

UNIVERSITY OF TARTU
DEPARTMENT OF ENGLISH STUDIES

SENTIMENT MINING IN GOODREADS REVIEWS
OF CLASSIC AMERICAN NOVELS

BA thesis

LOTTE PARKSEPP

SUPERVISOR: Prof. Raili Marling, PhD

TARTU

2019

ABSTRACT

This bachelor's thesis analyses the sentiment used in 1 and 5 star Goodreads reviews for 10 classic American novels. The sentiment analysis is done with the SentiStrength programme and the results are reviewed and compared to review lengths and average Goodreads ratings. This is done to answer two research questions:

1. Is more sentiment expressed in 1 or 5 star reviews?
2. Which of the chosen books has the highest sentiment in its reviews?

The thesis is divided into five sections: the introductions, theoretical background, overview of the research method, analysis of the sentiment scores and conclusion. The introduction discusses why the topic was chosen. The theoretical part is divided into four sections. The first gives an outline of the democratisation of expertise in different fields, including the humanities. The second part talks about digital humanities and distant reading. The third section talks about sentiment analysis the fourth introduces the SentiStrength programme.

The theoretical part is followed by an outline of the method, the research findings and a discussion of the findings. The thesis then discusses its limitations and ideas for future research.

TABLE OF CONTENTS

ABSTRACT	2
INTRODUCTION	4
1. Democratisation in humanities	6
1.1 The democratisation of expertise	6
1.2. Digital humanities and distant reading	9
2. Emotion mining using SentiStrength	13
2.1 Sentiment analysis	13
2.2 What is SentiStrength?	15
2.3 Research method	18
2.4 Sentiment analysis in Goodreads reviews	18
2.5 Discussion	20
2.6 Limitations and future research	24
CONCLUSION	26
LIST OF REFERENCES	28
Appendix 1. Using SentiStrength	30
RESÜMEE	31

INTRODUCTION

Goodreads is the world's largest and most popular social network site for readers and book recommendations, with over two billion books and 80 million reviews (Goodreads n.d.: About Us). The site, now owned by Amazon, allows individual users to access a database of books and book reviews and also to add their own reviews and reading lists. This means that every person with an internet access can become an online literary critic. Goodreads has been increasing its popularity throughout its existence and has reached the 404th spot in the global ranking of websites (Alexa: 2018). The popularity of the site in which readers rely on the recommendations of non-professional and non-academic reviewers is an example of the broader phenomenon of the democratisation of expertise. With everyday readers logging on to their Goodreads accounts after hearing the title of a book to check the book's rating on a five star scale, it is clear that Goodreads has a great influence on what is read. This raises the question of what the amateur reviewers are saying. Thus far, there is some research on Goodreads as well as sentiment mining but previous studies have not focused on a comparison of reviews of literary classics.

The aim of this thesis is to delve into the Goodreads reviews of classic American novels and to analyse the sentiment in these reviews, to convey a deeper understanding of the emotional intensity with which negative and positive reviews are written in the online book community. This is done to answer the following research question:

1. Is more sentiment expressed in 1 or 5 star reviews?
2. Which of the chosen books has the highest sentiment in its reviews?

The books chosen for this thesis came from the Goodreads list *100 Best Books of All Time: The World Library List*, which the person creating the list references as having got from the *Bokklubben World Library* Wikipedia page (Goodreads user 2011). The list

consists of 100 books chosen by 100 writers from 54 countries and was compiled and published in 2002 by the Norwegian Book Club (Wikipedia 2018). There is a plethora of ‘best books ever’ lists on Goodreads, but I chose this one because it had a good representation of different time periods, styles and authors. I picked the 10 American novels, that were on the list as I wanted to analyse the reviews of books that are considered classics and that are often mandatory for students. Therefore, it is likely that we would see a divergence here between the experts who create literature curricula for schools and amateur reviewers of Goodreads. The analysis of the reviews, thus, wants to contribute to the broader discussion of the democratisation of expertise.

The research is done by feeding the chosen reviews through the SentiStrength programme (provided by the author of the programme, Professor Thelwall, for academic purposes) in order to generate the data on emotional intensity. The data is then compared, to find the differences between the novels as well as the positive and negative reviews. The analysis also looks at other metrics for the reviews, such as length and the rating the book has received on Goodreads.

This thesis is divided into two main chapters. The first looks at the phenomenon of democratisation and how it appears in different fields including the humanities. The first chapter also provides background info about distant reading and its history and uses for literature analysis. The second chapter focuses on sentiment mining, discussing the way in which sentiment mining is done as well as SentiStrength — the programme used for the computational part of this thesis. The second chapter also introduces the research method, the research findings, and discussion.

1. Democratisation in humanities

1.1 The democratisation of expertise

Today, everyone with an internet connection can be a historian, philologist, doctor or detective, especially with the help of online communities consisting of enthusiasts in a specific field. The phenomenon has its positive and negative aspects that are context-dependent. That is, we should not rush to make generalisations without proper research. The democratisation of different fields of life is not in itself new, but has started to receive more academic attention in the recent years.

Robinson (2016: 1565, 1566), debating the democratisation of criminal law, noted that while democratisation generally sounds like a positive occurrence, it can lead to problems when it comes to criminal law as opinions can vary drastically between ethnicities, religions and communities. Thus, it is better to leave the decisions of crime prevention, detective work and punishments to the experts. In the case of psychiatry, as Little, Lobb and Atkins show, on the issue of involuntary admission the voice of the expert is secondary and decisions are made by a Mental Health Review Board that in some countries includes a lay person, who may or may not be trained in the field as well as a lawyer and the patient's family (Little et al 2007: 93-94). However, in the case of animal testing, Khoo (2018: para 3) argues that the democratisation of the field may be a good thing. It appears that since 2002, the approval of animal testing for medical research has decreased, possibly because of the democratisation of the field and public opinion on animal testing. It therefore seems that the effects and advantages of democratisation vary in different fields of study. However, whatever the field, the democratisation of expertise seems to have a direct impact on the people outside of the field as well.

