COMPUTATIONAL APPROACHES IN THE HUMANITIES: FROM SENTIMENT ANALYSIS TO DEEP LEARNING COLORIZATION

A Dissertation

by

HOYEOL KIM

Submitted to the Graduate and Professional School of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Maura Ives
Committee Members,	Amy Earhart
	Shawna Ross
	Ruihong Huang
Head of Department,	Maura Ives

May 2022

Major Subject: English

ABSTRACT

This dissertation is article-based, consisting of four chapters with two themes: colorization (chapters 1 & 2) and sentiment analysis (chapters 3 & 4). Chapter 1, "Victorian400: Colorizing Victorian Illustrations," reveals my methods for creating, curating, and validating the Victorian400 dataset for colorizing Victorian black-and-white illustrations. Victorian400 is a nineteenth century illustration dataset consisting of 400 colorful images that is helpful for testing and developing deep learning models. I tested the Victorian400 dataset with the pix2pix model, which is a conditional generative adversarial network (cGAN) model, to verify the dataset for colorizing black-and-white illustrations from the nineteenth century. Chapter 2, "Case Study: Using Machine-Colored Illustrations of Charles Dickens's Fiction in the Classroom," addresses the pedagogical usage of machine-colored illustrations in the English classroom based on my case study. I discovered that students often preferred either no illustrations or machine-colored illustrations to hand colored ones. In chapter 3, "Sentiment Analysis: Limits and Progress of the Syuzhet Package and Its Lexicons," I explore the limits and progress of the Syuzhet package, a sentiment analysis tool in R, for the sentiment analysis of literary texts. I compare the four sentiment lexicons (Syuzhet, Bing, Afinn, and NRC) used in the package. I also test Syuzhet, SentimentAnalysis, sentimentr, RSentiment, and VADER (Valence Aware Dictionary and sEntiment Reasoner) with seven different sentences to see how each lexicon-based sentiment analysis tool generates sentiment scores. In chapter 4, "Dickensian Sentiment and Sentiment Analysis of Victorian Novels," I perform sentiment analysis on three Victorian novels using BERT (Bidirectional Encoder Representations from Transformers) in tandem with a dataset I created for the sentiment analysis of Victorian fiction, in order to see if sentimentality is revealed

in Charles Dickens's *Our Mutual Friend*, George Eliot's *Middlemarch*, and Charlotte Brontë's *Jane Eyre*, since sentimentalism was a primary social value in Victorian society.

DEDICATION

가족을 위해 희생만 하다 하늘에 가신 아버지,
언제나 저를 묵묵히 지지하시는 어머니,
나의 사랑하는 아내에게
이 논문을 바칩니다.

ACKNOWLEDGEMENTS

I would like to express my appreciation and gratitude to my chair, Dr. Maura Ives for her time, advice, feedback, and support during my PhD. Her support, kindness, and encouragement helped me to successfully complete this dissertation. I also wish to thank Dr. Amy Earhart for her advice, support, and feedback. Finishing the Directed Studies with Dr. Earhart in Spring 2019 was a stepping-stone for me to become a digital humanist. I also wish to thank Dr. Shawna Ross and Dr. Ruihong Huang for their help and support.

I am grateful to the faculty and staff in the Department of English at Texas A&M University for giving me the opportunity to pursue my PhD with a graduate teaching assistantship. I wish to acknowledge *Digital Humanities Quarterly* and Edinburgh University Press for their generous policies to allow my articles to be included in this dissertation. I am also thankful to the digital humanities community for providing feedback on my projects.

I would like to express my appreciation to Lindsey Kim for editing suggestions and Kyle Fisher for technical advice. I would also like to thank Shawna Campbell for reading my drafts and giving me good feedback.

Finally, I wish to thank my wife, my mother, my brother, my sister-in-law, and my parents-in-law for their support and love.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This dissertation was supervised by a dissertation committee consisting of Professor Maura Ives (advisor) of the Department of English; Professor Amy Earhart of the Department of English; Professor Shawna Ross of the Department of English; and Professor Ruihong Huang of the Department of Computer Science and Engineering.

All other work conducted for the dissertation was completed by the student independently.

Funding Sources

For Chapter 1 “Victorian400: Colorizing Victorian Illustrations” I received the Professional Development Grant and Graduate Merit Funding from the Department of English, Texas A&M University and the Project Development Grant from the Center of Digital Humanities Research, Texas A&M University.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

The digital humanities are an interdisciplinary field that draws upon both qualitative and quantitative research methods. The term ‘digital humanities’ stems from ‘humanities computing,’ the term used to refer to the field before the ‘digital humanities’ was coined by John Unsworth in 2004 (Earhart, par. 2). Using text encoding was “typically seen as a core element of humanities computing” (Svensson, par. 45), whereas the digital humanities extend to a wider variety of scholarly fields such as literary studies, black studies, history, linguistics, information studies, and computer science and engineering; to describe the breadth of topics encompassed by the digital humanities, the term “big tent” was coined by Glenn Worthey and Matthew Jockers in 2011. The digital humanities have bloomed as an interdisciplinary field in the last decade, with interdisciplinary collaboration and the usage of both digital skills and humanities knowledge producing valuable and creative research. While computational methods can help researchers gather and find the latent meaning of data and can facilitate the creation of visualizations of complicated data and patterns, they cannot provide subjective interpretation and argumentation. The digital humanities therefore act as a space for the complementary relationship between digital and humanities research methods to flourish. Every year, a broad scope of interdisciplinary research has been introduced at DH conferences. In the DH2020 conference, for instance, topic categories across several disciplines under the big tent were presented, including “Digital Activism and Advocacy,” “Open Access Methods,” “Public Humanities Collaborations and Methods,” “Cultural Analytics,” and “Artificial Intelligence and Machine Learning.” Through interdisciplinary research between scholars, students, and professionals across a variety of disciplines, the digital humanities have made a great impact on the community, industry, and academia.

This dissertation is article-based, consisting of four chapters. There are two themes in this dissertation: colorization (chapters 1 & 2) and sentiment analysis (chapters 3 & 4). I principally deploy deep learning for computer vision and natural language processing as a research method to perform colorization and sentiment analysis, with an aim to contribute to literary studies and the digital humanities through arguments, suggestions, and new findings. The first two chapters, "*Victorian400*: Colorizing Victorian Illustrations" and "Case Study: Using Machine-Colored Illustrations of Charles Dickens's Fiction in the Classroom," are about the *Victorian400* dataset for colorizing Dickens's illustrations with deep learning and the pedagogical usage of machine-colorized Victorian illustrations in the classroom. In the last two chapters, "Sentiment Analysis: Limits and Progress of the Syuzhet Package and Its Lexicons" and "Dickensian Sentiment and Sentiment Analysis of Victorian Novels," I explore the limits and progress of the Syuzhet package, a sentiment analysis tool in R, for the sentiment analysis of literary texts and perform sentiment analysis on three Victorian novels using BERT (Bidirectional Encoder Representations from Transformers) in tandem with a dataset I created for the sentiment analysis of Victorian fiction.

Chapters 1 & 2: Colorization

Deep learning has been applied in a variety of domains such as medical science, business, education, and pertinently, the humanities. For example, advanced deep learning models can be used to transform input images stylistically (Karras et al. 2019), enhance image resolution (Dong et al. 2017), and colorize illustrations (Kim 2021). Deep learning is a type of machine learning under the umbrella of artificial intelligence. Deep learning, which consists of deep neural networks, can analyze texts, images, and sound based on trained models. Due to deep neural networks, deep learning can process more complicated data structures than traditional machine learning. The idea of machine learning began in 1949 with Donald Hebb, who introduced Hebbian theory, which describes how neural networks are shaped when our brains learn something new. In 1957, Frank Rosenblatt proposed the notion of perceptron, which is an algorithm for creating a result using various inputs. Based on his previous research, Arthur Samuel, who worked for IBM, coined the term "machine learning" in 1959. After that, however, there were few major breakthroughs in research on artificial neural networks until David Rumelhart, Ronnald Williams, and Geoffrey Hilton introduced how to train multilayer neural networks through backpropagation in 1986. Since then, due to constant research on neural networks that has worked to solve issues such as overfitting and long training times, deep learning has been deployed in a variety of domains involving images, texts, and sound.

Ever since GANs (Generative Adversarial Networks), which are designed with an adversarial discriminator and generator, were first introduced by Goodfellow et al. in 2014, generative models have shown promising results for the generation of images. Based on GANs, deep learning researchers have developed new GAN-derived models such as StyleGAN, CycleGAN, and SRGAN. Karras et al. introduced StyleGANs (Style Generative Adversarial Networks), which transform images stylistically based on datasets. Zhu et al. created CycleGANs (Cycle Generative Adversarial Networks), which create images based on the mapping between an input image and an output image. Ledig et al. presented SRGANs (Super-Resolution Generative Adversarial Networks), which upscale image resolutions. In the first two chapters, I deploy the pix2pix model, which is based on cGANs (conditional Generative

Adversarial Networks), with the *Victorian400* dataset in order to colorize black-and-white Victorian illustrations.

In the first chapter, "*Victorian400*: Colorizing Victorian Illustrations," I introduce and validate the *Victorian400* dataset, which was created to colorize black-and-white illustrations from the nineteenth century. The *Victorian400* dataset is a nineteenth century illustration dataset consisting of 400 colorful images that provides an opportunity for deep learning learners to run code easily without high performance devices. I first explain how I created and curated the *Victorian400* dataset, then I test the *Victorian400* dataset with the pix2pix model, which is a conditional generative adversarial network (cGAN) model, to verify the possibilities of colorizing black-and-white illustrations from the nineteenth century. The second chapter, "Case Study: Using Machine-Colored Illustrations of Charles Dickens's Fiction in the Classroom," covers the pedagogical usage of deep learning colorization in the literature classroom. In this chapter, I aim to explore the ways in which the *Victorian400* dataset shows possibilities for improving students' ability to better understand the intersections of text and image. I first introduce how Charles Dickens collaboratively worked on illustrations with his illustrators, then examine my pedagogical case study with original and colorized illustrations in the literature classroom. The case study in this chapter shows that deep learning-based colorized illustrations with the *Victorian400* dataset contribute to students' learning in the literature classroom by providing entertainment and facilitating students' visualization of the text.

Chapters 3 & 4: Sentiment Analysis

Similar to how the GAN model helped in the development of computer vision in deep learning, the transformer, which was introduced by Google Research through their article

“Attention is All You Need,” opened new avenues in natural language processing. Transformers work based on parallel operations, making it possible to calculate big data with more accuracy and speed, as well as not requiring the data to be processed in order, unlike with RNNs (Recurrent Neural Networks). Transformer language models have been used for a variety of domain tasks. One of the representative transformers, the BERT model, which is contextual and bidirectional, can be used for several tasks such as translation, sentiment analysis, and text summarization. Likewise, the GPT models developed by OpenAI had a great impact on industry and academia; news articles and blog posts written by GPT-3 are difficult to distinguish from human written ones. These transformer language models can be used together in domain tasks, such as for contextual question and answer generation tasks using both GPT-2 and BERT (Ahn et al. 2021). Furthermore, Linformer, an advanced transformer model, has been deployed for censorship on both Facebook and Instagram (Wang et al. 2020).

Along with exploring deep learning methods, in this dissertation I also deploy sentiment analysis as a research method and examine Syuzhet, which is a lexicon-based sentiment analysis tool for literary texts. There have been qualms about sentiment analysis by digital humanists since Swafford instigated debates on the limits of Syuzhet in 2015. Despite the imperfection of Syuzhet, it has been the most popular sentiment analysis tool in R, with a continuous increase in downloads over the last five years (see figure 1.1). While discussing the limits of Syuzhet, I impart that deep learning-based sentiment analysis methods using transformers, such as the BERT and RoBERTa models, produce more accurate results than Syuzhet when performing sentiment analysis. Some digital humanists have begun to apply deep learning-based sentiment analysis with small datasets, such as with poems, but the overall resistance in the field against sentiment analysis and the scarcity of literary datasets for deep learning remains a barrier for research. Therefore, over two chapters of this dissertation, I explore the limits of Syuzhet and perform sentiment analysis of Victorian fiction by deploying the BERT-Base model with the VictorianLit dataset, in order to suggest ways to improve sentiment analysis based on current research about deep learning-based sentiment analysis.

In the third chapter, "Sentiment Analysis: Limits and Progress of the Syuzhet Package and Its Lexicons," I impart the limits and progress of Syuzhet through my experiments with its lexicons and nineteenth-century British novels. I compare the results of sentiment analysis with the four different lexicons in the Syuzhet package, then compare the parsing results with different parsers based on the different versions of Syuzhet to show the improvements made over time. After that, I compare the sentiment analysis results of Syuzhet with other sentiment analysis tools such as SentimentAnalysis, sentimentr, RSentiment, and VADER (Valence Aware Dictionary and sEntiment Reasoner). Lastly, I delve into the limits of Syuzhet through the sentiment analysis of Charles Dickens's *Our Mutual Friend*, George Eliot's *Middlemarch*, and Charlotte Brontë's *Jane Eyre*. In the fourth chapter, "Dickensian Sentiment and Sentiment Analysis of Victorian Novels," I examine the relationship between sentimentalism in Victorian novels and sentiment analysis in the digital humanities. After briefly going over the definition of sentimental, sentimentality, and sentimentalism, I look into Charles Dickens's moral sentiments. I perform sentiment analysis, fine-tuning the BERT-Base model with the VictorianLit dataset, which consists of five Victorian novels: Charles Dickens's *Little Dorrit* and *Oliver Twist*, Elizabeth Gaskell's *North and South*, George Eliot's *Adam Bede*, and Mary Elizabeth Braddon's *Lady Audley's Secret*. When training the BERT-base model, I assigned five different sentimental values: 0 (very negative), 1 (negative), 2 (neutral), 3 (positive), 4 (very positive). I created the VictorianLit dataset for the deep learning-based sentiment analysis of Victorian literary texts, to see if sentimentalism is revealed through the sentiment analysis of three Victorian novels: Charles Dickens's *Our Mutual Friend*, George Eliot's *Middlemarch*, and Charlotte Brontë's *Jane Eyre*. While sentimentalism is difficult to detect through sentiment analysis, sentiment analysis can help distant readers grasp the flow of emotions in Victorian novels by providing meaningful results for analyzing sentiment plots and measuring sentiments within sentences.

