19th-century sentiment trends on rural vs urban life in Danish literature

Abstract

This study aims to better understand the people of the 19th century and how they experienced their contemporary lives. The 19th century is described as a decade of progress, but European fiction tells a tale of negative sentiment towards city life, according to cultural historian Hannu Salmi. In this study, we test this by looking at 19th century literary views on both the city and the countryside. The study uses digital sentimental analysis on Danish 19-hundreds short stories and poems to try and get a more transparent and neutral analysision, by using specific keywords affiliated with either the rural or urban setting to locate places in the literature where either the city or countryside have been mentioned. Then it uses N - words from keywords with a 5 word range and analyses these words with the Danish sentimental analysis tool sentida to understand if the keyword we found was described either positively, negatively or neutral. Our hypothesis was that the countryside, in the century of unseen urbanisation and innovation, would idolise the countryside, and would describe the rural area more positively, and the urban more negative. The results of the study shows the urban setting as being described with both more positive and negative words. Where the rural setting is being in more neutral tones. Therefore, we were not able to confirm our hypothesis. The study was limited by the lexicon based sentiment analysis combined and suffered from a lack of representatives through a small sample size, which makes our results somewhat unreliable.

Introduction

The 19th century is often described as a century of incredible progress because of the industrialisation and urbanisation movements. These changes are often described positively by political historians. Seldom do these political historians focus on how the contemporary people saw the 19th-century themselves. Studies that have had this focus, have not used digital methods, have therefore mostly been limited to the historiens own understanding/analysis of the literature they selected. Few of them have had any focus on Denmark, and almost none of them exclusively. We want to expand this area of study because different countries entered the Industrial Revolution at different times, which may have resulted in different experiences of the 19th-century.

This paper aims to explore how people in the 19th century viewed the countryside in comparison to the city, and whether these perceptions were favorable or unfavorable. We will do this by looking at how contemporary fiction authors portrayed these environments. To answer this we will use digital methods to get a more neutral and transparent analysis of Danish contemporary literature. Our inspiration for this question comes from the work of Hannu Salmi in his book *Nineteenth-Century Europe – A Cultural History*. In this book, he describes how people inside the city began to idolize the countryside, and how people with the capital to do so often took vacations there[1]. Salmi makes this argument by interpreting passages of popular European literature. We are interested in the objectivity of his analysis, and we wish to test whether we would reach the same conclusions using digital methods. We will use danish literature to further expand his analysis, since industrialization happened later and somewhat differently, for example with more focus on industrialization in the agricultural sector[2].

Background

The 19th century was defined by an unseen urbanization. Cities were not new, but industrialization created a rapid growth in both their size and their demand for workforce, while at the same time reducing demand in the countryside. In Denmark, 20% of the population lived in cities in 1814, a number that had risen to 49% by 1911.

The Industrial Revolution and the rise of capitalism led to a growth in tertiary professions, which in turn led to the production of new products and services. Cities soon became home to a new middle class who sought to define themselves through these new products, resulting in a new culture and lifestyle that differed significantly from that of the countryside[3]. These new lifestyles and cultures are often discussed from a modern viewpoint, but rarely from the contemporary people's point of view. Did people see the city as a shining place of invention, or a noisy place of overwhelming speed and impressions? And did these inventions change how people viewed the countryside in comparison? As Hannu Salmi states in the book *Nineteenth-Century Europe – A Cultural History*: "In historical literature and imagination, the nineteenth

---

[1] Hannu Salmi, *Nineteenth-Century Europe* (Polity Press, 2008), 79.

[2] Claus Møller Jørgensen, "Fra enevældig helstat til nationalstat, 1814-1914", i *En Danmarks Historie fra vikingetid til nutid*, af Thorsten Borring Olesen og Paulsen Bjørn, 2. udg. (Aarhus universitetsforlag, u.å.), 276–79.

[3] Jørgensen, "Fra enevældig helstat til nationalstat, 1814-1914", 288–89.

century is indeed portrayed more often as black and white or grey"[4] - it is very doubtful that people saw it the same way back then. To study how people in general felt in the 19th-century can be a difficult task, as this is not typically recorded in history books. Literature from contemporary authors does offer a look into the words and emotions that were used to describe the 19th century. From a purely historical standpoint, this is interesting because it allows us to understand what emotions people felt in a century of such innovation. The result of this study should enable us to get a better understanding of how people back then felt about a whole new world of invention—and thereby enable us to see if there are any modern comparisons. It can, however, be questioned to what degree fictional literature can reflect contemporary culture. It is also important to consider which literary genres are used in digital analysis. The 19th century, after all, contained a number of literary periods such as *Romantikken*, *Romantismen*, and *Det moderne gennembrud*, which each had different focuses and goals—and may have ignored certain parts of contemporary culture[5].

 We will, however, try our best to solve this issue by using a fair mix of these different genres, the selection of which will be explained later in the paper.