If it is possible for people to diagnose themselves on WebMD, does that mean that the internet leads to the democratisation of knowledge in general? Mößner and Kitcher argue that the epistemic practice to rely on experts and scientists is changing (2017:2). There are three general opinions that seem to follow the democratisation of knowledge. The first and more radical one is that the internet abolishes the whole notion of differential expertise and everyone has the same authority on issues, since people become increasingly autonomous due to the availability of information and simply no longer need expert opinions. This would change the whole structure on which the world has so far relied. The second way of viewing the democratisation of knowledge is that the internet does not do away with the epistemic structure but simply changes and reforms it. The third opinion is that the internet simply allows people to have more access to different sources of information and knowledge. (Mößner, Kitcher 2017: 3) The internet has created a new environment for knowledge and the people looking for it must adapt to its quirks. Mößner and Kitcher (2017:13) state that in this new environment, knowledge must be taken with a grain of salt, so to say.

Goodreads, however, is a different form of media than Wikipedia, for instance, where people searching for information would have to both trust the webpage to have high quality information and to use their own judgement in assessing the information. On Goodreads it is obvious that different users post the reviews and therefore that the opinions should not be regarded as the ultimate truth. There is also much less at stake in the case of book reviews, than let us say, medical knowledge or political decisions.

The general positive aspects of the democratisation of scholarship are collaborations with diverse groups of people with different backgrounds, mixing western and non-western approaches, using more activist and volunteer input as well as non-

academic knowledge and being able to apply all of this knowledge not only in the decision making but also in the work leading up to it (Koskinen 2017: 4673). As Koskinen (2017:4672) says, however, democracy does not guarantee objectivity and this is also evident in the context of Goodreads. Thelwall (2016: 1221) has viewed many aspects of Goodreads, one of them being the correlation between book genres and the genders of the authors. With the help of Goodreads he was able to look at 50 genres and the dominating gender of the authors as well as the readers within those genres. His research confirmed the gender bias when it comes to readers choosing a book based on the gender of the author (Thelwall 2016:1221). Maity, Panigrahi and Mukherjee (2017) have researched whether positive online reviews of books lead to larger sales numbers by comparing Goodreads ratings and Amazon best sellers list. They not only found a correlation but achieved 88.72% accuracy in predicting whether a book will *become* an Amazon best seller based on the reviews that it has on Goodreads within the period of 1 month (Maity et al 2017: 454). The reason to research and study Goodreads is, as Thelwall (2016: 1212) puts it, because there are differences between the publishing world and the world of the readers and Goodreads presents an opportunity to view the opinions of the readers all in one spot.

Goodreads is a prime example of an online community of non-experts bringing knowledge to others like them since anyone interested is able to find information on over 2.3 billion books — different editions of a novel, languages, names of main characters, settings and a short introduction to the book. There is also an abundance of voices and opinions often critiquing works that have been considered “great” for centuries. In fact, those critiques seem to have some power outside the online world, as Abdoli, Kousha and Thelwall (2017) suggest. Their analysis detected that Goodreads is a large enough platform, with enough reviews to have an impact, that can be assessed further, in a number

of fields. The mentioned article also suggests Goodreads for self-assessment for publishers since there were a number of differences between the Goodreads reviews and the author's peer reviews (Abdoli et al 2017: 2014), suggesting that in some cases the non-professionals have interesting insights when it comes to a specific subject or book and that their opinions can benefit the publishers and editors in determining what the readers are interested in.

1.2. Digital humanities and distant reading

In some ways, digital humanities and the democratisation of expertise go hand in hand. Hunter (2015) argues that the essence of digital humanities does not necessarily lie in the use of technology but in the possibility to collaborate with computer scientists in order to have digital archives of works that have previously been out of the limelight or not considered canonical. It is yet another way to bring to light a plethora of voices and opinions (Hunter 2015: 408, 409).

In 2005, Franco Moretti introduced the idea of distant reading, as an opposite to close reading, which had previously been the dominant method in the academic world for analysing literature. While close reading took a deep and detailed look at a literary work, Moretti's proposed distant reading would look at a great number of works at the same time, in order to find information that close reading would overlook. In *Graphs, Maps, Trees*, Moretti explains that close reading sees every word and sentence as unique, forcing it to have an extraordinary status. He proposes that literary historians should shift their focus from the extraordinary to a wider sense, viewing large masses of facts in order to find literature in there, too. (Moretti 2007: 3) Distant reading allows the researcher to capture either a closer or broader view of the text by changing both the object and the method of literary history (for instance, not looking at a passage from a novel but a whole genre or a

specific aspect in a text). Thus, a researcher could look at something very specific, for instance the use a single word or phrase, but on a much larger scale, consequently achieving a closer view of the text due to the specific search item and a wider view by enlarging the scale of context. Moretti (2007:1) also explains that “distance is not an obstacle but *a specific form of knowledge*” and that the specificity allows us to see connections and relations in a more clear-cut way.

Fleming states that the issue with close reading is that it relies not only on the hope that the reading itself is appropriately done, but that the passage or example of the text chosen is the *right* choice as without the well-chosen example the reading itself can turn out less than satisfying (Fleming 2017: 437). Moretti sees another issue with close reading, which is that firstly, no one could ever go through all of the literary canon to do a close reading of it all, much less works that are not in the canon. Even if it were possible, would the method be relevant? As Moretti states in *Graphs, Maps, Trees*: “/.../ A field this large cannot be understood by stitching together separate bits of knowledge about individual cases, because it *isn't* a sum of individual cases: it's a collective system, that should be grasped as such /.../” (Moretti 2007: 4).

An example of Moretti's work is his chapter “Style, Inc.: Reflections of 7000 Titles” from *Distant Reading*, where he takes a look at the titles of British novels from the years 1740 to 1850 and assembles his findings into graphs. He shows how the average length of book titles decreased rapidly from 15-20 words to 6, with long titles eventually disappearing due to more novels being published and literary reviews starting to appear. Eventually the long descriptive titles became obsolete. He is taking the base data of the titles and adding data such as the number of novels published as well as the types of titles,

new signals and compressed meanings in the titles and so forth, creating an interesting look at literature that would have been overlooked with close reading (Moretti 2013: 179-210).