I created, curated, and publicly shared every dataset used in this dissertation as an act of participation in the open data movement. The datasets that I created have been shared and used by students, practitioners, and scholars across several fields. As a digital humanist who deploys computational approaches in the humanities, my aim has been for this dissertation to contribute to the development of the digital humanities through the deployment of deep learning methods in the humanities, the creation of humanities datasets for deep learning, and the facilitation of students' learning through digital tools.

Bibliography

Ahn, Sungjin, et al. "The Automatic Generation of Contextual Questions and Answers for English Learners." *ACH 2021*. dx.doi.org/10.17613/pp84-yr61.

Dong, Hao, et al. "Tensorlayer: a Versatile Library for Efficient Deep Learning Development." Proceedings of the 25th ACM International Conference on Multimedia. 2017.

Earhart, Amy. *Traces of the Old, Uses of the New: The Emergence of Digital Literary Studies*. U of Michigan P, 2015.

Karras, Tero, et al. "A Style-Based Generator Architecture for Generative Adversarial Networks." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019.

Kim, Hoyeol. "*Victorian400*: Colorizing Victorian Illustrations." *International Journal of Humanities and Arts Computing*, vol. 15, no. 1-2, 2021, pp. 186–202.

Klein, Matthew, and Lauren Klein. "A DH That Matters." *Debates in the Digital Humanities 2019*, edited by Matthew Gold and Lauren Klein, 2019, pp. ix-xiv.

Ledig, Christian, et al. "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network." Proceedings of the IEEE Conference on Computer Vision. 2017.

Svensson, Patrik. "Humanities Computing as Digital Humanities." *Digital Humanities Quarterly*, vol. 3, no. 3, 2009. digitalhumanities.org/dhq/vol/3/3/000065/000065.html.

Wang, Sinong, et al. "Linformer: Self-attention with linear complexity." 2020. arXiv:2006.04768.

Zhu, Jun-Yan, et al. "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks." Proceedings of the IEEE Conference on Computer Vision. 2017.

CHAPTER I

Victorian400: Colorizing Victorian Illustrations[*]

1. Introduction

If we view data as a social and cultural artifact, datasets can be viewed as snapshots of the social and cultural climate of a time period. Tanya Clement and Amelia Acker assert that "[d]ata is not just given," but that culture is formed and changed by data. Data is inclusive of history, life, arts and literature, and reflects contemporary culture, society, and economy. Data, however, must be properly trimmed for usage in computation. "Data curation" includes 'the act of discovering a data source(s) of interest, [then] cleaning and transforming the new data" (Stonebraker at el. 1). The importance of data curation has been discussed by digital humanists, especially regarding the need to curate data in order to avoid faulty, biased or distorted results and reflect diverse and global data.[1] When data scientists look into data, they focus more on whether the data is sufficiently well curated enough to be trained for models.[2] Data scientists usually spend a large portion of their time curating datasets, as well-curated datasets are crucial to achieving state-of-the-art results in the deep learning field. Datasets curated by digital humanists, however, have been mainly created for other humanists in the form of resources such as digital archives, rather than for machine learning/data scientists. Publicly shared humanities datasets for machine learning/data scientists on Kaggle and GitHub are mostly created by non-humanists. Therefore, in order to for datasets to be tailored specifically to humanities-related deep learning tasks, I

[*] This article "*Victorian400*: Colorizing Victorian Illustrations" was published by Edinburgh University Press in the *International Journal of Humanities and Arts Computing*. The Version of Record is available online at: https://www.euppublishing.com/doi/abs/10.3366/ijhac.2021.0269.

believe it would be beneficial for digital humanists, with their knowledge of both fields, to create, curate, analyze, and share their own deep learning datasets.[3]

According to Goodfellow et al., "Generative adversarial networks (GANs) are a class of methods for learning generative models based on game theory" (1). Since GANs, which have proved to have an astonishing ability to create realistic high resolution images, were first introduced by Ian Goodfellow in 2014, they have been applied to a variety of imaging processes to predict the color of images, transfer colors for a style, or produce higher resolutions in the conditional setting based on either conditional or unconditional GANs, such as image-to-image translation (Isola et al.), text-to-image synthesis (Reed et al.), and image super-resolution (Ledig et al.). Famously, the *Portrait of Edmond de Belamy*, the artificial intelligence artwork created by Obvious, a group of artists consisting of three French students who borrowed from Robbie Barrat's code based on GANs, was sold for $432,500 in 2018, although there are technical defects in the *Belamy* portrait such as texture quality and low-resolution issues. In addition, there has been progress in colorizing photos and videos based on GAN-derived models, such as pix2pix (Isola et al.), DeOldify (Antic) and Paints Chainer (Taizan).

There have been attempts to colorize images by hand in order to improve the imaginational response in readers, as well as provide anticipation and interest, but hand-colored results vary depending on the style of the illustrator and such methods are time- and cost-ineffective. In contrast, deep learning-based colorization is cheap and fast, in addition to creating consistent results based on models trained with datasets. There are a variety of possible deep learning colorizing projects in the digital humanities; colorizing black-and-white Victorian images such as those found in the works of Charles Dickens, Thomas Carlyle, and John Ruskin would allow viewers to connect emotionally with the figures.[4] The colorization of black-and-white war photographs, as another example, could open up avenues for modern viewers to more easily conceptualize the past. For the project outlined in this article, I decided to tackle the colorization of Victorian illustrations, since as yet there has been no attempt to colorize illustrations from the nineteenth century using deep learning. Most illustrations from the Victorian era were printed in black and white due to the higher printing cost of color illustrations.[5] For example, the works of Charles Dickens, who helped pioneer Victorian illustrations by actively including illustrations in his fiction, include only four hand-colored illustrations, all printed for his story *A Christmas Carol* (1843). Colorizing the illustrations found in Dickens's works has the potential to enhance readers' understanding of the text and open up new interpretive possibilities. However, as yet no datasets of nineteenth-century illustrations have been made available for deep-learning-based colorization. As someone who has been trained for research and work in the fields of both the humanities and computer science, I have created, curated and publicly shared the *Victorian400* dataset for the deep learning colorization of illustrations from the Victorian era. The *Victorian400* dataset is a collection of Victorian color illustrations, which provides an opportunity for deep learning learners to run code easily without high-performance devices,[6] helps machine learning/data scientists to test and develop deep learning models,[7] and, as a pedagogical tool, shows the possibilities of improving students' ability to better understand the intersections of text and image.[8] In this article, I introduce the *Victorian400* dataset, reveal how I decided what to include and exclude in the process of curating the dataset, and examine the results of the test set with the trained set to see whether the *Victorian400* dataset produces reasonable results.[9]

2. *Victorian400* dataset

The *Victorian400* dataset, an open source shared on Kaggle[10] and GitHub,[11] is a nineteenth-century illustration dataset consisting of 400 illustrated images (see Figure 1.1). It has been downloaded, experimented with, and shared by data scientists and the deep learning community.

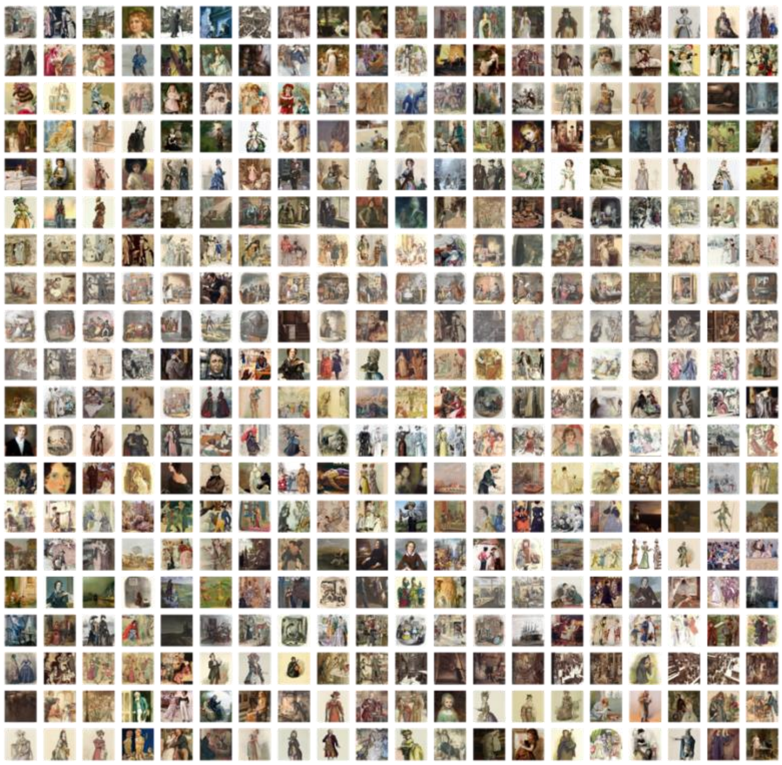


Figure 1.1: The *Victorian400* dataset.

Regarding the curation of the 400 images for the *Victorian400* dataset, I outlined four criteria: 1) the images should be from the Victorian era; 2) the painting style of the images should not be too idiosyncratic (Drawings with Aubrey Beardsley's style, for example, do not represent illustrations of the Victorian age); 3) the images should derive from a variety of illustrators and not focused on a single illustrator, and 4) the drawing style of the images should be close to that of the illustrations commonly found in Victorian novels.[12]

The *Victorian400* dataset was originally created to colorize Charles Dickens's black-and-white illustrations for a pedagogical purpose. While researching into the area of colorizing illustrations, I found deep learning datasets for pictures, shoes, animals, and anime, but was unable to find nineteenth-century illustration datasets created for deep learning.[13] Therefore, I had to create my own dataset, to include illustrations with similar drawing styles to those found in Dickens's fiction, in order to carry out imaginative colorizations of the Dickens illustrations. The scope of this dataset had to be expanded to all nineteenth-century illustrations due to the fact that most of the illustrations in Dickens's books exist only in black and white. I aimed to find colorful illustrations that were similar in style to the black-and-white illustrations drawn by George Cruikshank, Hablot Knight Browne (Phiz), and John Leech for Dickens's works.

To do this, I first coded web crawling tools to scrape together around 3,000 images from websites such as *The Victorian Web* and *The Charles Dickens Page*. Web scraping technology is beneficial to data scientists, but it requires a great deal of time to curate the scraped data. According to Aurélien Géron, "if your training data is full of errors, outliers, and noise (e.g., due to poor-quality measurements), it will make it harder for the system to detect the underlying patterns, so your system is less likely to perform well" (27). An extensive amount of experimentation is required to obtain reasonable results for training data. In general, the errors of training data such as overfitting and underfitting stem from poor-quality datasets, datasets that are too small, or models that are too complicated. In order to have credible data for deep learning, data must be curated based on specific and consistent criteria. Accordingly, using my pool of collected images, I spent the majority of my time in this project selecting appropriate images for the *Victorian400* dataset and curating the files in order to improve the accuracy of the colorization results. For example, I removed images that were clearly outliers, specifically those which included a preponderance of red or green colors, such as pictures of Santa Claus. I found that Victorian illustrations commonly feature red and green colors, but when, for example, the dataset included a series of illustrations colored with a lot of green, such as in scenes dominated by trees, the test results generated images with too much green, sometimes coloring in green what should not be green, for instance skin. Whenever the dataset was unbalanced, no matter how much data I tested with, the validation results were poor and biased.

The *Victorian400* dataset contains three different folders: original, resized, and gray. Images in the "original" folder were curated for significant parts such as faces and bodies, which were not to be cut in the process of resizing. In the "resized" folder were images resized to 256 x 256 for the process of deep learning. Based on these resized images, I created the "gray" folder, which includes black-and-white images converted from the resized color images.

3. Experiments

For colorizing illustrations from the nineteenth century, I chose the pix2pix model built based on cGANs by Isola et al. The pix2pix model performs automatic graphic operations on photographs based on cGANs by learning from datasets. The pix2pix model proved able to solve a number of issues when translating an input image into a corresponding output image by showing test results with a variety of datasets.[14] The GAN model suggested by Goodfellow et al. has the generator and the discriminator learn and compete with each other in order to generate the best outputs, which are referred to as adversarial nets. Likewise, the conditional GANs that pix2pix deploys consist of two main poles, the generator and the discriminator. In pix2pix, the generator colorizes input images by transforming them into output images through several steps using a series of encoders and decoders. A series of encoders, which include convolution and activation, helps the generator compress input images throughout the layers, and a series of decoders, which contain deconvolution and activation, decompresses them. However, the pix2pix model deploys a "U-Net" instead of simply using an encoder-decoder for the generator in order to improve performance when predicting the colors of input images (Isola et al. 3).

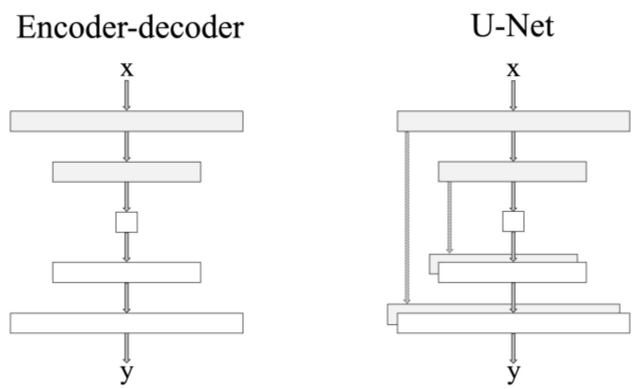


Figure 1.2: Simple depiction of differences between an encoder-decoder and U-Net.

