-choice item) were much more frequent, while those that entailed free text were less frequent. We suspect that this difference is governed both by the additional effort required to conceptualise, phrase and write a free text answer, and by the heterogeneity across items, with many items not providing an option for free text.

Readers varied in the degree of uncertainty/ambiguity they expressed in the questionnaire (see Fig. 2b). Of the 23 items analysed, 340 questionnaires (33%) contained no item with the above indicators of uncertainty/ambiguity, 337 questionnaires (33%) contained a single such item, and the maximal number of uncertain/ambiguous answers was 10, in a single questionnaire. The mean number of uncertainty reports per questionnaire, restricted to the above 23 items, was (mean$\pm$ standard deviation) 1.3 ± 1.5.

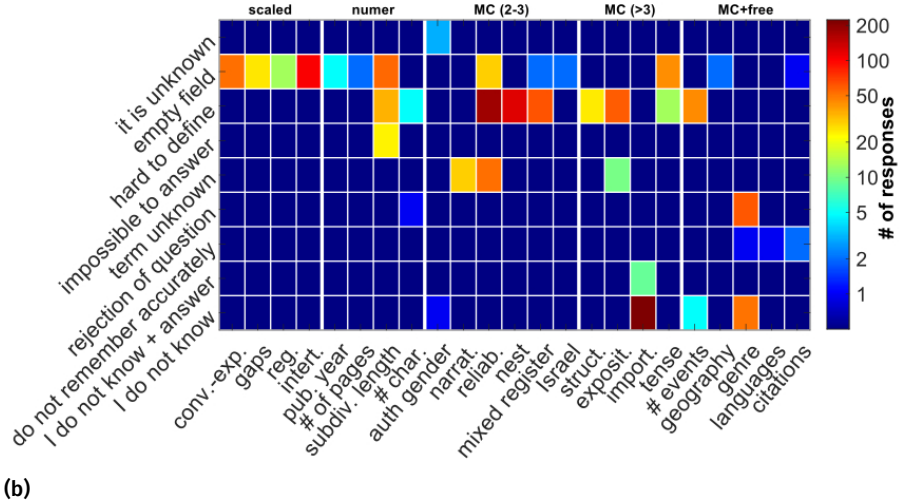
As explained above, the source of reader uncertainty is sometimes itself uncertain, and it is not always possible to determine if it stemmed from a property of the specific novel reported, the specific questionnaire item and the terminology it used, or from a property of the reader, assuming that different readers possess varying degrees of epistemic doubt. It is therefore informative that reports of uncertainty were not independently distributed across questionnaires. Like in Fig. 1b, statistical independence between reports of uncertainty would have resulted in almost no questionnaire with > 5 reports of uncertainty, and in our data, there is an excess of questionnaires with $6 - 10$ reports of uncertainty. This excess of uncertainty in some questionnaires may result from properties of the specific reader or the specific novel reported.

Last, we can see that within each item type, different items elicited varying degrees of uncertainty/ambiguity. Figure 2c summarises this data visually. Even within each item category, different items elicited varying degrees of uncertainty. For example, in the multiple choice questions with more than 3 suggested answers (MC (>3)), the item requesting readers to describe the tense of the narration elicited few instances of the uncertain response "hard to define" ($n = 13/1026$), while the item requesting readers to describe the location of the novel's exposition elicited almost a five-fold increase in the same response type ($n = 60/1026$). We must further stress that in the exposition item, we provided readers with yet another answer classified by us as uncertain / ambiguous: "I'm unfamiliar with the term". Thus, we can safely assume that in both these items, the "hard to define" answer reflects a difficulty in assessing the novel itself, and not in understanding the question, and that novels tend to ambiguate the location of the exposition more than ambiguating the grammatical tense.