It is not just Moretti who is taking advantage of the technique of distant reading. Belgum, Handley and Bott (2018) took 3000 titles of travel literature published from 1800 to 1900 and visualised that data by mapping it in order to have a wider look at the literature of the period in which travelling increased and to serve as an aid for anyone interested in a specific place and trying to find literature related to that place (Belgum et al 2018: 306, 320). Liddle analysed 20000 Victorian newspapers with the help of online newspaper databases. He analysed the storing of the digital files themselves and discussed differences in the pdf files of the papers as well as the word count in the leading article and its changes throughout the years (Liddle 2012: 230, 232, 234).

It is fair to say that the digitalising of different subfields of humanities has had an impact on the way we look at literature and has allowed for there to be more ways of conducting research, with distant reading being one of them. Computational research provides us with literary statistics that we have never had before, enabling us to have a wider view of every aspect of literature. One of those aspects that I will be looking at is emotion. Emotions have been mined in texts using distant reading techniques as well as statistics. For instance, Stanford Literary Lab, specifically, Heuser, Moretti and Steiner, took a look at the emotional intensity in describing London in novels. Their mappings show the emotional temperature and intensity of London as well as descriptions of streets and different areas and layers of London (Heuser et al 2016: 6-9). The mappings help understand the change from books being written about the upper class to the realistic depictions of everyday people, as they showcase the places that they would frequent. With

the mappings, we are also able to comprehend what London *felt* like at the time for the characters.

This thesis uses both the digitalisation of humanities and distant reading to analyse literature reviews. As Moretti says (2004:7), manually evaluating such large amounts of text is both impossible and not entirely relevant in today's world. Goodreads has given a platform for non-professionals to review books and as Abdoli, Kousha and Thelwall suggest (2017: 2014), those reviews seem to have power outside of the social media site itself. While digitalisation of humanities is not exactly a new phenomenon, it is rapidly increasing and thus provides interesting areas for research.

2. Emotion mining using SentiStrength

2.1 Sentiment analysis

Sentiment analysis is done automatically, with the help of a programme, detecting sentiment related information in texts using sentiment indicators. There are two types of sentiment analysis programmes. One uses direct sentiment indicators such as lexical indicators that have previously been categorised and marked for sentiment by a human. The other programmes use machine learning technology and indirect indicators, which are lexical indicators automatically marked for sentiment based on the context that they are in. The problem with the machine learning technology is that it marks words for sentiment based on a specific context and thus cannot be used in other contexts. To use an example provided by Thelwall, Buckley and Paltoglou (2012:165), a machine learning programme might mark words such as Iraq, for instance, as having a negative sentiment, based on comments on news stories. However, when applying the same programme to another domain, the context is different and the programme may provide incorrect analysis.

SentiStrength can be bought from the SentiStrength website or, when used for academic purposes, Professor Thelwall provides the programme for free. SentiStrength runs on Java and is available for both MacOS and Microsoft Windows. The programme consists of the .jar file, which is the programme itself, a user manual, and a file with the data. The data file has the entire lexicon used by the programme and the sentiment value for each word. I put all of these files on my desktop, as it is easiest for the code writing portion. I copied 20 positive and 20 negative reviews from each of the 10 books into my Word programme. SentiStreghth gives a separate score for every single paragraph of text or sentence, if the sentence is marked with a paragraph break, allowing the user to learn more detailed data, if they so please. I, however, modified the reviews by eliminating paragraph breaks giving me

one score per review, making the comparisons between reviews simpler and more comprehensible. I also decided to exclude quotes from the novels, as I felt that they did not directly indicate the reviewers' personal opinions and altered the results too drastically, for some of them had quoted entire paragraphs or even pages from books reviewed. I then fed the text through SentiStrength. I calculated the average positive sentiment (PS) and negative sentiment (NS) for each of the books. It is important to save the files in .txt format.

I used the Terminal programme on my MacOS operating system. An example of the Terminal commands can be seen in Appendix 1. For the example I used the 5 star reviews of *Absalom, Absalom!*, the P in the file name is for positive reviews. All of the files were on the computer desktop and the command `cd Desktop/` redirects Terminal to the computer desktop. The programme creates a separate file that will also go to desktop and is in .txt format but can be opened in other programmes including Excel.

Programmes using direct sentiment indicators, such as SentiStrength can be used despite the variations in context, as they are lexicon based and identify sentiment-bearing words and phrases, also known as *direct affective words* (Thelwall et al 2012: 165). The programme shows results for the words on a scale of 1 to 5 for positive emotion and -1 to -5 for negative emotion (1/-1=neutral and 5/-5=strong) (SentiStrength n.d: About). Some examples of strong positive sentiment from the SentiStrength sentiment lookup table would be *exquisite* (5), *overjoyed* (5) and *awesome* (4), while examples for strong negative sentiment are *catastrophe* (-4), *crying* (-4) and *heinous* (-5). For each sentence, SentiStrength gives a positive and negative sentiment score. For instance, in the sentence "I love you but hate the current political climate" SentiStrength generates the result of positive sentiment strength 3 and negative sentiment strength -4. The reason for two scores instead of one, is that it mimics the way a human brain processes mixed emotions

(SentiStrength n.d). The homepage for the programme links an article by Berrios, Totterdell and Kellett (2015) to further elaborate on this notion of mixed emotions. They discuss the varied opinions on the possibility of feeling two valenced emotions at the same time (Berrios et al 2015: para 1).

2.2 What is SentiStrength?

SentiStrength was created by Professor Thelwall from the University of Wolverhampton in order to detect emotion in short informal texts such as tweets or online comments. The programme extracts information that reflects sentiment from texts and compares it to the term weights in the programme that have been compiled and classified by humans. Similar programmes are TensiStrength, which, also created by Thelwall, evaluates the strength of stress in short informal texts, SentiMeter-Br, a Portuguese sentiment mining tool and SentiWordNet, which is built on a thesaurus-like programme WordNet.