The difference between an encoder-decoder and U-Net is that the U-Net has added skip connections. The skip connections help each path achieve better localization by combining context from the connection of encoder layers (the downsampling path) and decoder layers (the upsampling path). For example, the size of feature maps in the first layer of the encoder (256 x 256 x 3) is the same as in the last layer of the decoder. Each layer of the encoder is merged with the corresponding layer of the decoder. U-Net, a convolutional network for image segmentation, is faster than a sliding-window approach (Ciresan et al.), which is a brute-force solution and often considered inefficient. A sliding-window approach, which predicts the label of each pixel by providing a local patch from input images, has two drawbacks: slowness and a "trade-off between localization accuracy and the use of context" (Ronneberger et al. 2). In contrast, U-Net makes it possible to get both local and contextual information quickly. In addition, the overlap-tile strategy of U-Net allows "the seamless segmentation of arbitrarily large images" by extrapolating missing context through input images (Ronneberger et al. 3). U-Net is an appropriate approach to run the *Victorian400* dataset as U-Net deploys massive data augmentation in order to train a model with smaller datasets.

The relationship between the generator and the discriminator can be compared to the one between counterfeiters and the police: counterfeiters (the generator) produce “fake currency and use it without detection,” and the police “detect the counterfeit currency” (Goodfellow et al. 1). Within GAN models, the generator and discriminator compete with and learn from each other to bring about enhanced results. The discriminator role is to classify whether outputs from the generator are real, based on target images, and to generate a probability for outputs being categorized successfully through image-to-image translation tasks. The pix2pix model deploys a PatchGAN, which is “a discriminator architecture” that “only penalizes structure at the scale of patches” instead of through a deep convolutional neural network, which is commonly used for traditional GAN models. A smaller PatchGAN with fewer parameters proved to run faster when applied to “arbitrarily large images” (Isola et al. 4); the 70 x 70 PatchGAN was found to be most effective for image-to-image translation tasks.

In terms of validation, data scientists usually check outputs and add or delete data while running deep learning models with the datasets. There are no standards for validating datasets. Similarly, there is no agreement on which evaluation measures are be the best for GANs, although there have been studies that propose measuring GAN-derived models based on quantitative and qualitative methods (Lucic et al.). When evaluating GAN models, qualitative measures might “favor models that concentrate on limited sections of the data” such as overfitting, memorizing or low diversity, whereas quantitative measures “may not directly correspond to how humans perceive and judge generated images” despite being less subjective

(Borji 2). To validate the *Victorian400* dataset, I draw upon both quantitative (Figures 1.3, 1.4, and 1.5) and qualitative (Figures 1.6, 1.7, and 1.9) measures.

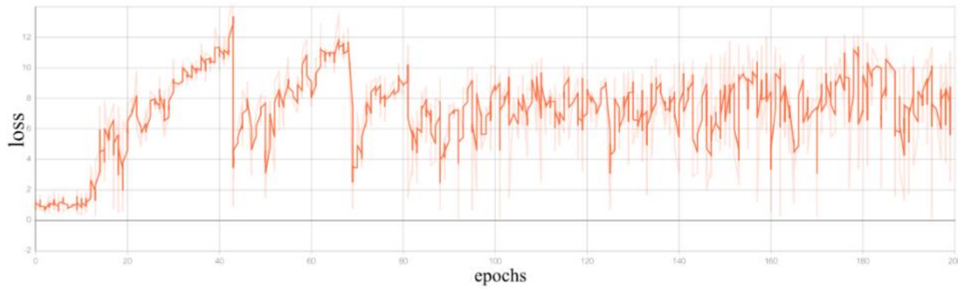


Figure 1.3: The generator loss of pix2pix with the *Victorian400* dataset.

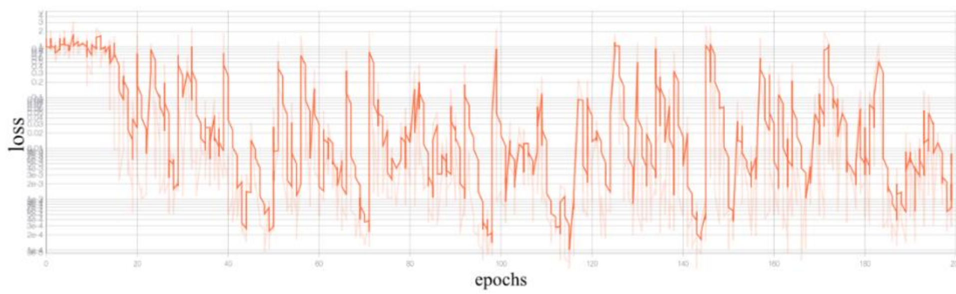


Figure 1.4: The discriminator loss of pix2pix with the *Victorian400* dataset.

I tested the pix2pix model with the *Victorian400* dataset to verify the possibilities of colorizing black-and-white illustrations from the nineteenth century. The *Victorian400* dataset consists of 400 images, and each image has 196,608 features (256 x 256 x 3 dimensions) as each image contains 256 x 256 pixels with RGB (Red, Green, Blue) values from 0 to 255. In Figures 1.3 and 1.4, both the generator loss and the discriminator become more stable as the number of epochs increased. The generator and the discriminator are trained with datasets aimed at having lower losses as training iterations progress. Although the discriminator loss seems inconsistent in Figure 1.4, the loss changes between 1 and 1e-4, which is a very small range. The generator loss at epochs=200 is 7.439, which is high compared to the discriminator loss at epochs=200, which scores 1.5325e-3. From epochs≈15, the discriminator loss drops drastically, while the generator loss increases, which signifies that the discriminator is beating the generator. The characteristics of GANs are such that the generator and the discriminator improve the accuracy of outputs by competing with each other. However, the pix2pix model with the *Victorian400* dataset reveals that the discriminator generally wins over the generator. GANs are based on a zero-sum game: the discriminator is stronger than the generator in this training model with the *Victorian400* dataset. In other words, the generator finds it difficult to fool the discriminator, which is the limit of GAN models: if the generator wins over the discriminator, or vice versa, it will be difficult to properly train a model with datasets due to overfitting from the unbalance between the generator and the discriminator.[15] Although the pix2pix model was not created with the aim of predicting illustration colors, Figures 1.5 and 1.6 reveal that the pix2pix model generates reasonable outputs when trained with the *Victorian400* dataset.

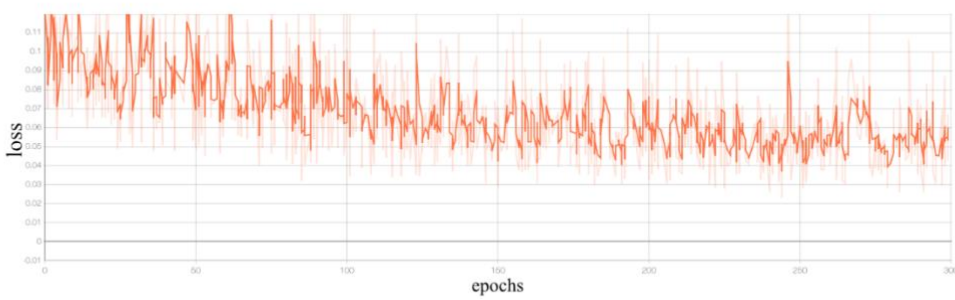


Figure 1.5: The L1 loss of pix2pix with the *Victorian400* dataset.

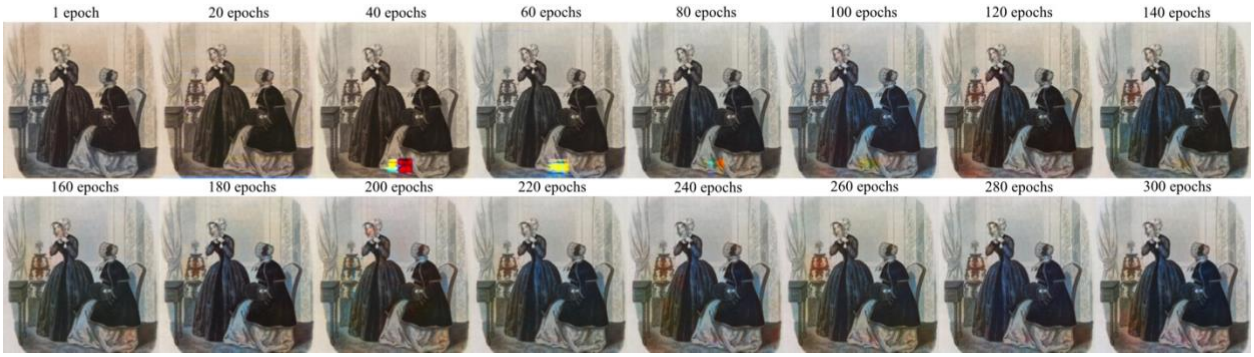


Figure 1.6: Test results of pix2pix with the *Victorian400* dataset based on every 20 epochs.

The L1 loss is the mean absolute pixel difference between generated outputs and target images in the training process. In Figure 1.5, the L1 loss gradually diminishes, which signifies that the model has been trained properly with the *Victorian400* dataset. From epochs≈160, the L1 loss no longer decreases but the L1 loss trajectory fluctuates around 0.05 between epochs≈160 and epochs≈240. Based on Figure 1.5, I concluded that the epochs between 160 and 240 would generate the best outputs. In Figure 1.6, a generated output was recorded for every 20 epochs. There are pixel scratches on the bottom of the output until 140 epochs, but from 160 epochs, the pix2pix model generates reasonable results without many improvements. Similarly, there are no remarkable improvements to the L1 loss in Figure 1.5 from around 160 epochs. Based on both the quantitative and qualitative methods (Figures 1.5 and 1.6), I decided to train the pix2pix model with the *Victorian400* dataset for 200 epochs to predict colors for black-and-white illustrations from the Victorian era.

Figure 1.7: The outputs of the pix2pix model with the *Victorian400* dataset.

Figure 1.8: Two images from the *Victorian400* dataset.

The *Victorian400* dataset was run with a test dataset that consisted of nine images. The L1 loss is around 0.044 at 200 epochs as a result of training the pix2pix model with the *Victorian400* dataset. Figure 1.7 shows the results of the test set after training with the *Victorian400* dataset. Despite the fact that the test results from Test1 to Test9 are not exactly the same as the original targets, the results provide natural-looking colored images based on the trained set. Looking into Test1, Test5, Test6, and Test9 from the column of Figure 1.7, however, the objects have a red tint. As mentioned above, this is the result of the fact that red is often found in Victorian illustrations. For example, Figure 1.8 shows Victorian women wearing colorful, reddish clothes. While testing the pix2pix model with the *Victorian400* dataset, I found the illustrations were colored with red, which was a case of overfitting due to the dataset being biased towards the color red. I had to remove some images which had a preponderance of red in them in order to balance the color palette. At the same time, I left a certain number of images using red so as not to disregard the color preferences of the Victorian age, as I deemed this important to reflect contemporaneous color trends. As the dataset significantly influences the test results of deep learning models, these Victorian color trends were still revealed through Test1 and Test9. In addition, limitations remain due to the small size of the *Victorian400* dataset.

After verifying the *Victorian400* dataset with the test dataset, I used the trained model with the *Victorian400* dataset in order to predict the colors of black-and-white illustrations from Charles Dickens's *Bleak House* (serialized 1852–1853). The illustrations used for the test are from the Bradbury & Evans book edition (1853) of Dickens's *Bleak House*, for which the plates were created by Hablot Knight Browne (Phiz). Browne illustrated ten of Dickens's books including *Dombey and Son* (serialized 1846–1848), *David Copperfield* (serialized 1849–1850), and *Bleak House*. *Bleak House*, which includes 40 black-and-white illustrations with the title-page etched on steel, is a good subject for testing colorization due to the fact that there is a good variety of illustrated depictions of characters and settings. In order to evaluate the outputs generated by the pix2pix model with the *Victorian400* dataset, I divided them into three categories: not good, fair, and good. As for validating GAN models, "visual examination of samples by human raters is one of the common and most intuitive ways" (Borji 30).

Figure 1.9: Test results for black-and-white illustrations from of Charles Dickens's *Bleak House*.

In Figure 1.9, there are 12 illustrations generated by pix2pix trained with the *Victorian400* dataset as examples. Figure 1.9 can be divided into three categories based on the quality of the output:

1. Not Good: Test1, Test2
2. Fair: Test3, Test4, Test8
3. Good: Test5, Test6, Test7, Test9, Test10, Test11, Test12

Test1 through Test6 are focused on characters, while Test7 through Test12 are focused on settings. Test1 reveals that the background and people in the background were not colored. There are a number of blurry faces in the background, and walls were not sketched with enough detail for the deep learning model to detect them. Similarly, Test2 was not colored thoroughly, with only two characters on the left painted in yellow and orange. In Test3 and Test4, characters were colored based on the trained set, but the backgrounds were not colored enough to show improvements on the black-and-white illustrations. Test3, Test4 and Test8 were categorized as fair since there were reasonable improvements in assigning the colors between the inputs and outputs. Test5 and Test6 were sufficiently colored in terms of both the characters and the background, although the colors were not always appropriate. Test7 to Test12 are colored based on black-and-white background illustrations, which Browne created by using the dark plate technique[16] in order to provide deep contrast for a higher quality of depth. The illustrations from Test7 to Test12 created with the dark plate technique turned out to be compatible with deep learning colorization by generating reasonable outputs due to the distinct contrast between darkness and highlights.

As proved by this experimentation, there is no single answer to the question of how to colorize black-and-white illustrations. Colorizing black-and-white illustrations is a double-edged sword since, while it may distort the original illustration, it nonetheless provides the opportunity for modern readers to expand their imagination and increase reading pleasure. To increase the accuracy of colorizing backgrounds, I would need a vast amount of colored background images.

In addition, it would require advanced deep learning layers dealing with a variety of different images. Without the support of grants, it would be impossible for an individual project like this to create a large dataset and to develop more complicated and enhanced deep learning models. The typical cGAN training requires a large training set in order to score reasonable results. As an individual project, the *Victorian400* dataset is a compromise between time and cost. It is true that the *Victorian400* dataset is small compared to other datasets for GAN training. Nonetheless, it produces reasonable enough results from the test sets to attest to its quality.