Related work

The work of determining culture with the help of literature, is not a new or unique task, our study is in this regard inspired by Hannu Salmi who, in his work *Nineteenth-Century Europe – A Cultural History*, shows that literary analysis can lead to a new understanding of culture history, his work is however, made only using close reading, and not digital sentimental analysis. Digital Sentimental analysis as a way to understand culture is a discipline that has moved a lot throughout recent years. Thomas Schmidt and Manuel Burghardts paper "An evaluation of Lexicon-based Sentiment Analysis Techniques for the plays of Gotthold Ephraim Lessing" is a shining example of this. The focus of the paper is on the limitations and abilities of computer based sentiment analysis, which is examined through German plays in the 1600-1800 hundreds. Which has been especially useful for us to learn about how it was smart to use the sentimental analysis on our text[6]. Another example is Kristine Nielsen Degn et al's work "Sentiment Classification of historical Danish and Norwegian Literary Text" where they use sentimental analysis to see the different expressions from female and male authors in Danish and Norwegian culture. Their paper, just like Thomas Schmidt and Manuel Burghardts, also shows the limits and workaround that are required to make the sentimental analysis work. The journal "Samlaren" from Sigrid Schottenius Cullhed and Jon Viklund, which names different options of sentimental analysis, and what weaknesses the different approaches have.

The limits of lexicon based sentiment analysis

Digital Sentimental analysis has some limits and difficulties, especially when they are lexicon based. The difficulty of understanding more complex emotions such as irony, sadness, anger and so on. like Thomas

---

[4] Salmi, *Nineteenth-Century Europe*, 1.
[5] Ukendt, "Litteraturhistoriske oversigter", Dansksiderne.dk, u.å., https://dansksiderne.dk/index.php?id=3440.
[6] Thomas Schmidt og Manuel Burgardt, "An Evaluation of Lexicon-based Sentiment Analysis Techniques for the Plays of Gotthold Ephraim Lessing", Sintemntal Analysis, u.å.

Schmidt and Manuel Burghardt say in "An Evaluation of Lexicon-based Sentiment Analysis Techniques for the Plays of Gotthold Ephraim Lessing "common problems of lexicon-based methods like the handling of irony and negations are certainly additional reasons for the mediocre accuracies. Based on our results, we consider the usage of general purpose lexicons alone as not sufficient to achieve acceptable accuracy scores"[7] Which is also why our focus only lies in if the countryside is portrayed good or bad, and aren't sorting things into more complex emotions. We however are still vulnerable to irony, and figurative language, especially because we are using text as data[8], which means our analysis isn't able to see the difference between how words are in relation to each other.

Sentiment analysis, if based on a lexicon, is also limited to the words that lexicon holds. Which can be especially limiting if data is used, from a different time period than the lexicon is created to. You can of course teach an AI to understand the language and thereby make it more robust when it comes to new words, this is however resource intensive, and is outside the scope of this project.

The lack of an understanding for a story's moral is another limit of sentimental analysis in general. It is hard to make Digital sentiment analysis understand when a story is trying to put a moral or lesson on itself. This may result in countryside positive literature being flagged as countryside negative, because only the end of the story is using countryside positive words, which the digital analysis can't understand actually changes the meaning of the story.

Hypothesis

Through a sentiment analysis of rural and urban keywords in Danish 19th-century fictional literature, we expect to find a trend of more positive sentiment towards the rural keywords in comparison to the urban keywords. We expect this trend to increase throughout the 19th century, correlating with the literary periods.

Research Design

This paper uses the context window of n words around keywords, with predefined urban and rural keywords, respectively. This allows us to use the Danish sentiment analysis tool Sentida to transform these qualitative values into either negative or positive numbers, thereby reforming the qualitative values of words into a quantitative value, which we can then use to describe the general sentiment of urban and rural life in the literature.

We have chosen to test our hypothesis on Danish literature from the 19th century, spanning different literary periods. We hope literature can provide insight into how people in the 19th century viewed and perhaps even idealized the countryside or the city. We will compare and contrast works spanning different authors, genres and themes to get a broad view. In this way, we can attempt to see patterns and potentially groupings of different views based on these factors.

---

[7] Schmidt og Burgardt, "An Evaluation of Lexicon-based Sentiment Analysis Techniques for the Plays of Gotthold Ephraim Lessing".
[8] Nina Tahmasebi og Simon Hengchen, "Samlaren" 140 (2019): 200–201.

N-words from keyword and Sentida

Our project is using the context window of N-words from keywords to find a negative or positive sentimental value, which in our case means that we look for words with positive or negative sentimental value 5 words in front and behind every keyword that our analysis finds[9]. To get our keywords we have asked chatGPT to make us a list of words associated with the city, and the countryside in the 19th-century. We have edited the keywords to eliminate words that have multiple meanings, so we can be sure that our keywords are actually related to the respective categories. Furthermore, we removed words that simply were not categorically urban or rural. This list will be stated in this studiests repository, a link to which you can find at the end of the paper. To determine the negative or positive value of the words around our keyword, we are using the danish lexicon based sentiment analysis tool Sentida. A limit of Sentida is that it is not designed for analysis of modern texts, and as such does not have many 19th century variations and spellings. Since our texts are long, we made the determination that there are still plenty of passages that have words included in the Sentida lexicon.