Driscoll used SentiStrength to look at the sentiment in the feedback for Melbourne Writers Festival from both the festival's self-conducted survey and tweets relating to the festival. She found that 40% of the responses expressed positive emotion and 20% negative emotion. However, the emotional strength was fairly weak for both of the emotions. She stated that emotional analysis can be a great tool for evaluating the emotional engagement of the audience by identifying the strong sentiment and then doing a close reading to learn appropriate feedback for the event (Driscoll 2015: 871). Scrivens, Davies and Frank (2018) used SentiStrength on Twitter to analyse the changes in right-wing Twitter posts. They analysed a total of 124,058 posts and looked at the negative messages about different ethnicities and minorities. With the help of SentiStrength they

saw that messages about the LGBTQ community were the strongest in their negativity, followed by those about Jews and blacks (Scrivens et al 2018: 7, 8). While sentiment analysis was only one part of their study, it helped with their study of the negativity trajectories.

Driscoll and Sedo (2018: 250, 253) also used Goodreads to assess the reviews and analyse contemporary book culture. Their goal was specifically to study intimacy in book reading and reviewing. They used SentiStrength in addition to the content analysis of emotional vocabulary. They specifically looked at Canadian and Australian bestsellers to focus on readers' choices and behaviour. When looking at the reviews, they did not separate them by rating or any other measure, but by time of publication. This is most probably why the average negative sentiment for all of the books is -1.86 out of -5, only slightly above neutral, which is -1. They also recognised the limitations that sentiment analysis has but commended it on being fast, relatively accurate and helping them to triangulate their findings.

For my analysis, I chose to use the SentiStrength programme. The programme has previously been used to study Goodreads and I wanted to be able to compare my results to some of the other studies. Since the programme has been tested quite a bit, it has had changes made to it, making it as accurate as possible. I decided to choose a programme using direct sentiment indicators, since is not domain and context dependent if I were to expand on this study later on.

In my empirical analysis, I focused on ten American literary classics reviewed on Goodreads: William Faulkner's *Absalom, Absalom!*, Toni Morrison's *Beloved*, Ralph Ellison's *Invisible Man*, Walt Whitman's *Leaves of Grass*, Vladimir Nabokov's *Lolita*, Herman Melville's *Moby-Dick; or, the Whale*, Mark Twain's *The Adventures of*

Huckleberry Finn, Edgar Allan Poe's *The Complete Stories and Poems*, Ernest Hemingway's *The Old Man and the Sea* and William Faulkner's *The Sound and the Fury*. The Goodreads ratings for the books varied from 3.49 (for Melville's *Moby-Dick*) and 4.38 (for *The Complete Stories and Poems* by Edgar Allan Poe). 8% was the highest number of 1 star ratings with a total of 38403 people giving the book a score of 1 star out of 5 and 55% the highest number of 5 star ratings with 106863 people rating the book 5 out of 5 stars on Goodreads. *Lolita* has the most reviews and *The Complete Stories and Poems* has the least reviews (for more details, see Table 1). For each of the books on Goodreads, the website breaks down the average rating by providing rating details which are the exact number of ratings and reviews as well as the breakdown of ratings in percentages. I decided to make note of those rating details as they might provide interesting correlations between the research findings.

Table 1. The chosen books and their rating details

Title	Author	Average rating	5 star rating details	1 star rating details	Reviews
<i>Absalom, Absalom!</i>	William Faulkner	3.96	39% (14137)	3% (1361)	1865
<i>Beloved</i>	Toni Morrison	3.80	32% (90691)	4% (13207)	9269
<i>Invisible Man</i>	Ralph Ellison	3.85	32% (46574)	3% (4631)	5033
<i>Leaves of Grass</i>	Walt Whitman	4.12	45% (35624)	2% (1759)	2244
<i>Lolita</i>	Vladimir Nabokov	3.89	35% (208207)	3% (23200)	21158
<i>Moby-Dick; or, the Whale</i>	Herman Melville	3.49	26% (116686)	8% (38403)	13206
<i>The Adventures of Huckleberry Finn</i>	Mark Twain	3.81	29% (320934)	2% (29962)	13988
<i>The Complete Stories and Poems</i>	Edgar Allan Poe	4.38	55% (106863)	0% (1490)	2107
<i>The Old Man and the Sea</i>	Ernest Hemingway	3.75	30% (212327)	4% (33055)	20618
<i>The Sound and the Fury</i>	William Faulkner	3.86	36% (53359)	5% (7323)	6284

2.3 Research method

For my analysis, I picked 20 positive and 20 negative reviews for each of the 10 books (with the exception of Edgar Allan Poe's *The Complete Stories and Poems*, as that book only had 9 negative reviews on the site). I specifically wanted to see the differences between the sentiment of positive and negative reviews and thus decided to exclude 2, 3 and 4 star ratings from my work. I decided to also include the books' rating details on Goodreads such as the percentages of 1 star and 5 star ratings as well as the total number of reviews on those books. I felt they would make interesting correlations with the sentiment data. The reviews were ranked from the most to the least popular on Goodreads, based on the number of likes that each review has. Therefore, I picked 20 of the first reviews on each page, as they seem to have the support of other readers, who perhaps did not wish to write a review themselves, but agreed with the points made.

2.4 Sentiment analysis in Goodreads reviews

SentiStrength generates a positive and negative sentiment score for each of the reviews, as negative and positive sentiment is processed in parallel instead of presenting it in one emotion (SentiStrength n.d.). I then calculated an average positive and negative sentiment for each book's 5 star reviews as well as 1 star reviews. The highest positive sentiment in 5 star reviews is 4.05 for *Lolita* and the lowest 2.85 for *Invisible Man*. The rest of the scores were between 3 and 4. The highest negative sentiment in 5 star reviews is also *Lolita* with -4.3 and the lowest is *Leaves of Grass* with -2.5. The highest positive sentiment in 1 star reviews is for *Lolita* and the highest negative sentiment in 1 star reviews is for *Beloved*. The average sentiment for all of the books is higher for 5 star reviews than it is for 1 star reviews, for both positive and negative sentiment. The strongest sentiment all together is in reviews of

Lolita and the lowest in reviews of *The Complete Stories and Poems* (for more details see Table 2).