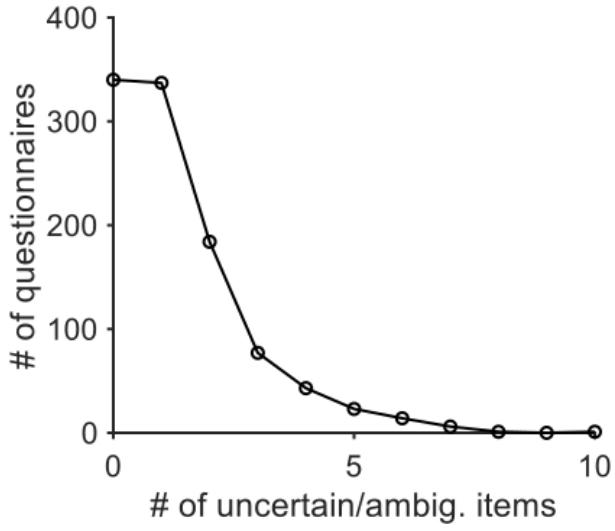
It is worth highlighting some of the reader contributions to the categories of uncertainty and ambivalence, which were provided in items that allowed free text answers. An interesting example is given in response to the item in which readers were asked to



(a)



(b)



(c)

Figure 2: Different modes of expressing uncertainty and ambiguity. Readers express their uncertainty in different ways. **(a)** Nine categories of uncertainty and their prevalence in the questionnaire corpus. **(b)** Distribution of the number of uncertain / ambiguous responses across all questionnaires ($n = 1026$; $mean \pm sd = 1.3 \pm 1.5$ per questionnaire, with a maximum of $n = 10/23$).**(c)** Distribution of different categories of uncertainty (y-axis) across 23 questionnaire items of different types. The types are separated by a thick, white vertical line. *scaled*: items with scale 1-5. *numer*: numerical items. *MC*: multiple choice items, further divided into binary/tertiary items (*MC (2-3)*); items with more than 3 choices (*MC (>3)*); and items that allow both MC and free text (*MC+free*).

report how many main and secondary characters the novel had. One reader wrote: "I claim that ... the novel is more complex than the framing of some of the questions aimed at. ... It is indeed possible that there is one main character and several secondary ones, but the structure of the novel challenges this by having the characters change parts ...". They thus questioned, or rejected, the relevance of the question itself, while still providing a hint to what their answer might have been if the issue was forced. This response was thus categorised as "rejection of question". Another item that elicited answers in the same category requested readers to describe the sub-genre of the novel. Four different readers replied with answers that rejected the suggested genres, and one even doubted the book is a novel (e.g. "none of the definitions [suggested] is accurate"). While a free text option may complicate analysis, and is often avoided in multiple-choice questions, the examples discussed suggest that they allow contributors not only to provide what they perceive as accurate answers, but also to comment on their own unease, uncertainty and ambivalence.

5. Discussion: Assessing Uncertainty— A Statistical-Phenomenological Approach

Integrating citizen science into projects whose primary objective is to collect data that cannot be efficiently gathered by other means, seems quite natural. In the so-called information age, where so many have access to the internet, and scientific endeavors are more data-driven than ever, it simply makes sense. The challenges arising from this method, as noted in the introduction, are offset by the advantages of its non-scientific added value. It is not surprising, then, that when citizen science has been integrated into CLS, it has primarily been used to collect data, and often in ways that contributed not only to the scientific work in the narrow sense of the term. It is also not surprising that in some cases that were described above in detail, it has been done in order to support computational work, in one way or another. However, we believe that this new research strategy offers an opportunity not only to collect—and preserve—cultural data, and not only to build a *useful* datasets that will enhance computational findings, but also to re-examine the *role* of data in CLS; and, more specifically, to rethink the place of reading-based data, in relation to prominent currents of literary criticism in the past century, whether empirical-oriented or theoretical-oriented. This approach can challenge how we perceive textual content, much like Ingarden's indeterminacy theory, Iser's (and others') reader-response criticism, and French poststructuralism have done before. By doing so, we adopt the very idea of operationalization in CLS, as described more than a decade ago by Moretti 2014: "the process whereby concepts are transformed into a series of operations—which, in their turn, allow to measure all sorts of objects. Operationalizing means building a bridge from concepts to measurement, and then to the world. In our case: from the concepts of literary theory, through some form of quantification, to literary texts."

Having said that, it is important to note that this method of operationalization might also challenge well-established CLS practices, such as annotation. While real readings are usually collected in CLS as in-line text annotation, we suggest comparing them with readings gathered as structured reflections on the literary text as a whole, as

an interpretative perspective that extends beyond mere details (Münz-Manor and Marienberg-Milikowsky 2023). After all, both methods indicate that the text is not just words on a page (or a screen), but a complex communicative act in which the recipient, not just the text itself, plays a part; they just treat this act differently.