Software Framework and Material

Hardware: Lenovo Yoga laptop

Processor   AMD Ryzen 7 7735HS with Radeon Graphics   3.20 GHz

Installed RAM   16,0 GB (13,7 GB usable)

System type   64-bit operating system, x64-based processor

Software: Project coded in R version 4,4,2 on Windows 11. We used the tidyverse (Hadley et al 2019), pdftools (Ooms, 2023), tokenizers (Mullen et al., 2018), sentida (Esben Kran et al., 2019), purrr (Henry & Wickham, 2023), dplyr (Wickham, François, Henry, & Müller, 2014), tibble (Müller & Wickham, 2023), ggplot2 (Wickham, 2007), and here (Müller, 2020) packages.

Data Acquisition and Processing:

We have used a combination of novels and poems from 19th-century Denmark as sources for our analysis, which can be found under "Analyzed Sources." Our goal was to get the most representative view of literature from the 19th century. To do this, we focused mainly on different literary periods in the century. We selected the literary periods *Romantikken*, *Romantismen*, and *Det moderne gennembrud*, which were the main dominant periods in Denmark at the time. To ensure equal representation, we tried to use an equal number of texts from each period. We mainly used short stories as our sources because we believe they have a greater chance of casually expressing an opinion on either the city or the countryside. Romantikken is, however, a special case because this period is especially defined by poetry. To compensate for this, we included more texts from this period due to their shorter length. The decision on how to select the most representative

---

[9] Tahmasebi og Hengchen, "Samlaren".

literature was not easy. Other options—such as focusing on when in the 19th century the literature was written—were considered. However, we chose to focus on literary periods instead, as we believe literature changes more significantly based on literary movements than on the exact year of publication. Additionally, the literary periods are already spread throughout the 19th century, which we hope will still make our results representative for the century as a whole.

Analyzed sources

We have used the following books/poems in our analysis:

1) *Kun en spillemand (1837)* – https://tekster.kb.dk/text/adl-texts-andersen05val-root

2) *Et dukkehjem (1879)* – https://www.ibsen.uio.no/DRVIT_Du%7CDuht.pdf

3) *Lykke-Per* – (1898-1904)
https://www.ronlev.dk/danske-klassikere/lykke-per-af-henrik-pontoppidan.html

4) A collection of Danish romantic poems and works 1803-1840:
   - Guldhornene - af Adam Oehlenschläger (1803)
   - Indvielsen - af Schack Staffeldt (1804)
   - Simon Peder - af Adam Oehlenschläger(1805)
   - Aften-sang - af Adam Oehlensläger (1805)
   - Morgen Vandring - af Adam Oehlenschläger (1805)
   - Platonisme - af Schack von Staffelt (1808)
   - Hans og Grethe - af Christian Winther (1828)
   - Moderen og Barnet - af H.C. Andersen (1829)
   - Angst - af Emil Aarestrup (1838)
   - Paa Sneen - af Emil Aarestrup (1838)
   - Drømmeverdnen - af Emil Aarestrup (1838)
   - Min Skat - af Christian Winther (1840)
   - Hvad er en Digter? - af Søren Kierkegaard (1843)
   - Klokken - af H.C. Andersen (1850)

https://dansksiderne.dk/index.php?id=3441#c11398

5) *Begrebet Angst (1844)* – https://www.kb.dk/e-mat/dod/111408014975-bw.pdf

6) fra hytterne (1887) - https://www.kb.dk/e-mat/dod/130020755076.pdf

7) Det forjættede land(1892) - https://books.google.com/books/download/Det_forj%C3%
A6ttede_land.pdf?id=SRwwAAAAMAAJ&output=pdf