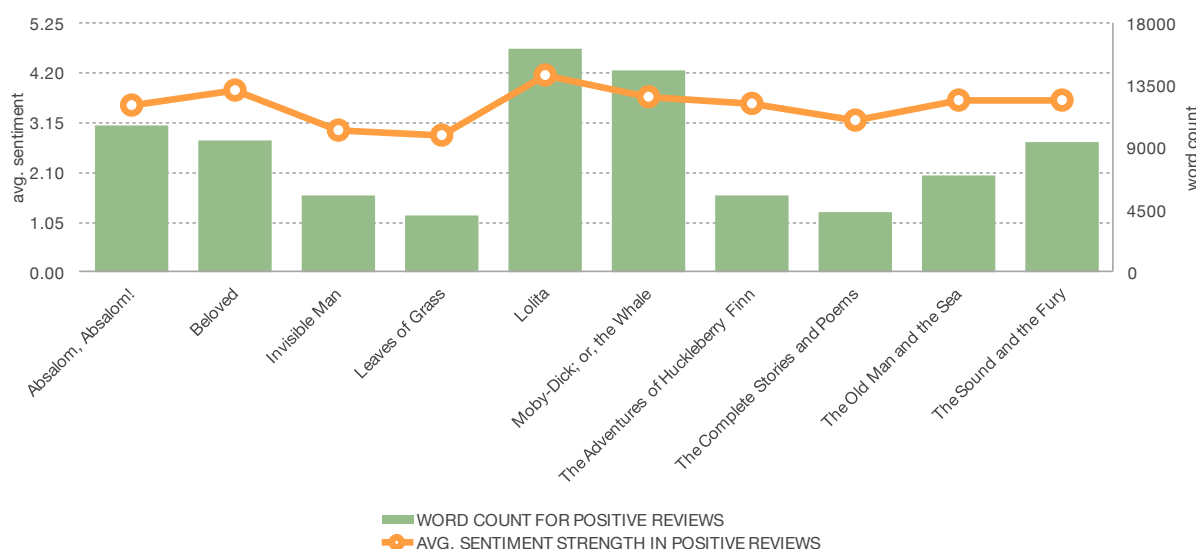
Table 2. Average sentiment for the reviews.

	5 star reviews PS	5 star reviews NS	1 star reviews PS	1 star reviews NS	avr strength
<i>Absalom, Absalom!</i>	3.35	-3.65	2.55	-2.9	3.1
<i>Beloved</i>	3.35	-4.3	2.65	-3.85	3.5
<i>Invisible Man</i>	2.85	-3.1	2	-2.85	2.7
<i>Leaves of Grass</i>	3.25	-2.5	2.35	-2.75	2.7
<i>Lolita</i>	4.05	-4.25	2.9	-3.75	3.7
<i>Moby-Dick; or, the Whale</i>	3.7	-3.7	2.45	-3.05	3.2
<i>The Adventures of Huckleberry Finn</i>	3.45	-3.6	2.55	-3.55	3.3
<i>The Complete Stories and Poems</i>	3.2	-3.2	1.89	-2.1	2.6
<i>The Old Man and the Sea</i>	3.55	-3.7	2.45	-3.45	3.3
<i>The Sound and the Fury</i>	3.45	-3.8	2.5	-3.15	3.2
average all together	3.42	-3.57	2.43	-3.14	3.1

One of my hypotheses at the beginning of the research project was that the sentiment would be stronger in the negative reviews than in the positive reviews. This hypothesis was not borne out by the findings. The sentiment (both positive and negative) is stronger in positive reviews, averaging a 3.5 out of 5 and the strength of sentiment in negative reviews is 2.8 out of 5. So it seems that reviewers, who write 5 star reviews, do so with more emotion than those who write 1 star reviews. While many of the negative reviews use pungent language, their reviews are not long enough to have that emotion registered as very strong. There appears to be a direct correlation between the strength of emotion and the length of the reviews, as evidenced in Figure 1. Thus, while one would assume that a short punchy review

emanates a stronger feeling of emotion, this is not demonstrated in the sentiment analysis here.

Figure 1. Comparisons between sentiment in positive reviews and word count in positive reviews



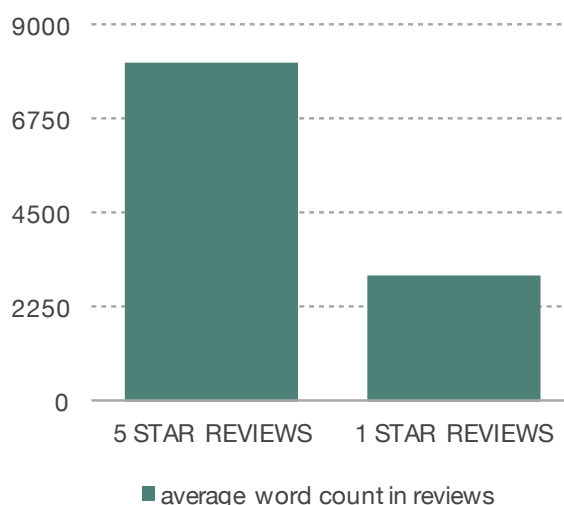
Another important finding is that for nine books out of the ten studied, negative sentiment in positive reviews is equal to or stronger than positive sentiment. In the research conducted by Driscoll and Sedo (2018: 253), it appears that only one of the books analysed by them also had a stronger average negative sentiment than average positive sentiment. However, their research did not separate reviews based on the reviewers' rating of the books and thus cannot be compared to my findings, as mine are focused on the star rating system.

2.5 Discussion

One possible explanation for the seemingly warped sentiment findings is the fact that positive reviews, more than negative reviews, describe the characters, subject matter and main problems presented in the novels in the reviews or at least give a general overview of the plot.