It should be emphasized, however, that the use of a structured, research-oriented questionnaire (rather than, for example, collecting reader impressions and reviews from commercial websites or reader communities forums), restricts the respondents' interpretive horizons. Hence, the potential perception of the text in computational literary citizen science, might seem closer (but not at all identical, as Gius and Jacke have shown) to approaches that were dominant around the mid-20th century and onwards, until the rise of post-structuralism (Gius and Jacke 2022).[8] Under such conditions of a standard questionnaire, the chance of getting a provocative and fruitful overinterpretation (Culler 2007), seems quite low. Yet, our findings suggest that forced, controlled, and data-oriented reading in which interpretive freedom is – at the same time – kept and limited, and restricted to the assessment of the text after its reading, contains valuable information.

Here is where a statistical-phenomenological approach comes into play. Considering different readings as definite data (so-to-speak), and, at the same time, as potentially undecided reactions, allows quantitative-conceptual analysis to better characterize indeterminacy. Indeed, as delineated above, uncertainty can be seen as relating to the complexity of literary characterization in general. This is demonstrated by figure 1a, rating a few literary concepts, in which some are easier to decipher (linguistic register) while others are perceived as more difficult (intertextuality). This is even more evident in the relationship between these specific expressions of uncertainty, and the explicit evaluation of the questionnaire as a suitable means of assessing the novel, as documented in the last, reflexive question of the entire questionnaire (Fig. 1c).[9]

Using the extent of item skipping as a proxy for item difficulty as experienced by readers, helps shed light on uncertainty or ambivalence as being consistent among certain readers and the ways in which the questionnaire resonates their reading experience. Taking this into account, we suggest that ambivalence should be evaluated as such, rather than being normalized for the sake of adjusting the results on the one hand, or validating them on the other. Moreover, the skipping of items may suggest that readers engaged thoughtfully with the challenges posed by the questionnaire. Based on the results, we suggest that skipping may not stem from inexperience in reading literature, but rather could imply a thoughtful and reflective engagement with the text.

We have to address the difficulty in the terminology used in this paper to describe a variety of engagements of readers with the questionnaire. The term uncertainty itself is ambiguous: It may reflect an epistemic uncertainty of the reader, but also

8. We refer here only implicitly to the "digital humanities-as-structuralism" narrative which Gius and Jacke engage with in their article, because, as they demonstrate, the title "structuralism" includes many variants that are productive to literary studies but cannot be described here. Moreover, some of our more explicit sources of inspiration (Ingarden, Iser) might have roots in structuralist thinking, but are not perceived as being under this umbrella. The Hebrew Novel Project, and especially our main concern here – namely, indeterminacy, uncertainty – echoes several (sometimes seen as contradictory) thinkers and approaches.
9. We use a similar method in annotation-based projects in our lab: When the annotation aim is conceptually complicated, we add a question in which annotators have to note if, or to what extent, they are sure about their annotations. The data that such a question provides is not only useful in the process of validation and re-examination of the annotations, but also in and of itself.

an uncertainty about the aptness of the question itself or the answers provided in the questionnaire. It would be useful to consider the variety of terms that may be applicable, to different extents, to the various cases we have presented here: uncertainty, ambivalence, ambiguity, epistemic doubt/humility, rejection. They all share a degree of defiance or an outside view on the question itself, even when not refraining from partially answering the question itself. They all, thus, share a degree of unease towards the question asked. An extreme instance of a combined answer and epistemic doubt can be observed in response to the question about the novel's significance, in which 9 readers chose the answer "I do not know", while marking an additional, informative answer. Future work would have to address and create a taxonomy of the different types of uncertainty and ambiguity, in the vein of Empson's "Seven Types of Ambiguity" (Empson 1973 [1930]).[10]

6. Conclusions

We presented in this paper an analysis of uncertainty in reader evaluations of novels, within the framework of The Hebrew Novel Project. While the obvious motivation for CLCS is extensive data collection and annotation, one should not ignore the subjective nature of individual contributions. The study of reader uncertainty and its enrichment of our understanding of reader engagement with literary texts, is not something that we set out to do when starting the project, but was revealed to us serendipitously when examining the resulting corpus of questionnaires. We believe that there is a lot to be learnt from adopting a prism that focuses on the phenomenological, subjective perception of literature by readers, irrespective of the theoretical framework it is cast within. We suggest that CLCS projects may gain something by considering, at the planning stage, providing participants with a variety of means to express their uncertainty, ambivalence, and other facets of their unease with the questions. We also believe that uncertainty and ambiguity can play a much larger role than typically done when collecting data in citizen science projects, in science, social studies and humanities alike. This article provides a step in this direction.