4. Conclusion

The *Victorian400* dataset was created for data scientists and digital humanists who create, train and test colorization deep learning models. The *Victorian400* dataset has been publicly shared on websites, namely Kaggle and GitHub, for data scientists and digital humanists, and it has been downloaded and experimented with using a variety of deep learning models.[17] Undeniably, due to its size, there is still scope for improving the *Victorian400* dataset. Although some scholars might think the *Victorian400* dataset is small, it proved sufficient to produce reasonable results. In addition, small, and high-quality datasets make it possible to run tests easily without spending too much time and money.[18] Data scientists spend a great portion of time curating datasets in order to make them valid and credible, in addition to spending funds to run deep learning models in cloud services such as Amazon EC2, Microsoft Azure, and Google Colab. The *Victorian400* dataset will not only save a tremendous amount of time for digital humanists who experiment with Victorian illustrations, but will also contribute to the development of deep learning-based research in the digital humanities. Lastly, the *Victorian400* dataset will assist digital humanists in the creation of datasets for machine learning/data scientists; it is difficult to find credible humanities datasets on the Internet since most humanities datasets shared on Kaggle and GitHub are created by non-humanists.[19] It is the role of digital humanists not only to create and curate humanities datasets, but also to provide EDA results which show that humanities datasets created by digital humanists are credible enough to be used for deep learning tasks. This will enrich the scholarship in both digital humanities and data science as well as lead to the development of collaboration between digital humanists and data scientists.

Notes

[1] Christof Schöch has argued that data in the humanities are unique artifacts. Trevor Muñoz has articulated relationships between publishing and data curation. Miriam Posner has claimed that we should have high standards when creating "data-based work that depicts people's lives." Katie Rawson and Trevor Muñoz revealed how to properly clean data for humanities researchers.

[2] Data scientists usually perform EDA (Exploratory Data Analysis) to find how datasets can be used for deep learning and to see whether they are credible. One way to perform EDA is to use t-SNE (t-distributed Stochastic Neighbor Embedding) with random sampling to see clustering shapes and the distance between features. In general, dataset creators provide test results with specific information such as parameters, evaluation metrics and network architectures, in order to show credibility of the datasets, as shown in this article.

[3] To analyze humanities datasets is both to provide EDA and to further expand data analysis into digital projects and articles.

[4] Eight of Charles Dickens's photographs were colorized by Oliver Clyde and released by the Charles Dickens Museum in London ahead of the 150th anniversary of Dickens's death.

[5] There were hand-colored additions of luxurious books until "the last decades of the eighteenth century," but the emergence of printing methods such as chromolithography began to bring colored illustrations to the public from the 1830s. Color printing was essential for "children's books, religious works, and gift books." Still, color printing was costly during the Victorian era (Allington et al. 293).

[6] The *Victorian400* dataset was used to introduce deep learning and Python tutorials on Pseudo Lab (https://pseudo-lab.github.io/Tutorial-Book-en/index.html) and Kaggle (https://www.kaggle.com/jiny333/tutorial-on-using-subplots-in-matplotlib), respectively.

[7] Several data scientists have contacted me through Kaggle to express their appreciation that I shared the dataset with the public.

[8] For two semesters, I used both hand-colored and machine-colored nineteenth-century illustrations in my English classes to see how different coloring influences readers and the possible usage of machine-colorized illustrations in facilitating students' learning and imagination in literature. Most of the students responded to the machine-colorized illustrations with surprise, curiosity and excitement. One made the observation that machine-colored illustrations helped depict objects in a way that is difficult to catch in hand-colored illustrations. Another noted that the hand-colored illustrations seemed too modern, whereas the machine-colored illustrations seemed more realistically Victorian.

[9] This article does not deal with the pedagogical side of deep learning colorization but focuses on the validation of the *Victorian400* dataset.

[10] https://www.kaggle.com/elibooklover/victorian400

[11] https://github.com/elibooklover/Victorian400

[12] See M. Hardie, *English coloured books* (London, 1906) which introduces English color illustrations by a variety of artists including W. Savage, T. Rowlandson, T. S. Boys, David Roberts, and Kate Greenaway. See P. Allingham, "The technologies of nineteenth-century illustration: woodblock engraving, steel engraving, and other processes" at *The Victorian web* (http://www.victorianweb.org/art/illustration/tech1.html), which examines illustration printing techniques in the Victorian era. See V. Finlay, *Color: a natural history of the palette* (New York, 2004), which explores the histories of colors.

[13] There are a number of Victorian databases/archives/platforms with Victorian illustrations: *Database of mid-Victorian illustration* by Julia Thomas et al. (http://www.dmvi.cardiff.ac.uk),

The illustration archive by Julia Thomas et al. (https://illustrationarchive.cardiff.ac.uk/), *The Victorian web* by George Landow (http://www.victorianweb.org/), *The Rossetti archive* by Jerome McGann et al. (http://www.rossettiarchive.org/), *The Charles Dickens page* by David Perdue (https://www.charlesdickenspage.com/), *The George Eliot archive* by Beverley Rilett (http://www.georgeeliotarchive.org/), and *Collaborative organization for virtual education* by Dino Felluga et al. (https://editions.covecollective.org/). These sites are historically and pedagogically helpful with open access for researchers, instructors and students.

[14] Since this article is not about the pix2pix model, which is a deep learning model for image-to-image translation, I focused on the process and results of training the *Victorian400* dataset with the pix2pix model. For more details on the pix2pix model, see P. Isola et al., "image-to-image translation with conditional adversarial networks."

[15] There are several issues with GANs such as non-convergence, mode collapse, and unstable gradients. This article does not deal with the limits of GANs. For discussions of the limits of GANs, see Lucic et al., "Are GANs created equal?"; M. Arjovsky and L. Bottou, "Towards principled methods for training generative adversarial networks"; and K. Li and J. Malik, "Implicit maximum likelihood estimation."

[16] Browne deployed the dark plate technique, which is a combination of engraving and etching, in order to convey thematic schemes in Dickens's works such as crime and hopelessness. This article does not delve into the interpretations of the illustrations in Dickens's novels but focuses on the validity of the *Victorian400* dataset.

[17] The DH community has been participating in an open data movement. Amy Earhart and Toniesha Taylor assert that allowing "the public to view the type of work that we accomplish is powerful, particularly within the current environment of distrust of the academy" (259). From

the perspective of a dataset creator, sharing work with the public yields feedback from a wider range of viewers that makes it possible to improve my work. The *Victorian400* dataset is currently ranked with a bronze medal on Kaggle.

[18] It took me around 13 hours to train the *Victorian400* set once with 200 epochs. There were difficulties when testing with the *Victorian400* dataset due to a poor research environment. In order to trim the data and avoid overfitting, I had to train with a variety of new datasets, which took a tremendous amount of time. Later, a project grant was given to this project by CoDHR (Center of Digital Humanities Research), so I was able to test the dataset in a better environment.

[19] When choosing datasets, some users consider who has created the datasets and how popular the datasets are, whereas others draw on EDA for the datasets, as these provide objective criteria, although it can be difficult to determine the credibility of some datasets until they are used for specific tasks. It is thus important for digital humanists who create and share datasets to provide EDA and thus enhance the credibility of their shared datasets.

Bibliography

Allingham, Philip V. "The Technologies of Nineteenth-Century Illustration: Woodblock Engraving, Steel Engraving, and Other Processes." *The Victorian Web: Literature, History and Culture in the Age of Victoria*, 2001. victorianweb.org/art/illustration/tech1.html.

Allington, Daniel, et al. *The Book in Britain: A Historical Introduction*. Wiley-Blackwell, 2019.

Amelia Acker and Tanya Clement, "Introduction: Data Cultures, Culture as Data - Special Issue of Cultural Analytics." *Journal of Cultural Analytics*. April 10, 2019.

Antic, Jason. *DeOldify*, 2019. *GitHub repository*, github.com/jantic/DeOldify.

Arjovsky, Martin, and Léon Bottou. "Towards Principled Methods for Training Generative Adversarial Networks." *ICLR*, 2017. arXiv:1701.04862.

Arora, Sanjeev, and Yi Zhang. "Do GANs Actually Learn the Distribution? An Empirical Study." 2017, arXiv:1706.08224.

Borji, Ali. "Pros and Cons of GAN Evaluation Measures." *Computer Vision and Image Understanding* 179, 2019, pp. 41-65. doi.org/10.1016/j.cviu.2018.10.009.

Ciresan, Dan, et al. "Deep Neural Networks Segment Neuronal Membranes in Electron Microscopy Images." *Advances in Neural Information Processing Systems*, 2012. papers.nips.cc/paper/4741-deep-neural-networks-segment-neuronal-membranes-in-electron-microscopy-images.pdf.

Earhart, Amy, and Toniesha Taylor. "Pedagogies of Race: Digital Humanities in the Age of Ferguson." *Debates in the Digital Humanities 2016,* 2016, pp. 251–264.

Finlay, Victoria. *Color: A Natural History of the Palette*. Random House, 2007.

Goodfellow, Ian, et al. "Generative Adversarial Nets." *Advances in Neural Information Processing Systems*, 2014. papers.nips.cc/paper/5423-generative-adversarial-nets.pdf.

Géron, Aurélien. *Hands-on Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media Inc, 2019.

Hardie, Martin. *English Coloured Books*. London: Methuen, 1906.

Isola, Phillip, et al. "Image-to-Image Translation with Conditional Adversarial Networks." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017. openaccess.thecvf.com/content_cvpr_2017/papers/Isola_Image-To-Image_Translation_With_CVPR_2017_paper.pdf.

Ledig, Christian, et al. "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network." *Conference on Computer Vision and Pattern Recognition*, 2017. openaccess.thecvf.com/content_cvpr_2017/papers/Ledig_Photo-Realistic_Single_Image_CVPR_2017_paper.pdf.

Lucic, Mario, et al. "Are GANs Created Equal? a Large-Scale Study." *NIPS*, 2018. papers.nips.cc/paper/7350-are-gans-created-equal-a-large-scale-study.pdf.

Muñoz, Trevor. "Data Curation as Publishing for the Digital Humanities." *Journal of Digital Humanities*, vol. 2, no. 3, 2013. journalofdigitalhumanities.org/2-3/data-curation-as-publishing-for-the-digital-humanities/.

Posner, Miriam. "What's Next: The Radical Unrealised Potential of Digital Humanities." *Debates in the Digital Humanities 2016*, edited by Matthew K. Gold and Lauren F. Klein, 2016, pp. 32–41.

Rawson, Katie, and Trevor Muñoz. "Against Cleaning." *Debates in the Digital Humanities 2019*, edited by Matthew K. Gold and Lauren F. Klein, 2019, pp. 279–292.

Reed, Scott, et al. “Generative Adversarial Text to Image Synthesis.” *International Conference on Machine Learning*, 2016. arXiv:1605.05396.

Ronneberger, Olaf, et al. “U-net: Convolutional Networks for Biomedical Image Segmentation.” *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015. arXiv:1505.04597.

Schöch, Christof. “Big? Smart? Clean? Messy? Data in the Humanities.” *Journal of Digital Humanities* vol 2, no. 3, 2013. journalofdigitalhumanities.org/2-3/big-smart-clean-messy-data-in-the-humanities/.

Shmelkov, Konstantin, et al. “How Good is My GAN?” *Proceedings of the European Conference on Computer Vision*, 2018. openaccess.thecvf.com/content_ECCV_2018/papers/Konstantin_Shmelkov_How_good_is_ECCV_2018_paper.pdf.

Stonebraker, Michael, et al. “Data Curation at Scale: The Data Tamer System.” *Conference on Innovative Data Systems Research*, vol. 1, 2013. cs.uwaterloo.ca/~ilyas/papers/StonebrakerCIDR2013.pdf.

Yonetsuji, Taizan. *Paints Chainer*, 2017. *GitHub repository*, github.com/pfnet/PaintsChainer.

Xu, Qiantong, et al. “An Empirical Study on Evaluation Metrics of Generative Adversarial Networks.” 2018. arXiv:1806.07755.

CHAPTER II

Case Study: Using Machine-Colored Illustrations of Charles Dickens's Fiction in the Classroom

1. Introduction

In 2021, eight of Charles Dickens's portraits, colorized by Oliver Clyde, were released to the public by the Charles Dickens Museum in London ahead of the 150th anniversary of Dickens's death. The attempt by the Charles Dickens Museum to hand-colorize Dickens's black-and-white portraits reveals that classical culture, when combined with the perspective of current technology such as colorization, can attract modern individuals by renewing enjoyment of and engagement with the material. As an extension of my colorization project, I decided to see the possible pedagogical usage of machine-colored Victorian illustrations through a case study where I examined if machine-colored illustrations could increase students' engagement with illustrated Victorian texts. Providing original illustrations that have been colorized can help readers picture characters' personalities and grasp events in the plot with more vitality. However, the British Library's commentary[1] on the effect of colored illustrations in the 1911 edition of Oliver Twist reveals that there are cons to colorizing illustrations from Victorian fiction. For instance, they mention that while colored illustrations gain "vitality," they "lose something of the gravity" present in black and white illustrations. Through my case study, I introduce students' responses to black-and-white, hand-colored, and machine-colored illustrations to examine the efficacy of using machine-colored Victorian illustrations in the literature classroom. Firstly, I colorize black-and-white Victorian illustrations using a deep learning model and the *Victorian400* dataset,[2] then ask students several questions while showing black-and-white, hand-colored, and machine-colored illustrations with texts. For my pedagogical case study, I use illustrations from Charles Dickens's *Oliver Twist* (serialized between 1837–8) and *A Christmas Carol* (1843), since both well-known works have frequently been reprinted with illustrations and adapted for television and film. In this paper, I briefly discuss the rise of Victorian illustrated texts, focusing on Dickens's role and influence, then introduce my research methods for the colorization of Dickens's illustrations, before lastly delving into my pedagogical case study of using original and colorized illustrations in the literature classroom. The case study in this article reveals that although over half of my students preferred hand-colored illustrations, both hand-colored and deep learning-based colorized illustrations helped them to apprehend the illustrations' details and settings and provided more entertainment during the reading experience. Furthermore, students who most preferred machine-colored illustrations were interested in the fact that machine-colorized illustrations reflect a contemporary color palette based on datasets. The case study shows the possibility of using machine-colored illustrations from Victorian fiction in the literature classroom.