This is somewhat evidenced in the fact that positive reviews are on average 63% longer than negative reviews, as shown in Figure 2. Positive reviews seem to go over the plot of the book, the author's writing style as well as the reviewer's personal experiences when reading. Thus, it can be argued that the sentiment detection is somewhat too swayed by the fact that all of these novels portray a serious subject matter, intense characters and storylines and subsequently create strong emotions in the readers, which make their way into the reviews. Driscoll and Sedo (2018: 248) also suggested that positive reviews go deeper into the reading experience and the readers describe their emotional states as they were reading, showing that 86.1% of the reviews they analysed describe the reading experience and 68% mentioned an emotional response to the text. Such reviews can be intense in their descriptions and SentiStrength marks that as intense negative sentiment. For instance, a quote from a positive review on *Beloved* reads: "I felt the hopelessness of Sethe and Denver who had no place else to go" reads as a neutral positive sentiment but is assigned a -4 negative sentiment. While generally having a reader relate to the characters and be emotionally invested in their journey and hardships is seen as a mark of good writing, in this case the programme is applying negative sentiment of -4 to the word *hopelessness*. Another interesting example comes from the five star review of *The Complete Stories and Poems* by Edgar Allan Poe and serves as an example of the machine not being totally up to date on new slang. Goodreads user writes: "Wig: snatched; Crops: cleared; Me: shook; Hotel: trivago". SentiStrength marked the word snatched as having negative sentiment of -2 while the top definition of *snatching wigs* on Urban Dictionary describes it as "deadass the highest form of reaction". SentiStrength also marked *shook* for negative sentiment of -2. While the word can indeed be used to express a negative reaction, based on the context, the reaction here seems to be overwhelmingly positive.

Figure 2. Comparison of word count in reviews.



When it comes to the specific books on the list, the emotion expressed about *Lolita* stands out. *Lolita* has the highest positive sentiment rating in the 5 star reviews, meaning the reviewers had the most intense positive affectivity towards that book. However, as noted, in the 5 star reviews the negative sentiment for *Lolita* was also the highest. *Lolita* also received the second highest negative sentiment in 1 star reviews. -3.7 is quite a high negativity score, meaning that *Lolita* is a book that creates the most contested opinions. Quite intense positive and negative emotions are tied to this book, which are perhaps best explained by the reviews themselves. One Goodreads user, having given *Lolita* a 5 star rating, writes: “There is an almighty conflict between morality and aesthetics happening between the pages”. Another 5 star review reads: “Sick, twisted and beautiful. Love this.” and another one says: “This book scared the living daylights out of me.” All of these reviewers gave the book a 5 star rating, yet they understand that the subject portrayed by the book is controversial and they themselves feel conflicted when reading the book and later when reviewing it. What I find to be a positive sign of democratisation of literature reviews is the fact that the reviewers express their conflicting feelings freely and are able to discuss all of the topics without fear of editors deeming some of the conversations unsuitable.

It seems that the people reviewing *Lolita* with a 1 star rating are not quite as conflicted. The reviews are much more to the point. For instance, one of the reviews reads: “Can’t do it. Vile. Offensive. Obscene. DNF”. They also seem to be addressing the book’s status as a classic as well as an awareness of the fact that the author is using beautiful language to make the reader confused about their feelings towards the subject matter. The difference is that the 5 star reviewers discuss their conflicts, almost as if they were defending their rating choice. Perhaps that is also the reason why negative reviews are shorter. When it comes to confrontational subjects in books, they do not feel the need to explain or defend their choices. It is also worth mentioning that while *Lolita* has the highest positive sentiment and a high negative sentiment score, it is neither the most or least popular book on the list. On Goodreads it has the average rating of 3.89, while *The Complete Stories and Poems* has 4.38 and the lowest rated book is *Moby-Dick* with a rating of 3.49. However, *Lolita* does have the most reviews written on it. As for *Moby-Dick*, it is the lowest rated book on the list and it has the second longest reviews of the books, after *Lolita*, indicating that in this instance, negative emotions do lend themselves to verbose reviews.

The book which elicits the least emotions in the reviews is more difficult to determine. The lowest score of the overall sentiment, in both 5 and 1 star reviews goes to *The Complete Stories and Poems* as it has the lowest sentiment in its negative reviews. It can be said that since Goodreads only had nine 1 star reviews for Poe’s book, it is more difficult to compare that score to the other books. However, since it appears that 1490 users have rated that book a 1 out of 5, but only nine on them decided to explain their thoughts in a review, it can be said that because of the low participation, the book automatically has the least negative sentiment in the 1 star reviews. From the books on my list, it also has the highest 5 star rating

on Goodreads, with 55%. Yet, it did not reach the highest positive sentiment in positive reviews.

It seems that the reviewers write 5 star reviews with the least emotion for *Invisible Man*. The overall sentiment strength score for *Invisible Man* is equal to that of *Leaves of Grass*. It appears that two out of the three books with the lowest sentiment for positive and negative reviews include poetry. While Poe's work also includes his short stories, both *Leaves of Grass* and *The Complete Stories and Poems* are collections, not novels, like the rest. It could indicate that people use less emotion language when reviewing poetry or short story collections and that research could be an interesting future addition to my work.

When looking at the average positive sentiment and average negative sentiment, it is clear that my decision to look at the 1 star and 5 star reviews plays a part in the strength of negative sentiment. Driscoll and Sedo (2018: 253) found that when looking at a random sample of all reviews they had a positive sentiment of 2.65 and a negative sentiment of -1.86. With my chosen sample of the reviews, positive sentiment for all of the books amounted to 2.92 and negative sentiment to -3.36. While the positive sentiment seems to be fairly similar for both my and their research, there is a strong difference in negative sentiment. Thus, it seems that it is possible to get a feel of the positive sentiment for all reviews, even when disregarding the 2, 3, and 4 star reviews. However, negative sentiment seems to be a bit more complicated and including those in-between ratings has an impact on the overall sentiment.

2.6 Limitations and future research

While SentiStrength was my chosen tool for sentiment analysis, there are limitations to the programme and analysing sentiment in this way. As previously stated, SentiStrength is not up to date on the most current slang and marks discussions of the problematic topics as

negative sentiment. However, SentiStregth is a great tool to use in conjunction with manual sentiment analysis. In future research, if the time and format permits, it might be interesting to use one or a few of those other sentiment detection programmes, to compare results and get a more well-rounded analysis.

Something that could be interesting for future research is that 76% of Goodreads' users are female (Thelwall 2016: 1218) and when studying sentiment on Goodreads, taking the reviewers gender and age into account might offer somewhat varying results. Other aspects that could vary the results are the ratings the books have on Goodreads, the specific books, and whether the review has received the approval (likes) of other users.