Uncertainty and ambiguity are but one facet of the complex data collected in the Hebrew Novel Project. The same corpus lends itself to multiple analyses and perspectives. One can, for example, focus on disagreements between different readers reporting on the same novel, and return to a close reading of novels that elicit divergent reactions; one can also examine what can be learned from *resolving* disagreements and employing a distant reading approach to the consensus dataset (two directions that we are currently pursuing simultaneously). The use of diverse, and at times conflicting approaches, to the same dataset, ultimately highlights the inherent complexity of literature and its reading, reminding us that, as in the past, nothing should be taken for granted. Data can be interpreted in multiple ways, and our article suggests that ambiguity itself can be treated as an additional dataset — one that is also open to interpretation.

10. Similar to Ingarden and Iser as mentioned above, Empson is another example of a theorist who worked long before post-structuralism, and even structuralism, and yet his theory might be highly relevant for computational literary studies, and used as an inspiration.

7. Methods

The questionnaire, its theoretical premises, creation and dissemination was previously explained (Dekel and Marienberg-Milikowsky 2021). The corpus of readings used in the current paper (n=1026) was extracted on August 12, 2024 into a spreadsheet format. Number of unique readers in this corpus, n=349; Number of unique novels, n=700. Data analysis was performed on Matlab, v. R2024b. The analytic fit in Fig. 1b was calculated using the Poisson binomial distribution in the following way. First, our sampling space is $\Omega = \{0,1\}^4$, whose elements are of the structure $\bar{b} = (b_1, b_2, b_3, b_4)$, $b_i = 1$ implies skipping and $b_i = 0$ implies not skipping, and define the random variable $X(b) = \sum_i b_i$. The probability of skipping is different for each item i and denoted by p_i, and these values are estimated from the data. Then the probability of a certain outcome is:

$$P\left\{\bar{b}\right\} = \prod_{b_i=1} p_i \cdot \prod_{b_i=0} (1 - p_i)$$

and the probability of a certain number of outcomes k is given by:

$$P(X = k) = \sum_{\bar{b} \in (X=k)} P\left\{\bar{b}\right\} = \sum_{\bar{b} \in (X=k)} \prod_{b_i=1} p_i \cdot \prod_{b_i=0} (1 - p_i)$$

8. Appendix A: questionnaire items

The Hebrew Novel questionnaire includes the following scaled (1-5) questions, some are skippable (indicated below) and others which are compulsory:

1. Where along the conventional-experimental axis would you locate the novel? If you don't know, please skip the question). [from 1: the most conventional to 5: the most experimental]
2. How would you define the pace of events in the novel's plot? [from 1: very slow plot to 5: very quick plot]
3. To what extent do you think the novel's plot leaves gaps for the reader to fill in using their own knowledge, reasoning, or imagination? This refers to fundamental gaps between events, to unclear causal connections, or to essential gaps in the description of characters, landscapes, and occurrences. If you do not know, please skip the question. [from 1: very little to 5: very much]
4. Try to characterize the key events in the novel's plot. If there are multiple key events, refer only to the central ones. To what extent did they surprise you? [from 1: did not surprise at all to 5: I was really surprised]
5. To what extent, in your opinion, does the novel end in an open-ended way (where it is unclear what happens to the characters, the conflicts remain unresolved, the questions unanswered, etc.) or has a closed ending (such as a marriage, death, or 'and they lived happily ever after)? [from 1: completely open to 5: completely close]
6. If you marked 'yes' in the previous question (was Hebrew a spoken language at the time the novel was written?), how would you assess the typical linguistic

register in the novel in relation to the spoken language of the time when it was written? If you marked 'no' in the previous question, please skip this question. [from 1: very colloquial to 5: very literary]

7. To what extent do you think the novel employs intertextuality? That is, to what extent does the novel maintain a linguistic, formal, or thematic connection — direct or indirect, explicit or implicit — to other texts? If the concept is unclear, please skip this question [from 1: little usage to 5: extensive usage]

8. How readable was the novel for you? That is, did you find it easy to read, was the plot easily understood, and was the reading experience not challenging? [from 1: very readable to 5: very unreadable]

9. To what extent was it easy for you to characterize the novel using the questionnaire? [from 1: very easy to 5: very difficult]

Figure 1 provides an analysis of the skipping of items in the scaled items 1,3,6 and 7 in the above list.