2. Charles Dickens's Illustrations

As Vincent van Gogh claimed, "There is no writer, in my opinion, who is so much a painter and black-and-white artist as Dickens." Charles Dickens, one of the most prominent Victorian authors, guided, published, and commercialized illustrations in his fiction. Dickens was "the great popular entertainer" (Levine 1), actively using illustrations in his fiction due to his belief in the power of images. Illustrations help readers of varied educational backgrounds understand texts within context, in addition to being both enlightening and entertaining. Dickens actively deployed illustrations in his fiction as a device of visual storytelling. His novels are no longer “properly appreciated” without the illustrations which are essential to understanding the text (Solberg 129). His illustrations, which provided information about his texts and aided readers’ imaginations, resulted in the success of his novels and thus increased the number of his readers. Illustrations also enabled children to access Dickens’s stories and helped readers to recall important events in the stories. The accessibility to his stories created by the illustrations led his novels, such as *A Christmas Carol* (1843), to great success among a broad spectrum of readers. Illustrations in the original prints of Dickens’s fiction are mainly black-and-white, except for four colorful illustrations in *A Christmas Carol*. The inclusion of hand-colored illustrations in *A Christmas Carol* shows that Dickens had a desire to provide colored illustrations to readers. Black-and-white illustrations in his fiction, along with portraits of Dickens, have been hand-colored by illustrators. Twenty-four original illustrations by George Cruikshank in *Oliver Twist* were colored in the 1911 edition.[3] The illustrations of *Oliver Twist* were newly produced in color by Frederick Pailthorpe in the 1885 edition.[4]

James Reitter delved into the three stages of Dickens’s illustrations, in addition to examining the influence of William Hogarth, an eighteenth-century English illustrator, on them. Reitter notes that Dickens’s illustrations in his earlier fiction, especially in *Oliver Twist*, provided a “caricatured and satirical counterpart to the text” due to William Hogarth’s influence, whereas in his later work, *Our Mutual Friend*, the illustrations were more beautiful and realistic. Illustrations by George Cruikshank in *Oliver Twist* “provided the reader with insight into characters and enhanced the narrative,” whereas Marcus Stone’s work in *Our Mutual Friend* was “simply a reflection of the text” (33). Although both men were deeply influenced by Hogarth, Dickens fought with George Cruikshank over the ways in which his characters in *Oliver Twist* were depicted, not wanting them to be portrayed in caricature.[5] Therefore, Cruikshank blended "his talent for caricature with the realism that Dickens desired" when illustrating *Oliver Twist* (Reitter 41).[6]

As Cohen notes, *Pickwick*, which was "in an illustrated serial format […] revolutionized the publication of new fiction between 1836 and 1870" (4). Dickens's bold attempt to publish illustrations in his serialized fiction brought a change to the publication market in the Victorian age as books or serials including illustrations were found to boost sales. It is also evident that Dickens's illustrations contributed to a rise in the number of fiction readers during the Victorian era, which has lasted into modern times. Charles Dickens's illustrations have been continuously discussed by modern Dickensian scholars.[7] Stein claims illustrations are "description redescribed from the perspective of a second medium and a second imagination" (171). In addition, illustrations play the role of providing images that align with the text, in order for readers to visualize events and characters (175). Catherine Golden points out the importance of illustrations in *Oliver Twist* by noting that *Oliver Twist* would be a different book "from the one Dickens intended and the Victorian readers experienced" without "the twenty-four plates George Cruikshank designed for the novel" (117). Whether the illustrations have been black-and-white or colored, they have played a role in helping contemporary readers understand characters, scenes, and plots through visualization.

Dickens worked with 16 different illustrators, always carefully considering the style of his illustrations. As a successful writer, Dickens did not want illustrators to taint or change the plot of his fiction through their illustrations. Dickens possessed unprecedented power in controlling his illustrators, which was not normal at the time. He always guided his illustrators to create illustrations that matched exactly what he had planned. Dickens was knowledgeable of arts and had his own point of view. In his fiction, the style of illustrations varies depending on the illustrator, due to their different techniques and styles. Illustrators such as George Cruikshank, Hablot Knight Browne (Phiz), and John Leech played significant roles in depicting many of the characters, specific settings, and complicated plots in Dickens's fiction. Still, Dickens was strict about his illustrations. He supervised his illustrators closely with consideration towards details and visual effects. He involved himself by thoroughly inspecting and providing feedback on the illustrations, from the "preliminary sketches" to the "final drawings" (Cohen 5). Dickens provided specific guidelines for each illustration, such as "the number of the characters as well as their position, gestures, expressions, dress, and settings," in addition to what color should be used, although they were usually printed to be black-and-white (Cohen 5). Dickens also considered exactly where illustrations should be placed, believing that the illustrations should be "integrated into the text instead of appearing on separate pages" (qtd. in Hughes 94). Michael Steig infers that "Dickens chose the subjects, specified the lettering for each plate, and gave directions with varying degrees of thoroughness regarding composition, descriptive details and characterization" (124). His illustrators often found it difficult to satisfy him due to time constraints and technical issues. Dickens sometimes delivered guidelines to his illustrators at the last minute with "marked-up proofs, hastily written précis, or brief verbal instructions" (Cohen 8), which resulted in substituting circulated illustrations. His requirements were "exacting even beyond what is ordinary between author and illustrator" (Kitton vii-viii). With strict control over his illustrations, Dickens was able to describe characters, settings, and plots from his own perspective through his illustrations.

3. Research Methods

To create the machine-colored illustrations for *Oliver Twist* and *A Christmas Carol*, I first curated the black-and-white illustrations to fit into image processing. I trained the pix2pix model (Isola et al.) with the *Victorian400* dataset (Kim). The pix2pix model performs automatic graphic operations on images based on cGANs (conditional Generative Adversarial Networks). The *Victorian400* dataset, which consists of 400 images from the Victorian era, is a deep learning dataset for colorizing black-and-white illustrations from the Victorian age.

The pix2pix model, which was released by Isola et al. in 2017, is a cGAN model based on the GAN (Generative Adversarial Network) model, which was first introduced by Goodfellow et al. in 2014. In the GAN model, adversarial nets, which consist of a generator and a discriminator, create images by learning updated vectors through the differences between fake and real images. Goodfellow et al. explained the relationship between the generator and the discriminator in the GAN model, by using the analogy of the relationship between a thief, who produces fake currency as the generator, and the police, who detect counterfeit currency as the discriminator (1). Although the architecture of the pix2pix model is similar to the GAN model, the pix2pix model contains a 'U-Net,' which skip connections are added to, instead of using an encoder-decoder for the generator. With U-Net, the pix2pix model can colorize images efficiently with local and contextual information. In addition, unlike previous conditional generative adversarial networks models, the pix2pix model deploys a convolutional PatchGAN classifier for the discriminator (Isola et al. 2).

Deep learning models need to be trained with curated and customized datasets for specific tasks. The *Victorian400* dataset, which consists of 400 illustrated images from the nineteenth century, was created for the specific task of colorizing black-and-white illustrations from the Victorian era. The *Victorian400* dataset is publicly shared for humanists as an open data source on Kaggle and GitHub. The *Victorian400* dataset was tested with the pix2pix model and produced reasonable enough results, although it still has room for improvement. Images in the *Victorian400* dataset are already curated to fit a standard size of 256 x 256 for the process of deep learning.

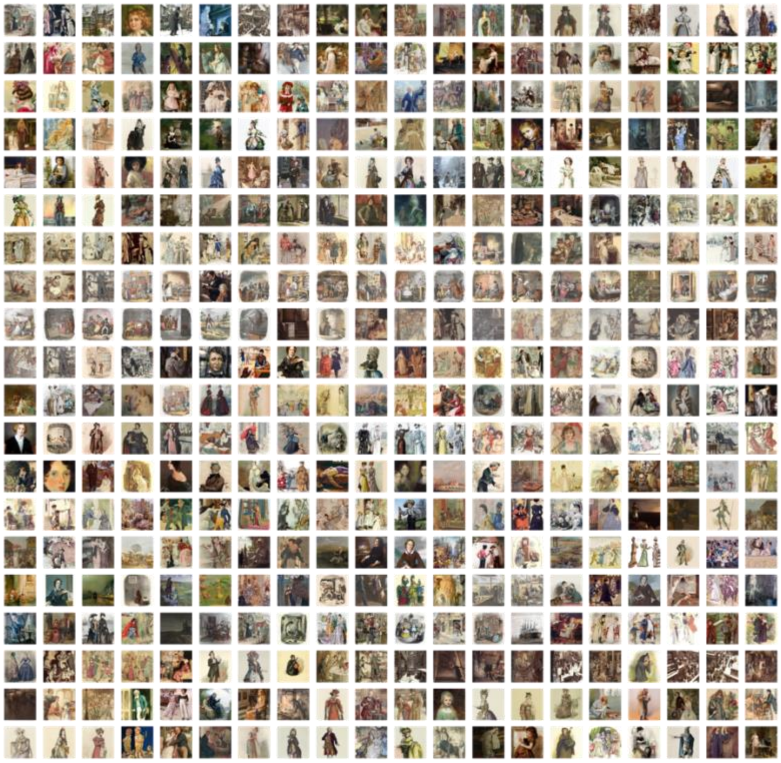


Figure 2.1: The *Victorian400* dataset.

By training the pix2pix model with the *Victorian400* dataset, I created machine-colored illustrations from Dickens's *Oliver Twist* and *A Christmas Carol* to use in literary classrooms. I trained the pix2pix model with the *Victorian400* dataset for 200 epochs to adequately colorize black-and-white illustrations from *Oliver Twist* and *A Christmas Carol*. Figure 2.2, from top to bottom, shows black-and-white, machine-colored, and hand-colored illustrations from *A Christmas Carol*. The hand-colored illustrations are from the original edition, illustrated by John

Leech in 1843. In *A Christmas Carol*, there are four black-and-white and four hand-colored illustrations in total.

Figure 2.2: Black-and-white (top), machine-colored (middle), and hand-colored (bottom) illustrations from Charles Dickens's *Christmas Carol*.

In Figure 2.3, there are four black-and-white, machine-colored, and hand-colored illustrations, each from *Oliver Twist*. In total, there are 24 black-and-white illustrations in the original edition of *Oliver Twist*, illustrated by George Cruikshank. I used the original, hand-colored illustrations, which were reproduced in around 1911 from the original edition and publicly shared by the British Library.

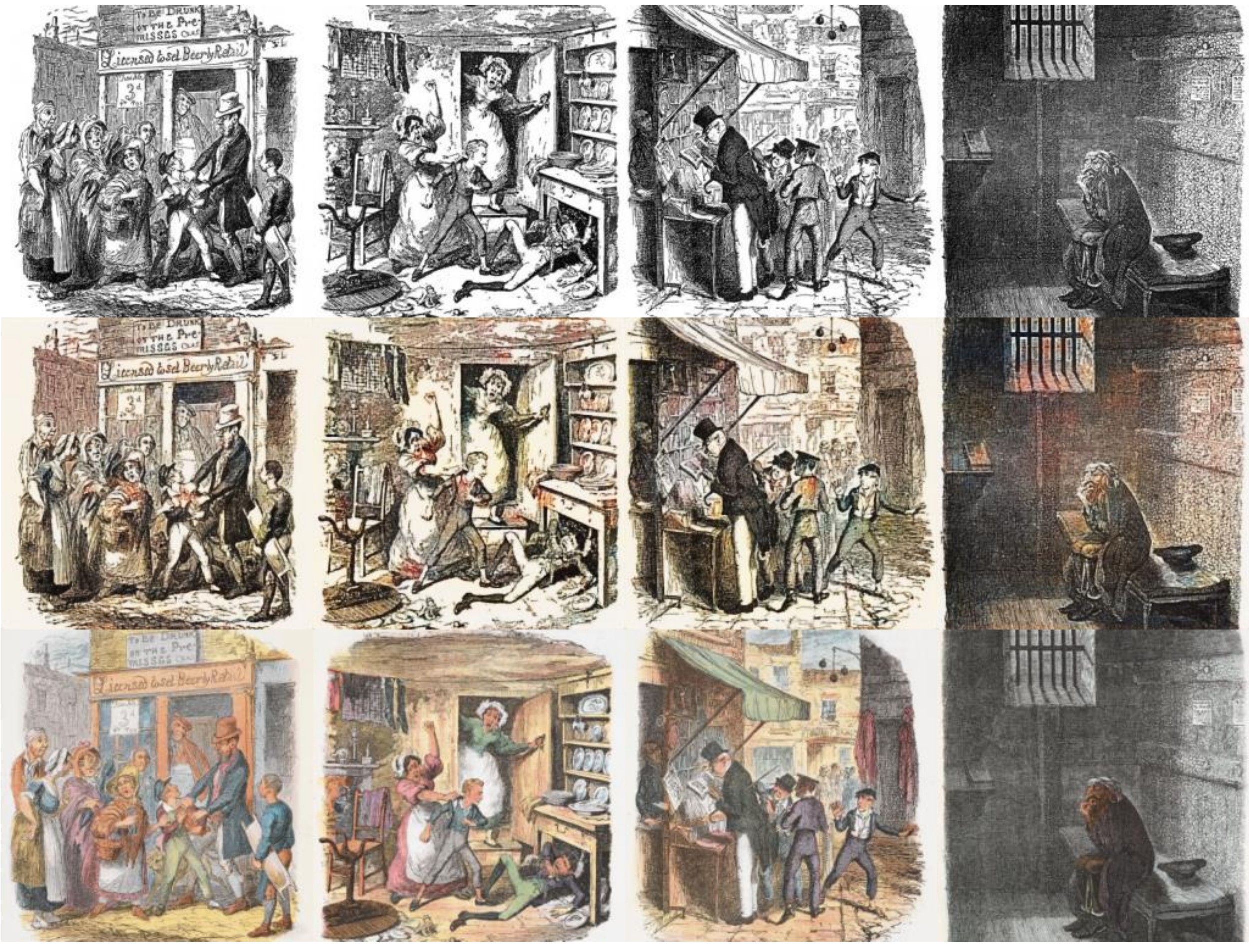
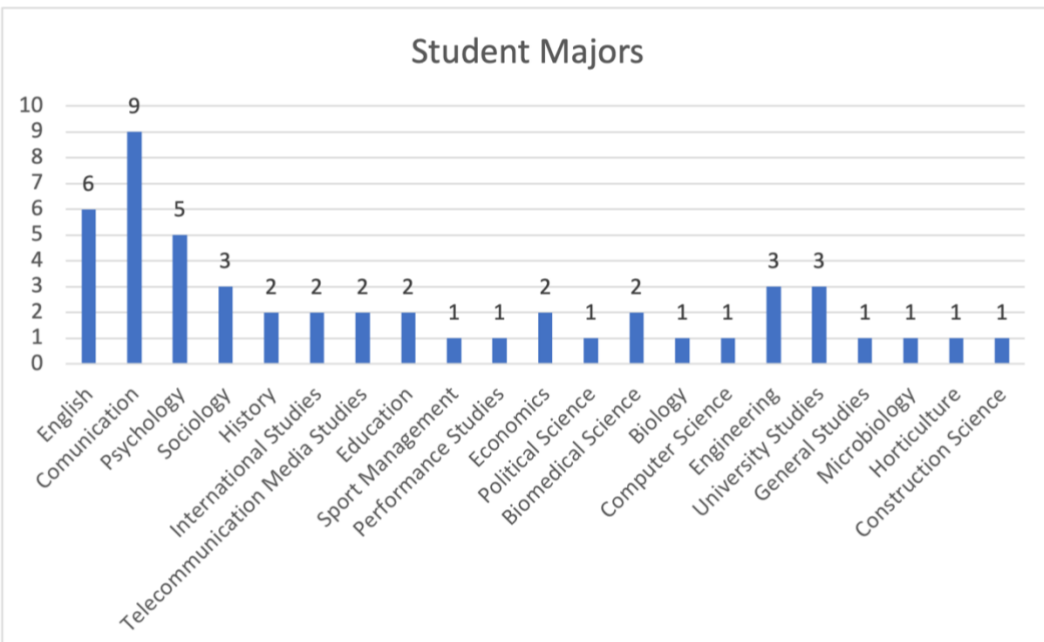