Findings

Average sentiment for rural vs urban themes
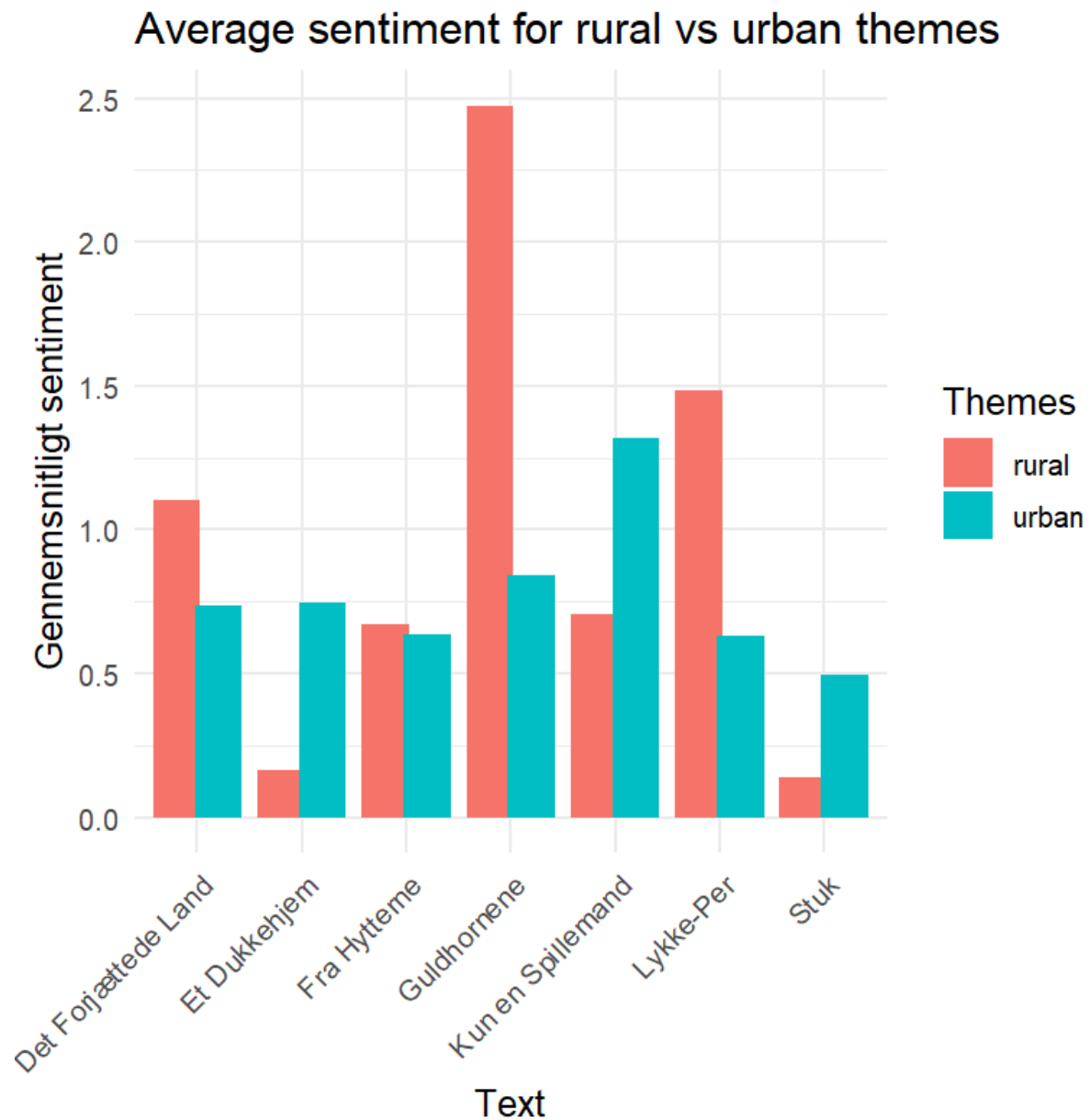
This is a graph showing the mean sentiment scores of our corpus. As an example, we can see that the Adam Oehlenschläger poem 'Guldhornene' shows a significant positive sentiment score of over 2. The rural sentiment score is much lower, at around 0.8. Using Sentida, a positive value is anything above zero, and a negative score below zero. In the case of 'Guldhornene', the rural mean sentiment is still positive, but it is significantly lower, suggesting some animosity or neutrality. It is important to remember that this specific text, as it is a poem, is quite short, and the data is more polarised because of the low word count. In turn this

makes it easier to identify trends in comparison to novels which have more words and complex narrative arcs that can blur the analysis.

Looking more broadly at the graph, we unfortunately cannot identify any general trends, as the sentiment scores are quite close to each other. Indeed, some of the texts even show a more positive score in the urban category.
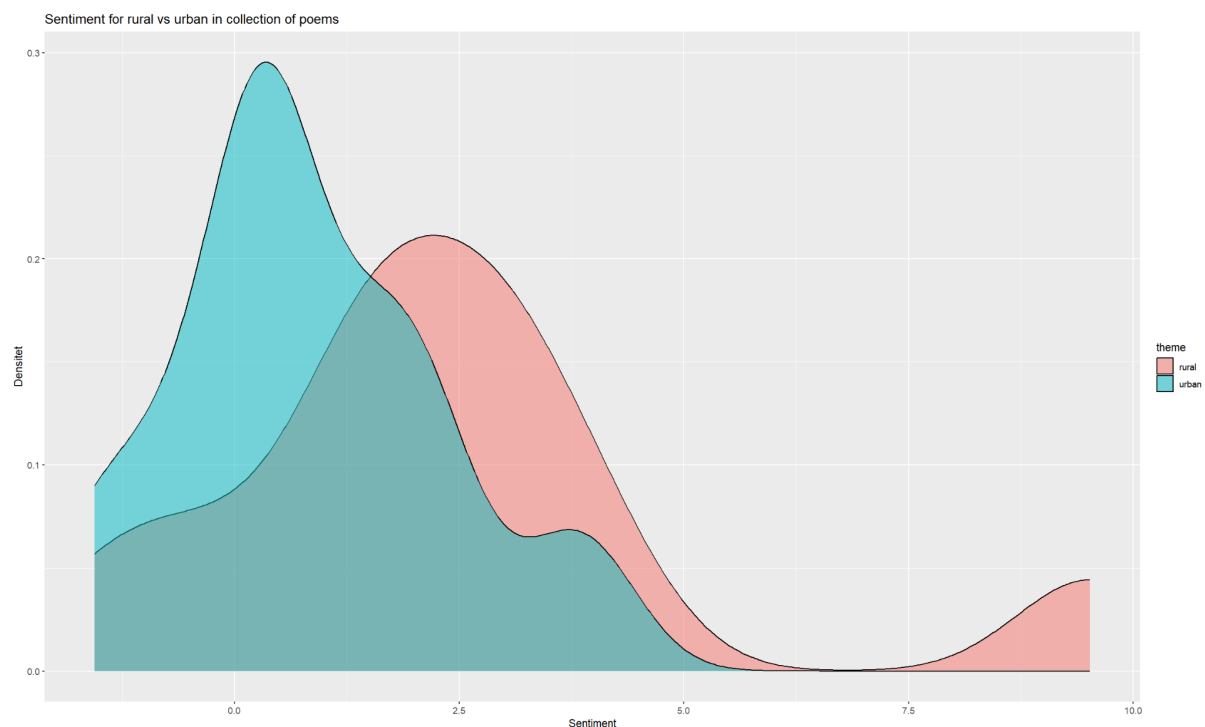