Figure 2 provides an analysis of 23 items that represent the different types of questions in the questionnaire (scaled; numeric; multiple choice questions with 2-3 options; multiple choice questions with more than 3 options; multiple choice and free text). All 23 items enable the reader to express at least one type of uncertainty.

Scaled items:

- Items 1,3,6,7 in the above list.

Numeric items:

- Year of publication
- Number of pages
- Length of subdivisions
- How would you describe the network of characters in the novel? In your answer, please refer only to the main characters and to significant secondary characters, not all the characters appearing in the novel.

Multiple choice questions with 2-3 answers:

- Author's gender
- Type of narrator (diegetic, non-diegetic, alternating narrators, term unknown)
- How would you assess the reliability of the narrator? The reliability of the narrator is usually determined by the degree of alignment between the narrator's value system and knowledge framework and that of the implied author, which is perceived as the value system underlying the text.
- Is the novel structured as a nested story?:
- Does the novel distinctly mix different registers of the Hebrew language? For example, when a certain character uses a colloquial form of language while the narrator uses a literary form, or vice versa.

- To what extent Israel is central to the novel?

Multiple choice questions with more than 3 answers:

- How would you describe the division of the novel into units and sub-units?
- How can the exposition in the novel be characterized? Exposition is the part of the story that presents the background necessary for understanding the plot.
- In your opinion, what is the importance of the novel? You may mark more than one option.
- What is the main grammatical tense in which the story is narrated? The question refers to the primary tense used by the narrator.

Multiple choice and free text:

- What is the nature of the events in the plot? According to the common distinction between 'key events' that are important for advancing the plot and 'filler areas,' which include simple everyday events, descriptions of landscapes and characters, pauses, etc., try to characterize the density of key events in the plot.
- Geographically, where does the main plot or the main plots take place? You can mention more than one possibility.
- Try to define the subgenre of the novel. You may mark more than one option.
- Are languages other than Hebrew being used in the novel? If so, which are they?
- Does the novel include elements from different artistic genres? The question refers to elements that are distinctly separate from the main plot and/or form of the novel, yet are still an integral part of it.

9. Data Availability

Data and software can be found here: https://github.com/ga-jacobson/JCLS2025a/

10. Acknowledgements

This research was generously supported by grant No. 1223 from the Israeli Ministry of Science and Technology.

11. Author Contributions

Gilad Aviel Jacobson: Conceptualization, Data curation, Formal analysis, Methodology, Project administration, Resources, Software, Visualization, Writing – original draft, Writing – review & editing

Itay Marienberg-Milikowsky: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing

Yael Dekel: Conceptualization, Data curation, Methodology, Project administration, Resources, Validation, Writing – original draft, Writing – review & editing

References

Auerbach, Erich (2012). "Philology and Weltliteratur (1952)". In: *Trans. Maire Said and Edward Said. The Centennial Review* 13, 1–17.

Bonney, Rick, Tina B Phillips, Heidi L Ballard, and Jody W Enck (2016). "Can citizen science enhance public understanding of science?" In: *Public understanding of science* 25.1, 2–16.

Brossard, Dominique, Bruce Lewenstein, and Rick Bonney (2005). "Scientific knowledge and attitude change: The impact of a citizen science project". In: *International Journal of Science Education* 27.9, 1099–1121.

Culler, Jonathan D (2007). *The literary in theory*. Stanford University Press.

Dalen-Oskam, Karina van (2023). *The riddle of literary quality: a computational approach*. Amsterdam University Press.

Dekel, Yael and Itay Marienberg-Milikowsky (2021). "From distant to public reading the (Hebrew) novel in the eyes of many". In: *Magazén* 2.2, 225–252.

Dickinson, Janis L, Benjamin Zuckerberg, and David N Bonter (2010). "Citizen science as an ecological research tool: challenges and benefits". In: *Annual review of ecology, evolution, and systematics* 41.1, 149–172.

Dixon, Peter and Marisa Bortolussi (2011). "The Scientific Study of Literature: What Can, Has, and Should Be Done". In: *Scientific Study of Literature* 1.1, 59–71.