Figure 2.3: Black-and-white, machine-colored, and hand-colored illustrations from Charles Dickens’s *Oliver Twist*.

4. Case Study

For three semesters from Spring 2020 to Spring 2021, I have used black-and-white, hand-colored, and machine-colored illustrations from *Oliver Twist* and *A Christmas Carol* in three different literature classes. Throughout the semester I discussed the novels with illustrations, and at the end of each semester, I had an illustration session. I asked the following questions after showing black-and-white, hand-colored, and machine-colored illustrations from the novels: Q1) When you read Victorian novels, do you prefer those with illustrations or those without illustrations? Why? Q2) How do you feel about black-and-white illustrations? Q3) How do you feel about machine-colored illustrations? Q4) Which one do you prefer between black-and-white and machine-colored illustrations? Why? Q5) Which one do you prefer between machine colored and hand-colored illustrations? Why? Q6) What do you wish to see from the future of machine-colored illustrations? The total number of participants was 50 undergraduate students. Students' majors were various, from English to Mechanical Engineering. In Figure 2.4, English, communication, psychology, sociology, and history majors comprise 50% of the total responders.



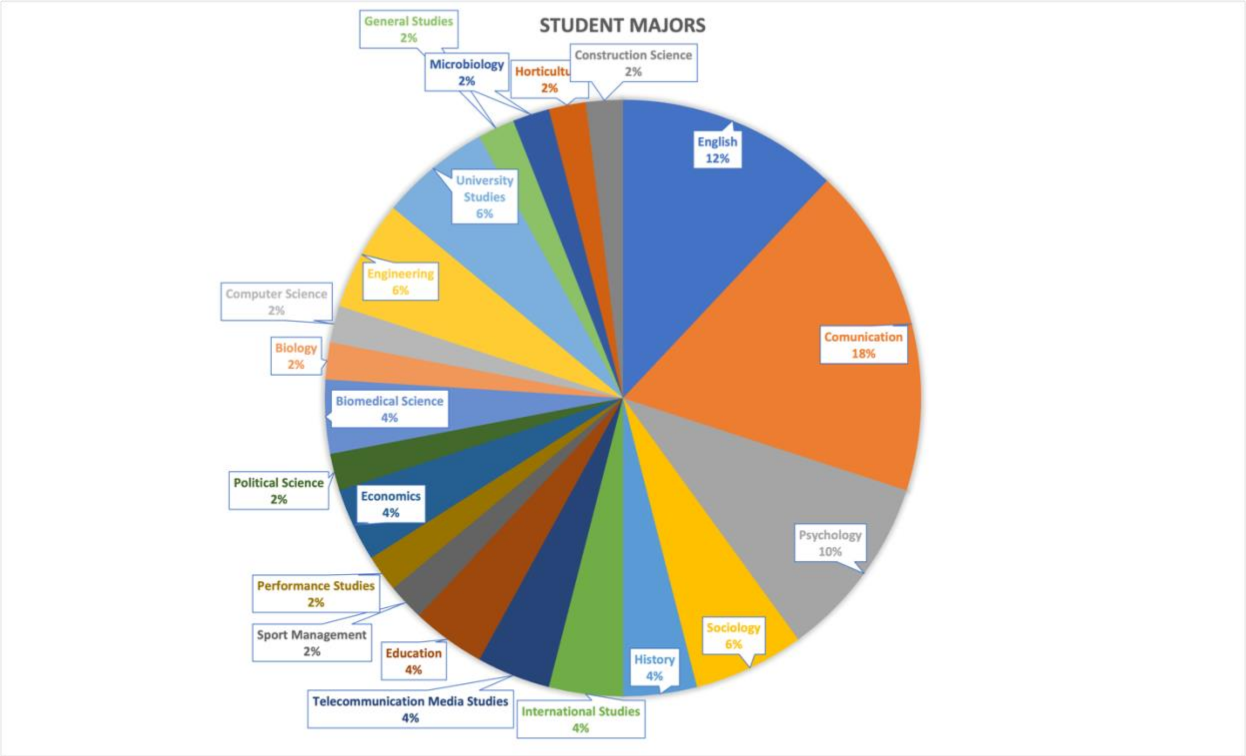


Figure 2.4: Survey of participants' majors.

As for the first question, "When you read Victorian novels, do you prefer those with illustrations or those without illustrations? Why?" 88% students positively responded to having illustrations when reading Victorian novels. Students who preferred illustrations narrated that illustrations are helpful for picturing characters, settings, and scenes, as well as better understanding the story, while students who preferred to read Victorian novels without illustrations answered that illustrations sometimes did not match up with the image they had created in their mind. Furthermore, illustrations prevented them from using their own imagination, as one student noted, "I prefer reading without illustrations because I usually have something more creative in my head."

In the second question, "How do you feel about black-and-white illustrations?" 74% students responded positively. They noted that black-and-white illustrations match the atmosphere of the Victorian era, since black-and-white illustrations align with readers' ideas of antiquity. Students who responded negatively mentioned that black-and-white illustrations are less clear and that details such as characters' emotions and appearances, in addition to settings, were more difficult to observe due to the lack of colors. This is likely because modern readers are used to colorful images, where details are easier to spot. The number of negative responses was similar to the first question, while 12% of students responded that they feel neutral towards the black-and-white illustrations.

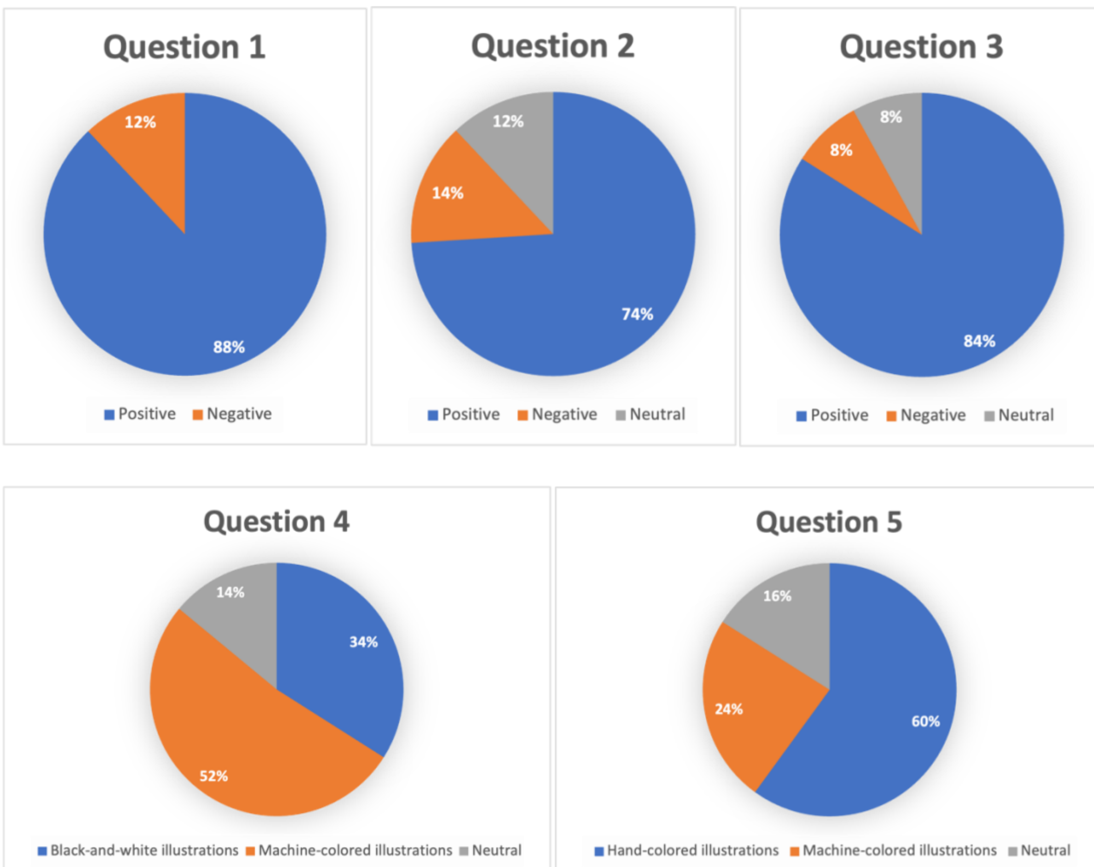


Figure 2.5: Survey results.

While the majority of students preferred to read Victorian novels with illustrations, I expected that there would be resistance against machine-colored illustrations among those surveyed. However, as the pie chart in Figure 2.5 shows, most students had positive impressions on machine-colored illustrations. One of the most impressive responses was that the responder preferred machine-colored illustrations since it is easier to understand and differentiate between details in colored illustrations. Furthermore, the student liked “the fact that these machine-colored illustrations color the illustrations based on datasets with colors used during those times.” Another student mentioned that machine-colored illustrations added more interesting perspectives to scenes and helped them to see the images more clearly. However, some students who responded negatively mentioned that machine-colored illustrations had much room for growth in recognizing fine details. Although students were aware that machine-colorization could be improved, overall, they were not resistant to machine-colored illustrations, and they perceived them as illustrations that represented Victorian colors.

When it comes to the comparison between black-and-white and machine-colored illustrations (Q4), students tended to prefer machine-colored illustrations over black-and-white illustrations. A student noted that black-and-white illustrations lacked detail in terms of color, whereas machine-colored illustrations showed what color of clothes were used during that time. Another student responded that overly detailed black-and-white can be hard to read sometimes, while they found the machine-colored illustrations more fun to look at. Another response was that the machine-colored illustrations brought more life to the scene portrayed in the illustrations. Responses from students who preferred black-and-white illustrations over machine-colored illustrations provided a different perspective on colorizing illustrations in Victorian fiction. For example, one student mentioned that black-and-white illustrations added to the sadness of the novel’s theme. Another student preferred how black-and-white illustrations let him pick the colors he imagined in his head. Some other students answered that black-and-white illustrations seemed to fit better with Victorian novels by noting that black-and-white illustrations seemed to be straight from the illustrator's thoughts while colored illustrations felt ingenuine.

While the students showed preferences towards machine-colored illustrations over black-and-white illustrations (Q4), responses to which was preferred between hand-colored and machine-colored illustrations (Q5) revealed that hand-colored illustrations were favored over machine-colored illustrations. Students who preferred hand-colored illustrations responded that the human touch made them more realistic, accurate, and authentic, even though they admitted machine-colored illustrations had arguably more consistent coloring. In addition, they perceived that hand-colored illustrations were more vibrant, but that they sometimes included the wrong colors. Most students who preferred machine-colored illustrations pointed out the accuracy of machine-coloring. For instance, one response read "machine-colored illustrations are more precise and accurate to the time compared to hand-colored illustrations due to the fact that artists choose the color in hand-colored illustrations and in machine-colored illustrations that machine chooses the color based on datasets from that time." Some expressed that the hand-colored illustrations looked fake or unrealistic; one student mentioned that hand-colored illustrations seemed to represent a different era, and that they felt too modern, causing distraction and confusion while reading. Therefore, wishing that the hand-colored illustrations had been less vibrant or modern, these students ended up answering that they preferred machine-colored illustrations. Still, twice as many students preferred hand-colored illustrations over machine-colored illustrations.

The last question asked for feedback on improving machine-colored illustrations. Most students hoped to see more colors in the machine-colored illustrations, in addition to more realism and detail. Students perceived that the machine-colored illustrations had a limited number of colors due to the dataset, which reflects the color palette of the Victorian age, and wished to see more vibrant colors. There is still room for improvement when colorizing Victorian illustrations by reflecting on the Victorian color palettes used at the time. To improve the quality of the machine-colored illustrations, the deep learning model used to colorize needs to be developed and tailored to colorizing black-and-white illustrations. In addition, the dataset trained with the model needs to be expanded and curated to reflect more accurate and varied colors and details.