This graph represents the sentiment analysis of a collection of poems from the early half of the 19th century, written by different Danish authors. This lets us know if there could be a difference in sentiment in poetry compared to novels. The graph is a density plot, which shows the sentiment scores on the x-axis and the y-axis the *density*, i.e a measurement of how many data points are around a specific sentiment value. The urban peak is just above a sentiment score of zero, which indicates that city life more often is described very neutrally. On the other hand, rural life is broadly described positively, with the peak between 1 and 2. Generally, the rural category is shifted more towards the higher value sentiment scores, indicating our rural keywords being described more positively.

Sentiment density for rural vs urban comparison
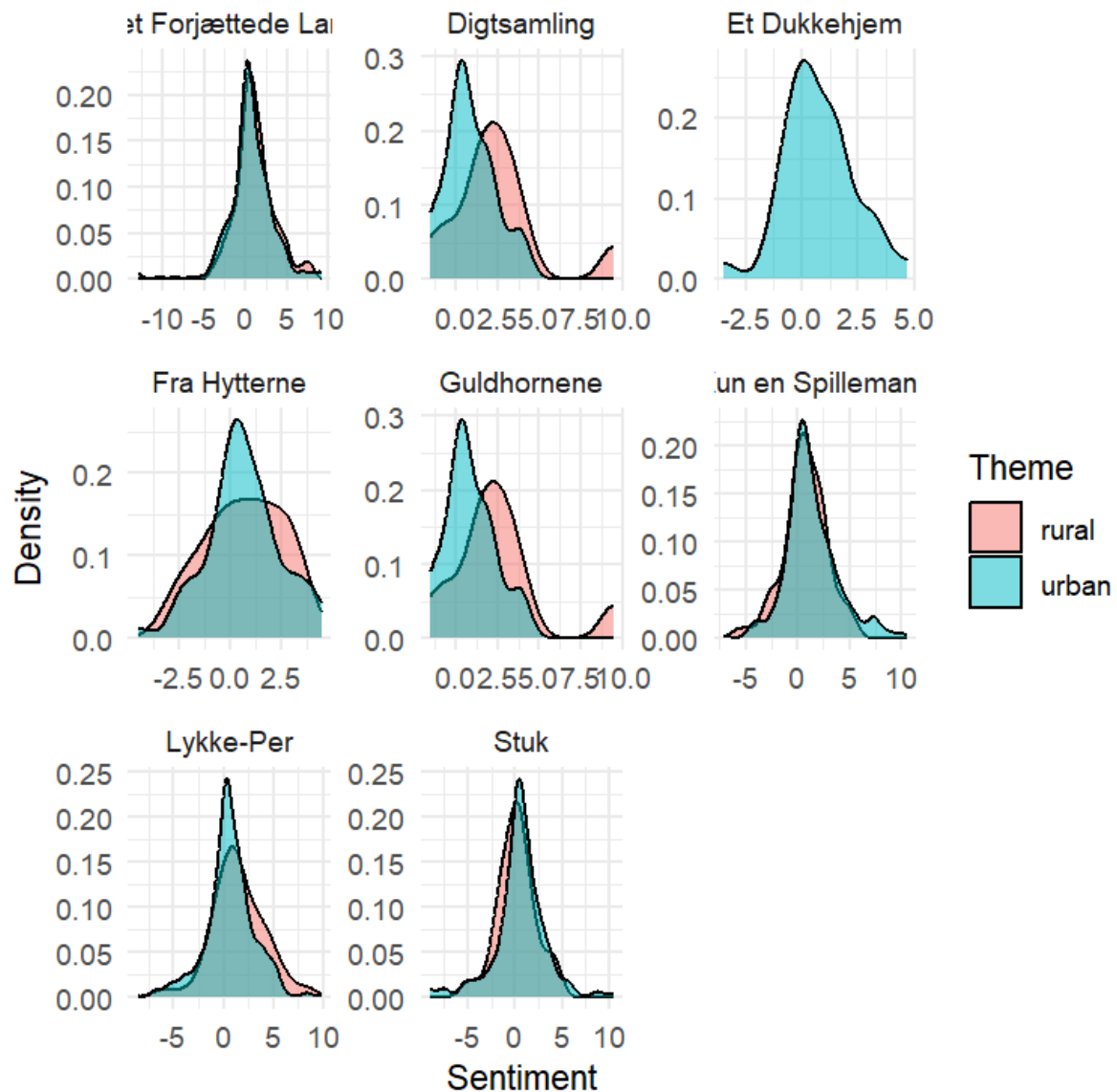
These density plots of our texts show that there, in general, is a lot of overlap between our two categories. This indicates that there is no strong general trend towards rural or city life. In the case of 'Fra Hytterne', rural keywords are more polarised, while urban keywords are more neutrally described. This goes against our thinking, since urban life is often described as being chaotic, but this is not represented in the data. Our collection of poems and 'Guldhornene' by Adam Oehlenschläger show the expected result, i.e rural keywords having more positive sentiment scores.