CONCLUSION

The democratisation of expertise can be seen in all areas of life with both positive and negative impacts and literature reviews are not immune to the changing world. With 85 million users (Goodreads n.d.: About Us), Goodreads is the most popular social media site for book-lovers and thus offers an interesting platform for literary research. One of the possibilities is sentiment mining, which has been used to look at Goodreads reviews before. The aim of this thesis, however, was to separate 5 star and 1 star reviews in order to analyse the sentiment of people writing negative reviews versus positive reviews. This type of research uses the distant reading approach, which was first introduced by Franco Moretti as an alternative to close reading. Distant reading often uses the tools of digital humanities in order to look at a great number of texts at the same time. In the case of this research, I was able to determine the sentiment of 400 reviews mechanically.

I began this thesis to answer two research questions based on the analysis of the sentiment findings. The first was whether more sentiment was expressed in positive reviews or negative reviews. The second, which of the 10 American novels chosen had the highest sentiment in its reviews. After going through the reviews and removing book quotes and paragraph breaks, I was able to feed the reviews through SentiStrength, my chosen sentiment detection programme. SentiStrength uses lexical indicators for its sentiment calculations instead of a context-dependent machine learning programme. Once the programme had calculated its results, I had one positive sentiment and one negative sentiment score for each of the reviews. SentiStrength uses a scoring method, where each sentence gets a positive sentiment score of 1 to 5 and a negative sentiment score of 1 to 5 (1/-1=neutral and 5/-5=strong). The programme uses two scores as that mimics the way a human brain would analyse sentiment — positive and negative emotions are parallel.

In order to answer my first research question, I calculated the mean positive and negative sentiment score for each of the books' 1 and 5 star reviews. Surprisingly, both positive and negative sentiment was stronger in 5 star reviews. To figure out what might cause these seemingly warped sentiment scores, I looked at the length of the reviews. As it turned out, 5 star reviews are on average 63% longer than 1 star reviews and in these longer reviews, the people writing positive reviews, not only give their opinion on the book, but discuss the plot, writing style and their own emotional reactions to the book. Since all of the books on this list have serious themes, with the longer reviews, they affected the sentiment score.

As for the second question, I looked at the sentiment scores of the 10 individual books. The overall highest sentiment in both positive and negative reviews was for *Lolita*. The lowest overall sentiment however, was more difficult to see. The lowest overall sentiment score was for *The Complete Stories and Poems*. However, *The Complete Stories and Poems* did not have the lowest sentiment in 5 star reviews. That spot went to *Invisible Man*. The overall sentiment for *Invisible Man* was equal to *Leaves of Grass*, meaning that of the three lowest rated books, two were collections and included poetry.

There are aspects of this research that could be viewed in future works, such as the difference including 2, 3 and 4 star ratings, viewing the reviewers gender, age and the specific books, dividing genres, fiction and non-fiction.

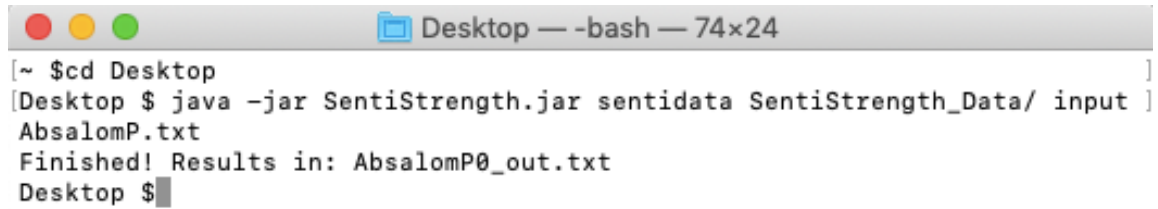
LIST OF REFERENCES

- Abdoli, Mahshid and Kousha, Kayvan and Thelwall, Mike. 2017. Goodreads Reviews to Assess the Wider Impact of Books. *Journal of the Association for Information Science and Technology*. 68:8, 2004–2016.
- Alexa. 2018. [goodreads.com](https://www.alexa.com/siteinfo/goodreads.com) Traffic Statistics. Available at: <https://www.alexa.com/siteinfo/goodreads.com>, accessed November 20, 2018.
- Belgum, Kirsten and Handley, Keith and Bott, Rachel. 2018. Mapping Travel Writing: A Digital Humanities Project to Visualise Change in Nineteenth-century Published Travel Texts. *Studies in Travel Writing*. 22:3, 306–324.
- Berrios, Raul and Totterdell, Peter and Kellett Stephen. 2015. Eliciting Mixed Emotions: A Meta-Analysis Comparing Models, Types, and Measures. *Frontiers in Psychology*. Available at: https://www.frontiersin.org/articles/10.3389/fpsyg.2015.00428/full?utm_source=newsletter&utm_medium=email&utm_campaign=Psychology-w18-2015#h8, accessed May 20, 2019
- Driscoll, Beth. 2015. Sentiment Analysis and the Literary Festival Audience. *Journal of Media & Cultural Studies*. 29:6, 861–873.
- Fleming, Paul. 2017. Tragedy, for Example: Distant Reading and Exemplary Reading (Moretti). *New Literary History*. 48:3, 437–455.
- Goodreads. n.d. About Us. Available at: <https://www.goodreads.com/about/us?rel=nofollow>, accessed November 22, 2018.
- Goodreads user. 2011. 100 Best Books of All Time: The World Library List. Available at: https://www.goodreads.com/list/show/9440.100_Best_Books_of_All_Time_The_World_Library_List?page=1, accessed November 22, 2018.
- Hallyuhoo. 2018. Snatching Wigs. Urban Dictionary. Available at: <https://www.urbandictionary.com/define.php?term=snatching%20wigs>, accessed April 25, 2019.
- Heuser, Ryan and Moretti, Franco and Steiner, Erik. The Emotions of London. Pamphlets of the Sanford Literary Lab. Pamphlet 13. Available at: <https://litlab.stanford.edu/LiteraryLabPamphlet13.pdf>, accessed April 29, 2019.
- Hunter, Andrea. 2015. The Digital Humanities and Democracy. *Canadian Journal of Communication*. Vol 40:3, 407–423.
- Khoo, Shaun Yon-Seng. 2018. Justifiability and Animal Research in Health: Can Democratisation Help Resolve Difficulties? *Animals* 8:2, 1–12.
- Koskinen, Inkeri. 2017. Where is the Epistemic Community? On Democratisation of Science and Social Accounts of Objectivity. *Synthese*. 194:12, 4671–4686.
- Liddle, Dallas. 2012. ‘Distant Reading’ *The Times* using the *Times Digital Archive*. *Journal of Victorian Culture*. 17:2, 230–237.
- Little, John and Lobb, David and Atkins, Martin. 2007. The democratisation of psychiatry. *Australasian Psychiatry*, 15:2, 93–96.
- Maity, Suman Kalyan and Panigrahi, Abhishek and Mukherjee, Animesh. 2017. Book Reading Behavior on Goodreads Can Predict the Amazon Best Sellers.

- Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. 451–454.
- Moretti, Franco. 2013. *Distant Reading*. London and New York: Verso.
- Moretti, Franco. 2007. *Graphs, Maps, Trees*. London and New York: Verso.
- Moßner, Nicola and Kitcher, Philip. 2017. Knowledge, Democracy, and the Internet. *Minerva*. 55:1, 1–24.
- Princeton University. 2010. About WordNet. Available at: <https://wordnet.princeton.edu>, accessed April 29, 2019
- Robinson, Paul H. 2016. Democratizing Criminal Law: Feasibility, Utility, and the Challenge of Social Change. *Northwestern University Law Review* 111:6. 1565–1596.
- Scrivens, Ryan and Davies, Garth and Frank, Richard. 2018. Measuring the Evolution of Radical Right-Wing Posting Behaviors Online. *Deviant Behavior*.
- SentiStrength. n.d. About SentiStrength. Available at: <http://sentistrength.wlv.ac.uk/#About>, accessed April 29, 2019.
- TensiStrength. n.d. About TensiStrength. Available at: <http://sentistrength.wlv.ac.uk/TensiStrength.html#About>, accessed April 29, 2019.
- Thelwall, Mike. 2016. Book Genre and Author Gender: Romance>Paranormal-Romance to Autobiography>Memoir. *Journal of the Association for Information Science & Technology*. Vol 68:5. 1212–1223.
- Thelwall, Mike and Kevan Buckley and Georgios Paltoglou. 2012. Sentiment strength detection for the social web. *American Society for Information Science and Technology*. 63:1, 163–173.
- Wikipedia. 2018. Bokklubben World Library. Wikimedia Foundation, Inc. Available at: https://en.wikipedia.org/wiki/Bokklubben_World_Library, accessed November 22, 2018.

Appendix 1. Using SentiStrength

Image 1. Terminal commands to run SentiStrength

A screenshot of a macOS terminal window. The title bar shows three colored window control buttons (red, yellow, green) on the left, followed by a folder icon and the text "Desktop — -bash — 74x24". The terminal content shows a sequence of commands and their output: first, a prompt "[~ \$]" followed by the command "cd Desktop" and a closing bracket; then, a prompt "[Desktop \$]" followed by the command "java -jar SentiStrength.jar sentidata SentiStrength_Data/ input" and a closing bracket; then, the output "AbsalomP.txt"; then, the output "Finished! Results in: AbsalomP0_out.txt"; and finally, a prompt "Desktop \$" with a cursor. The text is in a monospaced font.

```
[~ $cd Desktop ]  
[Desktop $ java -jar SentiStrength.jar sentidata SentiStrength_Data/ input ]  
AbsalomP.txt  
Finished! Results in: AbsalomP0_out.txt  
Desktop $
```

RESÜMEE

TARTU ÜLIKOOL
ANGLISTIKA OSAKOND

Lotte Parksepp

Sentiment Mining in Goodreads Reviews of Classic American Novels

Sentimendikaeve Goodreadsi Ameerika kirjandusklassika arvustustes

bakalaureusetöö

2019

Lehekülgede arv: 31

Annotatsioon:

Käesolev bakalaureusetöö uurib sentimedi olemasolu ja tugevust veebilehe Goodreads kümne Ameerika kirjandusklassikasse kuuluva teose arvustustes. Töö eesmärgiks oli välja selgitada, kas tugevama emotsiooniga kirjutatakse positiivseid või negatiivseid arvustusi ning milline kümnest kirjandusklassikust tekitab lugejate arvustustes tugevaimaid emotsioone.

Töö teoreetiline osa annab ülevaate ekspertiisi demokratiseerumisest erinevates valdkondades, digihumanitaariast ning sentimendianalüüsi teostamisest. Lisaks sellele ka ülevaade töös kasutatud programmi SentiStrength taustast ja kasutamisest.

Töö empiiriline osa analüüsib 400 arvustuse sentimente, kasutades programmi SentiStrength. Analüüsis leiti, et positiivsetes arvustustes on rohkem emotsiooni kui negatiivsetes ja selle põhjuseks võib olla positiivsete arvustuste märgatav detailsus. Analüüsis leiti ka, et kümnest valitud raamatust kõrgeima sentimendi hinnanguga on Vladimir Nabokovi “Lolita”.

Märksõnad

sentiment, emotsioon, Ameerika kirjandus, digihumanitaaria, Goodreads

Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina, Lotte Parksepp,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose

Sentiment Mining in Goodreads Reviews of Classic American Novels,

mille juhendaja on Raili Marling,

1.1.reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;

1.2.üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace'i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.

2. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.

3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Lotte Parksepp

Tartus, 24.05.2019

Autorsuse kinnitus

Kinnitan, et olen koostanud käesoleva bakalaureusetöö ise ning toonud korrektselt välja teiste autorite panuse. Töö on koostatud lähtudes Tartu Ülikooli maailma keelte ja kultuuride kolledži anglistika osakonna bakalaureusetöö nõuetest ning on kooskõlas heade akadeemiliste tavadega.

Lotte Parksepp
Tartus, 24.05.2019