When the survey was handed to the students, they were mostly surprised by the fact that the black-and-white illustrations were colorized by computers using deep learning. After the survey session, they began to be more interested in illustrations than before when reading fiction in class, and posed questions about the ways in which the black-and-white illustrations were colorized. Although 60% of students preferred hand-colored illustrations over machine-colored illustrations, I was astounded by the fact that around 24% students preferred machine-colored illustrations. The results of this survey imply that most students prefer to read Victorian novels with illustrations, although 12% students showed negative responses on illustrations, citing that they enjoy using their own imagination when reading. The survey also shows that students tend to prefer machine-colored and hand-colored illustrations over black-and-white illustrations through the responses to questions 2, 3, 4, and 5, and that they generally prefer hand-colored illustrations over machine-colored illustrations.

5. Conclusion

Through the case study, I found promising possibilities for machine-coloring illustrations in Victorian fiction. Colorizing black-and-white illustrations might distort the original illustration, but it ultimately provides more benefit than loss. For example, machine-colored illustrations can help readers better understand text, can satisfy their imagination, and can make the text more entertaining. In addition, there was less resistance than I expected against machine-colored illustrations from the survey participants. After the case study, I observed that the students became more engaged with the illustrations in the texts they read. The next step for this project would be to make it possible for students to colorize black-and-white illustrations on their own using a digital tool. When asked what they wished to see from machine-colored illustrations, most participants answered that they would want to see more vibrant and realistic colors. To satisfy contemporary readers with more colors in the future, I will need to create more well-curated and larger datasets, and fine-tune a more advanced deep learning model for colorization. Through my case study, I found that machine-colored illustrations from Victorian fiction can be used in the literature classroom, despite the limited capacity of colorizing black-and-white illustrations. I believe that machine-colored illustrations, with improvement, will bring more enjoyment and imagination for students in the literature classroom.

Notes

[1] https://www.bl.uk/collection-items/colour-illustrations-from-1911-edition-of-oliver-twist

[2] See H. Kim, "*Victorian400*: Colorizing Victorian Illustrations."

[3] https://www.bl.uk/collection-items/colour-illustrations-from-1911-edition-of-oliver-twist

[4] https://www.bl.uk/collection-items/colour-illustrations-from-1885-edition-of-oliver-twist

[5] Dickens called himself a caricaturist (Harvey 2), but he wished that illustrations in *Oliver Twist* were not illustrated in a caricature.

[6] The different techniques resulted in different styles of illustrations in Dickens's fiction. Hablot Browne mainly drew upon the roulette, and his etching techniques revealed "his power to capture a living human face in a few quick, deft lines, and his ingenuity" with "the ruling machine and stopping-out varnish" (Harvey 184–5). Cruikshank rarely used stopping-out varnish, whereas Browne sometimes took advantage of stopping-out varnish for some effect; illustrators had characteristics within their own technique as well as palette preferences.

[7] In her article "Dickens Extra-Illustrated" Luisa Calè explored extra-illustrations in *Nicholas Nickleby*. Calè argues that the extra-illustrations by Peter Palette and Minss La Creevy reshape "the acts of reading and viewing" Victorian literature (24). In her article "Illustrating Pip and the terrible strange" Jolene Zigarovich examined illustrations by both British and non-British artists in the several editions of *Great Expectations*. For instance, Zigarovich argues that John McLenan's original illustration for the American serialization of *Great Expectations* (1860–1) "intuitively portrays the novel's themes of identity, life, and death" (27). Furthermore, Zigarovich notes that F. A. Fraser's illustration "*And you know what wittles is?*" (1877) accurately depicts Pip's appearance and expression of terror and emphasizes "the church's significance" in the background, while F. W. Pallthorpe's "The Terrible Stranger in the

Churchyard" (1900) luminously depicts Pip's surprise and fear. Richard Stein's essay "Dickens and Illustration" focuses on two illustrators who illustrated Dickens's fiction, Hablot K. Browne (Phiz) and George Cruikshank.

Bibliography

Calè, Luisa. "Dickens Extra-Illustrated: Heads and Scenes in Monthly Parts (The Case of *Nicholas Nickleby*)." *The Yearbook of English Studies*, 2010, pp. 8–32.

Golden, Catherine. "Cruikshank's Illustrative Wrinkle in *Oliver Twist*'s Misrepresentation of Class." *Book Illustrated: Text, Image, and Culture*, 2000, pp. 117–46.

Goodfellow, Ian, et al. "Generative Adversarial Nets," *Advances in Neural Information Processing Systems*, 2014. papers.nips.cc/paper/5423-generative-adversarial-nets.pdf.

Halsall, Alison. "'What is the Use of a Book … Without Pictures or Conversations?': Incorporating the Graphic Novel into the University Curriculum." pp. 87–101.

Hughes, William Lee. "Impersonal Grief: Charles Dickens and Serial Forms of Affect." *differences*, vol. 29, no. 3, 2018, pp. 86–106.

Isola, Phillip, et al. "Image-to-Image Translation with Conditional Adversarial Networks." *IEEE Conference on Computer Vision and Pattern Recognition*, 2017. openaccess.thecvf.com/content_cvpr_2017/papers/Isola_Image-To-Image_Translation_With_CVPR_2017_paper.pdf.

Kim, Hoyeol. "*Victorian400*: Colorizing Victorian Illustrations." *International Journal of Humanities and Arts Computing*, vol. 15, no. 1-2, 2021, pp. 186–202.

Kitton, Frederic. *Dickens and His Illustrators*. London: George Redway, 1899.

Reitter, James. "Dickens and Three Stages of Illustration Evolution." *Interdisciplinary Humanities*, vol. 25, pp. 33–53.

Solberg, Sarah. "Dickens and Illustration: A Matter of Perspective." *The Journal of Narrative Technique*, vol. 10, no. 2, 1980, pp. 128–137.

Steig, Michael. “Martin Chuzzlewit’s: Progress by Dickens and Phiz.” *Dickens Studies Annual*, vol. 2, 1972, pp. 119–362.

Stein, Richard. “Dickens and Illustration.” *The Cambridge Companion to Charles Dickens*, 2001, pp. 167–188.

Zigarovich, Jolene. “Illustrating Pip and the Terrible Stranger.” *Dickens Quarterly*, vol. 32, no. 1, 2015, pp. 21–43.

CHAPTER III

Sentiment Analysis: Limits and Progress of the Syuzhet Package and Its Lexicons

1. Introduction

Text mining is no longer an uncommon research method when it comes to analyzing texts in the digital humanities. Once limited to the research field, text mining now influences "our lives, our teaching, and our scholarship, and digital humanists" (Binder 213) as "a logocentric practice" (Clement 534). Sentiment analysis, also known as opinion mining, shares common features with text mining when parsing, detecting, and locating words or sentences. Sentiment analysis is "the process of extracting an author's emotional intent from text" (Kwartler 85). Sentiment analysis has historically focused on product reviews, such as those of movies, hotels, cars, books, and restaurants, in addition to blog data, but current sentiment analysis has expanded to "stock markets, news articles, [and] political debates" (Medhat et al. 1094), and serves a variety of purposes. There have been attempts at employing sentiment analysis in literature, mainly grounded on lexicon-based approaches, but sentiment analysis in literature has been a target of attack in digital humanities due to its limits as a research method: Swafford's critique of the Syuzhet package made a great impact on the digital humanities field by alerting readers to the danger of choosing faulty tools, although her criticism rehashed already existing issues in sentiment analysis. Along with Swafford's critique of Syuzhet, other digital humanists shared erroneous results found through Syuzhet and expressed uneasy feelings about sentiment analysis in literature.[1] In reality, perfect codes/tools cannot exist, so we need to "embrace 'problems'" with Syuzhet "as a feature rather than a flaw" (Rhody). Ted Underwood asserts that if we "use algorithms in our research," we should "find out how they work" (Underwood 69). Similarly, when using digital tools, it is important to understand their functions, algorithms, and programming syntax, instead of simply drawing upon the visualized results, in order to avoid creating faulty results.

Sentiment analysis is a subfield of natural language processing, which classifies the sentiments of texts. Sentiment analysis researchers traditionally used lexicon-based and machine learning approaches. The machine learning approach uses machine learning algorithms with training datasets to classify sentiments based on linguistic features, whereas the lexicon-based approach draws upon the collection of precompiled sentiment lexicons to label words with sentiment scores. However, these traditional approaches revealed the limits of dealing with complex syntaxes and semantics. Recently, sentiment analysis researchers have proposed deep learning approaches such as transformers, cognition attention-based models, and sentiment-specific word embedding models. Deep learning approaches for sentiment analysis have been considered "as efficient methods due to their capability of learning the text without manual feature engineering" (Habimana et al.). Traditional sentiment analysis approaches mainly drawing upon lexicons have around 70% accuracy, while recent deep learning approaches for sentiment analysis create state-of-the-art results. Sentiment analysis for literary texts, however, is still based on traditional approaches: Kim and Klinger note that "[i]t is true that much digital humanities research (especially dealing with text) uses the methods of text analysis that were in fashion in computational linguistic twenty years ago" (18). Although sentiment analysis has been commonly employed in a variety of fields, mainly for commercial purposes, in addition to testing sentiment analysis with literary texts, sentiment analysis for literature in the digital humanities is relatively new and received little attention until the Syuzhet package was first released, aimed at providing a proper tool for literary analysis. Syuzhet 0.2.0 was released on February 22, 2015 and was soon critiqued by Swafford, who pointed out problems with Syuzhet on her personal blog on March 2, 2015, such as (1) splitting sentences, (2) negators, (3) parts of speech, such as ‘well’ and ‘like,’ (4) lexicons being based on contemporary English words, (5) counting a word once for a sentence even if it is repeated, (6) scoring subjectivity, (7) satire and sarcasm, (8) foundation shapes Despite the effort by Jockers’ lab to create a useful tool for sentiment analysis tailored to analyzing literary texts, the limits of Syuzhet that Swafford pointed out caused digital humanists to have qualms about sentiment analysis in literature. After Swafford’s criticism against Syuzhet 0.2.0, Syuzhet 1.0.0 was released on April 28, 2016, followed by another release on December 14, 2017 of the 1.0.4 version. After almost three years since 1.0.4, Syuzhet 1.0.6 was released with minor updates on November 24. 2020.[2]

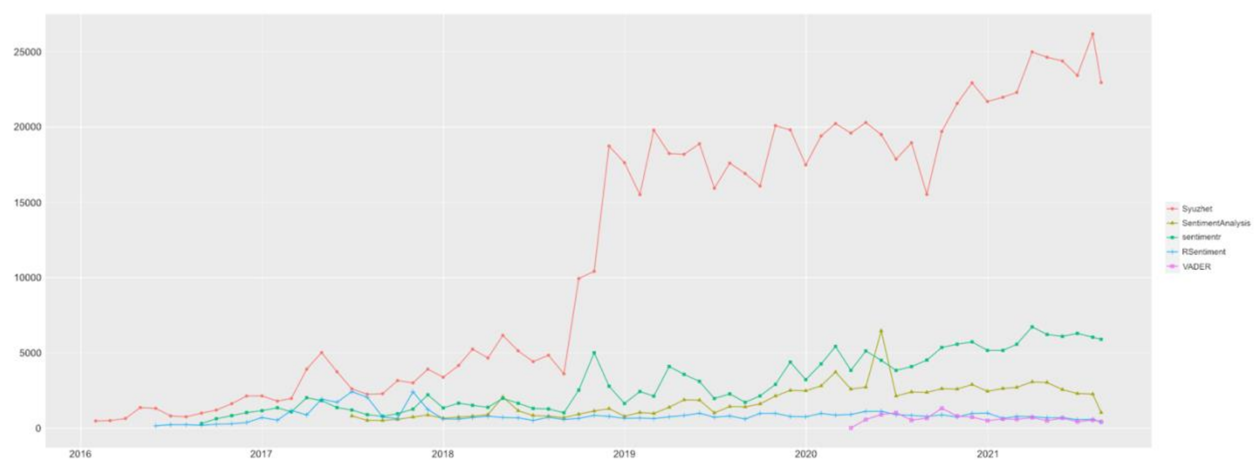


Figure 3.1: The number of monthly downloads for sentiment analysis R packages (created on August 17, 2021).

Figure 3.1 reveals that Syuzhet has been continuously downloaded as the most popular package for sentiment analysis in R.[3] In 2021, it has been downloaded more than 20,000 times monthly, but due to its limits, Syuzhet still remains difficult to validate as a research tool for sentiment analysis in the humanities. In the past, sentiment analysis researchers tested sentiment analysis with literary texts: Saif Mohammad created and tested the NRC lexicon with literary texts such as Shakespeare's *Hamlet* and *As You Like It*, based on the basic emotion models of Ekman and Plutchik. Reagan et al. suggested the "six core emotional arcs" (rise, fall, fall-rise, rise-fall, rise-fall-rise, and fall-rise-fall) for fictional stories. Haider et al. performed sentiment analysis with poems in English and German, using word embeddings as features and manually multi-labeling sentiments. Evgeny Kim and Roman Klinger provided a survey of sentiment analysis in computational literary studies and examined the difficulties of detecting sentiments due to indirectly expressed emotions in literary texts. Michelangelo Misuraca et al. validated Syuzhet, using confusion matrices and macro-averaging with the course_evaluation dataset, of which each sentiment was manually labeled by Charles Welch and Rada Mihalcea. In their test, the overall accuracy of Syuzhet was 0.671, and with the education dataset, the averages for precision, recall, and F-measure were 0.605, 0.526, and 0.526, respectively (Misuraca et al. 22). Jockers asserts that "current benchmark studies suggest that [sentiment detection] accuracy" is "in the 70-80% range and that depends on genre" (Jockers 2015), but the accuracy of sentiment detection in the validation test of Syuzhet by Misuraca et al. was 67.1% (Misuraca et al. 22), which is a little lower than the 70-80% range Jockers argued to defend Syuzhet.