'Kun en Spillemand' progression
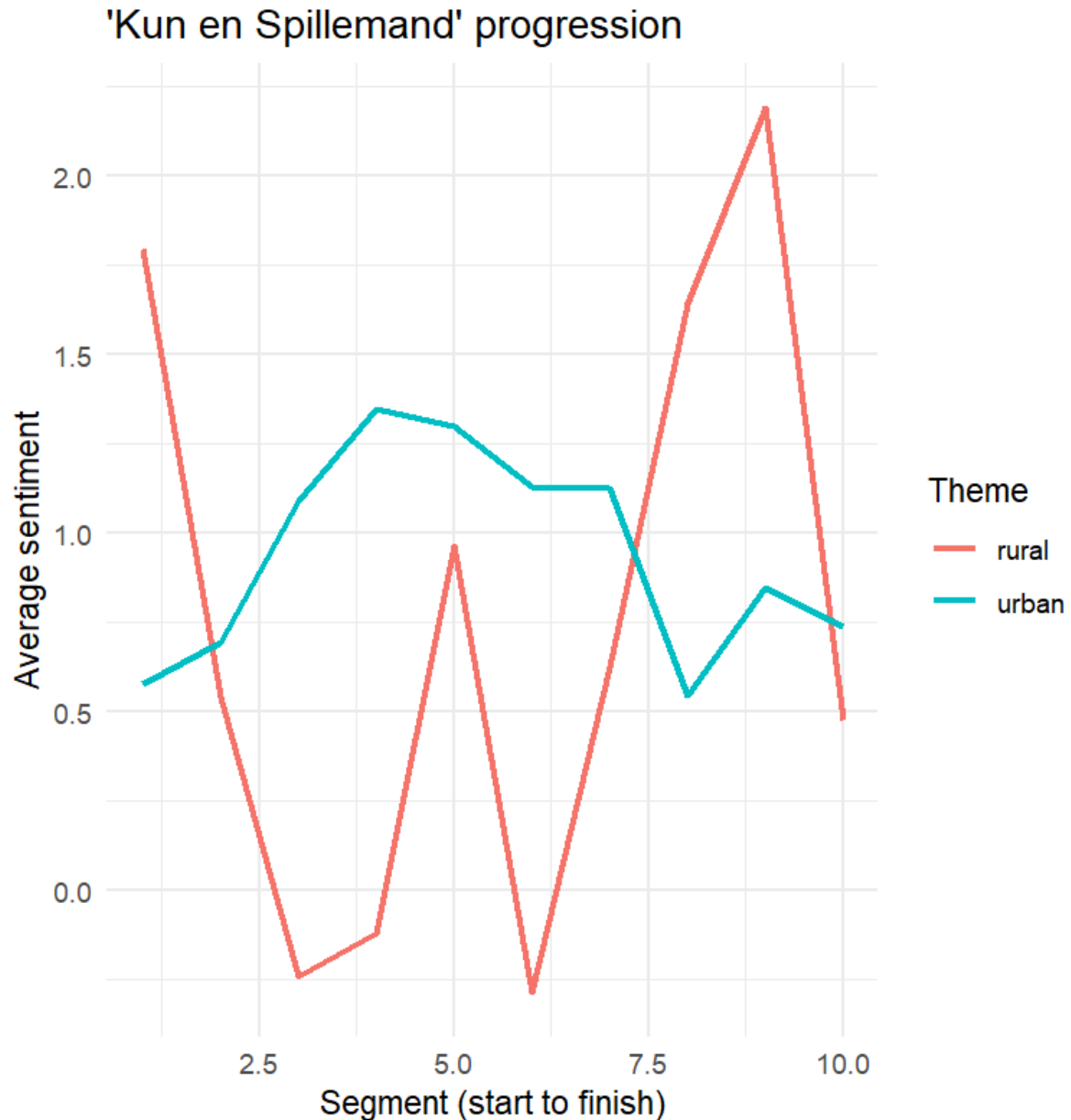
In fiction, there can be a development of themes throughout the books. With this graph, we are attempting to find out the sentiment towards urban and rural life changes throughout the book. To do this, we split the text into 10 segments, and calculated a mean sentiment score for each segment, one for the rural keywords and one for the urban ones. This line plot shows the scores for each segment. Reading the plot, we can see that the rural sentiment score does peak significantly towards the end of the book. This could indicate that the moral of the book is positive towards rural life. However, each segment is shifting dramatically in sentiment score, and by segmenting the book we are reducing the sample size of keywords in each segment, so we cannot say for certain. This segmentation lead to a result that is evidence that can contribute to a confirmation of our hypothesis. Therefore, we tried to segment our other texts, but these did not produce strong results one way or the other, and have thus been left out of this report.

```
keyword context                                                                      sentiment
<chr>   <chr>                                                                          <dbl>
jord    hævn det skulde været at høste denne jord som fædrene har pløjet der nede i       NA
landet  de omliggende herreder ja rundt om i landet ligger der broder og søsteranstalter der ikke  2.71
jord    stod paa bænken i en tør askeagtig jord kunde for den sags skyld godt se        1.28
hede    og snorksov med ansigtet mod det glo hede sand tunge og søvnige med bare ben    -2.11
ager    rigsort til hvilken husmænd og jordløse indsiddere ager op til deres fædre for nu slet  -1.5
jord    en stiv klokketime foran kisten og efter jord paakastelsen ude paa kirkegaarden blev han ved  -2.67
```