Empson, William (1973). *Seven types of ambiguity*. Chatto & Windus.

Gius, Evelyn and Janina Jacke (2017). "The hermeneutic profit of annotation: On preventing and fostering disagreement in literary analysis". In: *International Journal of Humanities and Arts Computing* 11.2, 233–254.

— (2022). "Are Computational Literary Studies Structuralist?" In: *Journal of Cultural Analytics* 7.4.

Gius, Evelyn, Marcus Willand, and Nils Reiter (2021). "On organizing a shared task for the digital humanities–conclusions and future paths". In: *Journal of Cultural Analytics* 6.4.

Haklay, Muki (2012). "Citizen science and volunteered geographic information: Overview and typology of participation". In: *Crowdsourcing geographic knowledge: Volunteered geographic information (VGI) in theory and practice*, 105–122.

Herrmann, J. Berenike, Arthur M. Jacobs, and Andrew Piper (2021). "Computational Stylistics". In: *Handbook of Empirical Literary Studies*. Ed. by Kuiken and Jacobs. Berlin and Boston: De Gruyter, 451–486.

Ingarden, Roman (1973). *The literary work of art: an investigation on the borderlines of ontology, logic, and theory of literature: with an appendix on the functions of language in the theater*. Northwestern University Press.

Iser, Wolfgang (1980). "Texts and readers". In: *Discourse Processes* 3.4, 327–343.

Koolen, Corina, Karina van Dalen-Oskam, Andreas van Cranenburgh, and Erica Nagelhout (2020). "Literary quality in the eye of the Dutch reader: The National Reader Survey". In: *Poetics* 79, 101439.

Moretti, Franco (2000). "Conjectures on world literature". In: *New left review* 2.1, 54–68.

Moretti, Franco (2014). ""Operationalizing": or, the Function of Measurement in Modern Literary Theory". In: *The Journal of English Language and Literature* 60, 3–19. 10.15794/jell.2014.60.1.001.

Münz-Manor, Ophir and Itay Marienberg-Milikowsky (2023). "Visualization of categorization: How to see the wood and the trees". In: *Digital Humanities Quarterly* 17.3.

Piper, Andrew (2020). *Can we be wrong? The Problem of Textual evidence in a Time of Data*. Cambridge University Press.

Piper, Andrew, Michael Xu, and Derek Ruths (2024). "The Social Lives of Literary Characters: Combining citizen science and language models to understand narrative social networks". In: *Proceedings of the 4th International Conference on Natural Language Processing for Digital Humanities*, 472–482.

Rebora, Simone, Peter Boot, Federico Pianzola, Brigitte Gasser, J. Berenike Herrmann, Maria Kraxenberger, Moniek M. Kuijpers, Gerhard Lauer, Piroska Lendvai, Thomas C. Messerli, and Pasqualina Sorrentino (Oct. 2021). "Digital humanities and digital social reading". In: *Digital Scholarship in the Humanities* 36.Supplement_2, ii230–ii250. 10.1093/llc/fqab020. https://doi.org/10.1093/llc/fqab020.

Salgaro, Massimo (2021). "The History of the Empirical Study of Literature from the Nineteenth to the Twenty-First Century". In: *Handbook of Empirical Literary Studies*. Ed. by Don Kuiken and Arthur M. Jacobs. Berlin and Boston: De Gruyter, 515–542.

Tauginienė, Loreta, Eglė Butkevičienė, Katrin Vohland, Barbara Heinisch, Maria Daskolia, Monika Suškevičs, Manuel Portela, Bálint Balázs, and Baiba Prūse (2020). "Citizen science in the social sciences and humanities: The power of interdisciplinarity". In: *Palgrave Communications* 6.1, 1–11.

Wecker, Alan J, Uri Schor, Dror Elovits, Daniel Stoekl Ben Ezra, Tsvi Kuflik, Moshe Lavee, Vered Raziel-Kretzmer, Avigail Ohali, and Lily Signoret (2019). "Tikkoun sofrim: A webapp for personalization and adaptation of crowdsourcing transcriptions". In: *Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization*, 109–110.

Wiggins, Andrea and Kevin Crowston (2011). "From conservation to crowdsourcing: A typology of citizen science". In: *2011 44th Hawaii international conference on system sciences*. IEEE, 1–10.