Above is a tibble showing 5 examples of where our rural keywords were found in the text 'Fra Hytterne'. The tibble is divided into 3 columns, showing which keyword was found, the 'context', i.e the 5 words in front and behind the keyword, as well as the calculated sentiment score. In the first row, there is an example of one of the issues with our analysis; if a word is not present in the sentida lexicon, it cannot calculate a sentiment score, therefore returning NA. It is likely the old spelling of "skulde" instead of "skulle" that is the issue. Sentida has not been built for analysing 19th-century texts, and the lexicon lacks many old variations of words. The keyword "ager" could be another false positive, as in this context it might mean something akin to "looking up to" rather than the traditional definition.

The results

We were not able to confirm our initial hypothesis of rural life sentiment being more positive in comparison to urban life sentiment. Our analysis points towards a tendency of the rural area being described in a more neutral tone and the urban area being more polarised. Relative to rural keywords, urban keywords more often have a positive or negative sentiment score. In general, we could not conclude a general trend of urban or rural life having a more positive sentiment in the literature we chose. The exception is the collection of poetry, which had a strong positive rural sentiment, which makes sense considering the genre definition of Romantikken. Perhaps the relatively shorter poems, in comparison to the novels, allowed sentiment trends to show up more clearly because they weren't diluted by a higher word count.

Representativeness

One of the clearest flaws in our analysis is the relatively low amount of literary data we have used. Since we have only used 9 novels and 12 poems, our sample size is not big enough that we can confidently call it representative of Danish 19th century literature. Multiple biases may have been introduced, for example in the selection of texts, which has been done with a 21st century view of which texts are important and representative of the 19th century. The limited amount of data also reduces our ability to generalize, as it reflects only a small fraction of the literature produced in the 19th century, and it thereby becomes difficult to say if this limited amount of data does represent the literature as a whole.

The literature represents some of the most popular novels and poems from some of the most known authors, like H.C. Andersen, Adam Oehlenschläger, and Søren Kierkegaard. Even though this could point towards a greater representativeness of 19th century Denmark, we lack metadata to answer how popular these works actually were in the 19th-century.

You can question if our data should have been from later in the 19th century, as some were from before the 1840s, which is where the earliest signs of industrialisation can be seen in Denmark. As such, these texts can give us an idea of the view of urban and rural life before industrialisation, but not during or after. This is an interesting point of comparison to the rest of the texts written during the industrialisation.

Issues and lessons learned

We ran into multiple issues during our project. In the initial phase, we wished to measure which feelings certain words evoked as can be done with some lexicons. We quickly found out that the Danish lexicon Sentida does not have this capability. Manually assigning feelings to words or training a model was outside the scope of the project, so we had to abandon this idea.

During data selection, we found multiple novels we wanted to use that were simply unavailable in a digital format, which was necessary for our analysis. This introduced a bias in our project, because our data selection was influenced by what was available.

Keyword selection has a tremendous impact on the results, therefore this should be done with much care and some trial and error to identify which keywords aren't working. We experienced some issues with designing the list of keywords for our rural and urban categories. As words often have multiple meanings, our sentiment analysis was flawed. Words such as "lade" had to be removed from the rural keyword list. Another issue was Sentida not recognising old variations and spellings of words. This issue in particular has had more impact on the analysis than we would like, but there unfortunately is no reasonable alternative that would be feasible. In a bigger project you could use machine learning or compile words manually. The key lesson learned is that Sentiment analysis is a cool and useful tool, that often has some issues you need to be aware of to make it as accurate as possible.

Usefulness

Although our results do not match up with our Hypothesis, they are still useful. Our results point towards the city, which in the 19th-century for most people was a new place, or atleast a place of great change to also be a subject of an increased polarization when it comes to how it is described. In spite of that it's very important to be aware of the limited scope of our research and the lag of representativeness, and sentimental problems which makes generalisation hard.

Our results do anyhow also go against Hannu Salmies point in the book "19th Century Europe
- A Cultural History" where he determined that literature generally points towards a more positive view of the countryside compared to the city[10]. This could simply be a result of the weakness in our sentimental analysis, or our lag of representativeness, nevertheless could this also point towards a difference between how the Danish population saw the city in the 19th-hundrede compared to other parts of europe.

---

[10] Salmi, *Nineteenth-Century Europe*, 79.

Conclusion

Through a sentimental analysis of literary text made through the 19th-hundreds in Denmark, we looked for a pattern of how the countryside is talked about compared to the city, to see if especially the countryside should have become idolized in the new era of urbanisation and industrialization.

The results of our analysis shows that the urban area is being described more polarizing than the countryside, by both being described more positively and more negatively in 19th-century literature than the rural area, which in kind generally is being described with more neutral words. Our Hypothesis is thereby disproved by our results. This could point towards a different vision of how the rural and urban setting is seen in 19th-century Denmark. Our analysis has some representativeness issues, largely because of low sample size and selection bias of our data. Further work could be done to increase the accuracy of our analysis using machine learning.

Recommendations

We would recommend caution when choosing literature as data for your sentimental analysis, the resources required to get enough data and at the same time making sure that there is an equal representativeness between the different kind of literature genres can be surprisingly high, we would therefore recommend sticking to one genre of literature for more resource limited studies.

Future studies using sentiment analysis on fictional literature could benefit from more focus on the value of the moral of the story, and figurative language that may be lost in a digital analysis. Here we recommend using a sentimental analysis tool which tries to use the combined sentence as data instead of just using individual words. Machine learning could be used to understand the texts in a broader context, leading to a more accurate result. As we attempted with segmentation, the possibility of giving more sentimental value to the end of the text than the rest of the story. as a way to try and catch the moral of the story could also be interesting.

Sources

Jørgensen, Claus Møller. "Fra enevældig helstat til nationalstat, 1814-1914". I *En Danmarks Historie fra vikingetid til nutid*, af Thorsten Borring Olesen og Paulsen Bjørn, 2. udg. Aarhus universitetsforlag, u.å.

Salmi, Hannu. *Nineteenth-Century Europe*. Polity Press, 2008.

Schmidt, Thomas, og Manuel Burgardt. "An Evaluation of Lexicon-based Sentiment Analysis Techniques for the Plays of Gotthold Ephraim Lessing". Sintemntal Analysis, u.å.

Tahmasebi, Nina, og Simon Hengchen. "Samlaren" 140 (2019).

Ukendt. "Litteraturhistoriske oversigter". Dansksiderne.dk, u.å. https://dansksiderne.dk/index.php?id=3440.

Thanks to

We would like to thank Fellow students Asger Brylle, Nikolaj Vørstein, Laurits Johansen and Sejer Christensen. who have helped a great deal with developing parts of our portfolio in digital methods.

We would also like to thank Adela Sobotkova and her assistants who have helped us understand R studio, and sentimental analysis. Her teachings and consulting have been an instrumental part in this projekt.

We would also like to thank Søren Poder, who has helped us understand the more theoretical part of using sentimental analysis. His teaching in the many drawbacks and potential problems to be aware of when using sentimental analysis have been invaluable.

Conceptualization: Casper
Methodology: Marcus Horsager Olsen
Digital data wrangling, analysis and visualisation: Casper
Writing - original draft preparation: Marcus Horsager Olsen
Writing - critical review and overall editing for uniform voice: Casper

Software metadata description	*Please fill in this column*
Permanent link to Github repository where you put your script(s), R project, and data	*Link to Github:* *https://github.com/Flakdk99/Casper-og-Marcus-DAM-Final-projekt.git*
Software License	*CC BY-NC-SA 4.0*
Data License	*List one of the approved licenses, e.g. Creative Commons 4.0; see Week 17 lecture for more*
Software versions, Installation requirements & dependencies for software not used in class	*You need to have ImageMagick 3.8 installed*
If available Link to software documentation for special software (only	*Example* *http://mozart.github.io/documentation/*

relevant if you go outside the scope of class)	
Support email for questions	marcushorsagerolsen93@gmail.com // ccommerou@gmail.com