Despite the low accuracy of Syuzhet, it is one of the most popular sentiment analysis tools for R, as Figure 3.1 shows. After the criticism against Syuzhet, it was difficult to find new sentiment analysis research in the digital humanities, although Syuzhet users have drastically increased in the meantime. The problem is that sentiment analysis tools in R heavily draw upon lexicons, which are far from deep learning approaches in regards to methodology. Recently, despite the criticism against Syuzhet, which resulted in digital humanists having reservations towards sentiment analysis as a research method in the humanities, there were a couple of digital humanists who presented at the ACH2021 conference about sentiment analysis in the humanities using VADER (Valence Aware Dictionary and sEntiment Reasoner) for sentiment analysis with humanities data. VADER is a lexicon and rule-based sentiment analysis tool tailored to the sentiment analysis of social media. As Stéfan Sinclair, Stan Ruecker, and Milena Radzikowska emphasize, cultivating a sufficient understanding of digital tools is important since “the interpretive work is being guided and biased by the data and software” (par. 54). While Syuzhet has been controversial as a research method due to its limits, it is still meaningful for helping literary critics grasp what they should consider when performing sentiment analysis.

Therefore, I decided to closely examine Syuzhet 1.0.6 to impart the limits and progress of Syuzhet, with the subjects of my experiment being mainly from 19th century British novels, since they are not under copyright, are long enough to produce valid analyses, and are credited for their well-structured plots. I begin by exploring similar and dissimilar results of sentiment plots, the similarity of deciding positivity and negativity between the lexicons, and the percentage of shared words between lexicons with four lexicons for sentiment analysis: Syuzhet, Bing, Afinn, and NRC. As there are currently no validation datasets for the sentiment analysis of Victorian fiction, I examine the results of sentiment analysis with Charles Dickens’s *Our Mutual Friend*, George Eliot’s *Middlemarch*, and Charlotte Brontë’s *Jane Eyre*. I conclude that Syuzhet needs to be improved in order to capture semantic and syntactic information, that the usage of DCT (Discrete Cosine Transformation) for sentiment analysis plots creates distorted results. Finally, I suggest that we should use deep learning approaches for sentiment analysis in the humanities.

2. Lexicons

The term Syuzhet stems from "the Russian Formalists Victor Shklovsky and Vladimir Propp who divided narrative into two components, the 'fabula' and the 'syuzhet'" to depict narrative structures of story. Syuzhet intends to provide "the latent structure of narrative by means of sentiment analysis" and specifically "the emotional shifts that serve as proxies for the narrative movement between conflict and conflict resolution" (Jockers 2017b). Jockers' explanation of Syuzhet describes it as a sentiment analysis tool for the analysis of literary texts. Syuzhet is a lexicon-based package, mainly drawing upon four standard lexicons: Syuzhet, Bing, Afinn, and NRC.

	Syuzhet	**Bing**	**Afinn**	**NRC**
No. of Positive Words	3587	2006	878	2312
No. of Negative Words	7161	4783	1598	3324
No. of Other Words	-	-	1	8265
Total	10748	6789	2477	13901

Table 3.1: Number of sentiment words in lexicons used in the Syuzhet package.

The Bing, Afinn, and Syuzhet lexicons provide polarity which sorts words into positive or negative positions with numeric values. The Bing lexicon[4] has a binary categorization, which simply has two values of –1 and 1. The Afinn lexicon[5] grades words between –5 and 5. The Syuzhet lexicon has more specific values for each sentiment word, ranging between –1 and 1, which are –1.0, –0.8 –0.75, –0.6, –0.5, –0.4, –0.25, 0.1, 0.25, 0.4, 0.5, 0.6, 0.75, 0.8, 1.0. The NRC lexicon[6] sorts sentiment words into categories consisting of positive, negative, anger, anticipation, disgust, fear, joy, sadness, surprise and trust. The other words from the NRC lexicon in Table 3.1 consist of anger (1247), anticipation (839), disgust (1058), fear (1476), joy (689), sadness (1191), surprise (534), and trust (1231). A number of words from the NRC lexicon are included in different categories at the same time, but the Syuzhet package can only work with positive and negative lexicons from the NRC lexicon. Excluding duplicate words in the different feeling categories of the NRC lexicon, there are 6,468 unique words. Among these, there are 81 words which belong to both positive and negative categories, such as ‘boisterous,’ ‘endless,’ and ‘revolution.’ The Syuzhet package processes those 81 words with a score of 0. In addition, if a word was not categorized as positive or negative, it will score 0. For example, ‘confront’ falls into two categories: anger and anticipation, but scores 0, whereas ‘annoy’ scores –1, which is categorized as negative, anger, and disgust in the NRC lexicons.

Figures 3.2 and 3.3 were created through the `get_dct_transform` function of Syuzhet using four different lexicons, Bing, Afinn, NRC, and Syuzhet, for sixteen novels. In Figure 3.2, the emotional valence of each lexicon is similar over the narrative time from eight novels: Charles Dickens’s *Oliver Twist* and *Little Dorrit*, George Eliot’s *Adam Bede, The Mill on the Floss* and *Middlemarch*, Thomas Hardy’s *The Return of the Native*, Elizabeth Gaskell’s *North and South*, and Mary Elizabeth Braddon’s *Lady Audley’s Secret*.

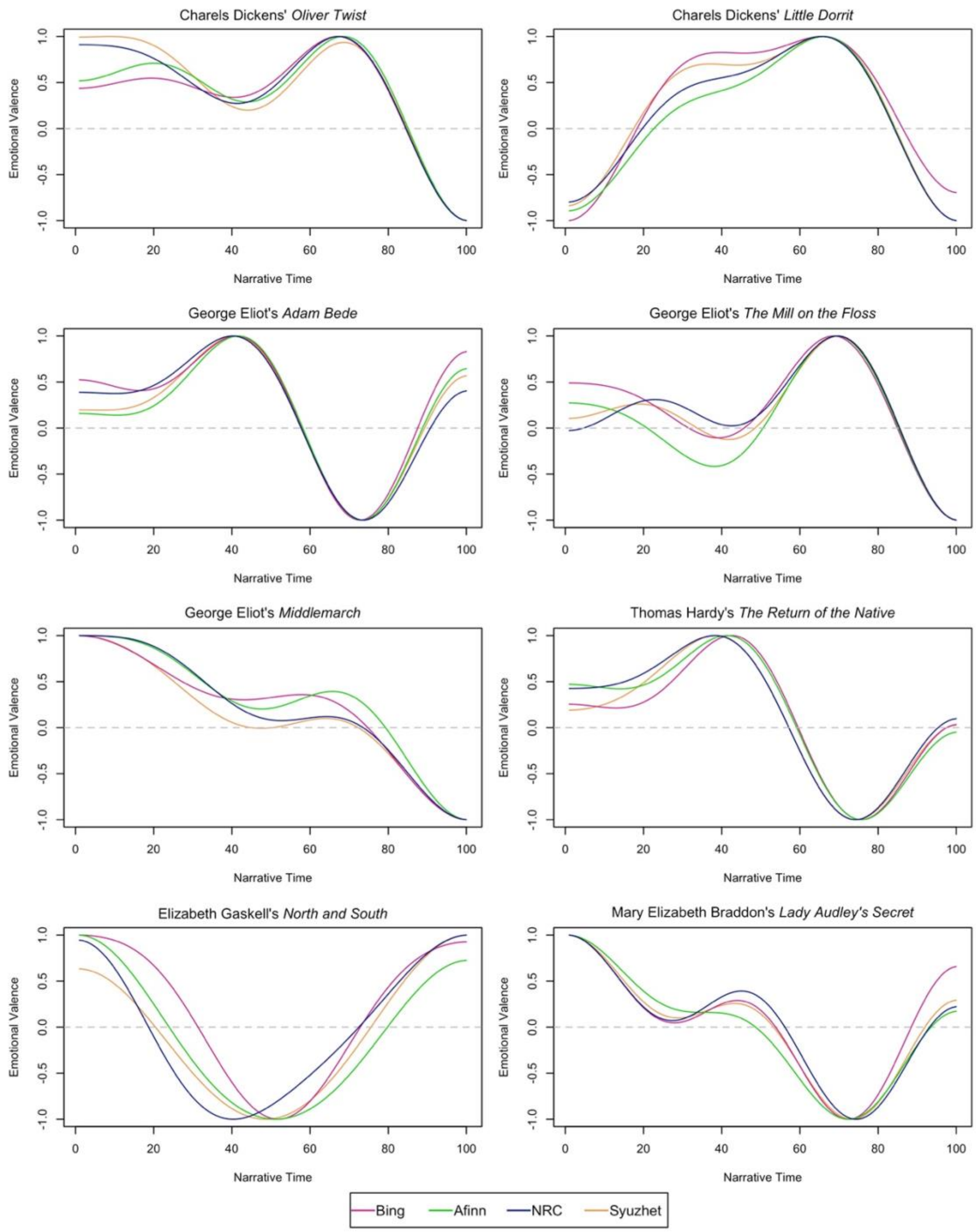


Figure 3.2: Similar results from four different lexicons.

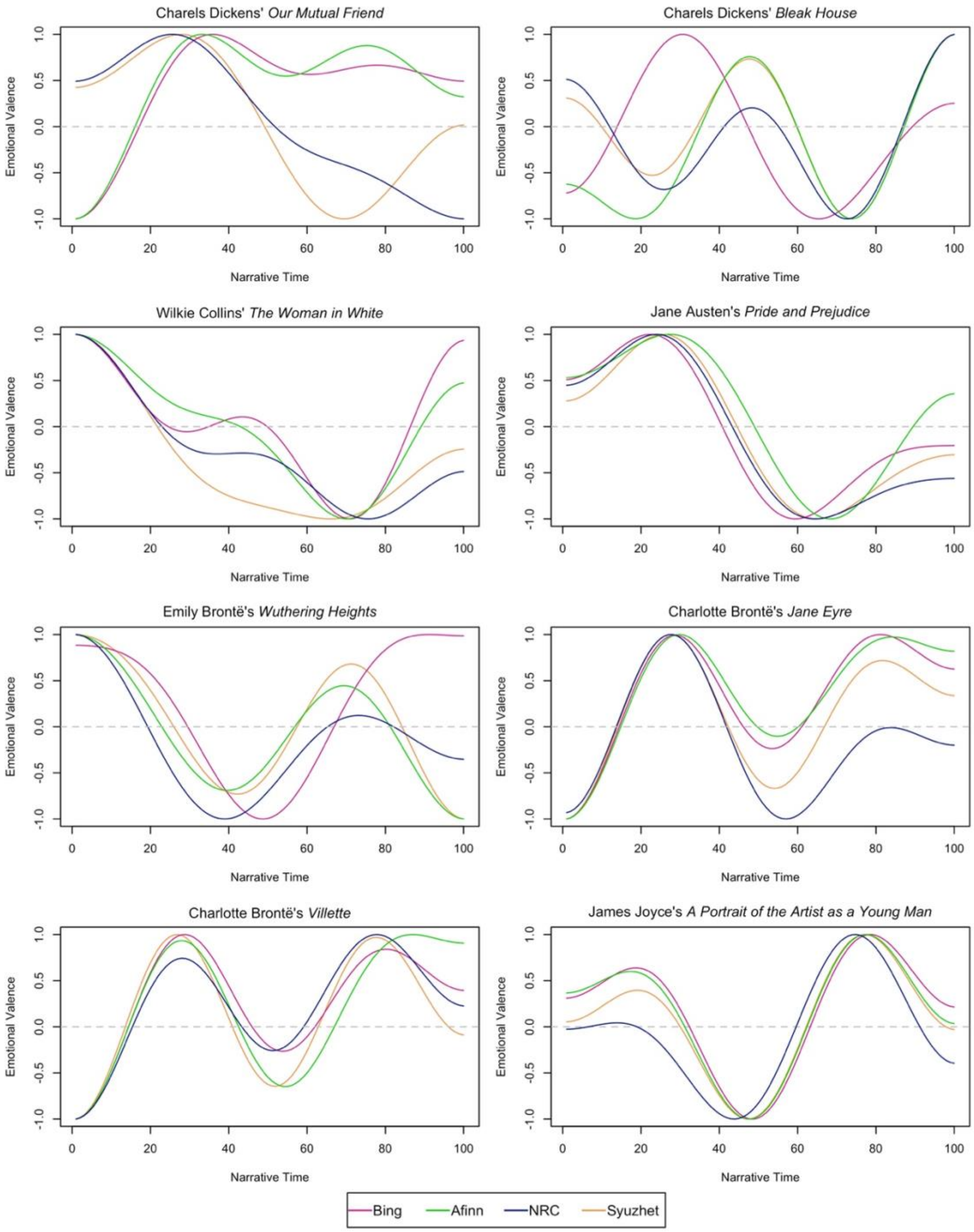


Figure 3.3: Differing results from four different lexicons.

Figure 3.3, however, reveals inconsistent emotional valences from four lexicons for eight novels: Charles Dickens's *Our Mutual Friend* and *Bleak House*, Wilkie Collins' *The Woman in White*, Jane Austen's *Pride and Prejudice*, Emily Brontë's *Wuthering Heights*, Charlotte Brontë's *Jane Eyre* and *Villette*, and James Joyce's *A Portrait of the Artist as a Young Man*.

What causes different sentiment analysis results to be generated depending on the lexicon? I examined the differences between the four lexicons based on positivity and negativity in order to find the reasons why sentiment trajectories could be different between them. Table 3.2 reveals that the Bing and Afinn lexicons have the highest similarity of deciding positivity and negativity, whereas the Syuzhet and NRC lexicons have the lowest number between the results, although the number is still high.

Lexicons (No. of Words)	**Syuzhet-Bing (5,910)**	**Syuzhet-Afinn (2,285)**	**Syuzhet-NRC (4,783)**	**Bing-Afinn (1,315)**	**Bing-NRC (2,396)**	**Afinn-NRC (990)**
Similarity of Deciding Positivity and Negativity	98.26%	98.47%	96.59%	98.71%	98.33%	98.18%

Table 3.2: Similarity of deciding positivity and negativity between lexicons used in the Syuzhet package.

The percent similarity for giving the same words positive or negative values between two different lexicons are the followings: Syuzhet-Bing (98.26%, 5,910 words), Syuzhet-Afinn (98.47%, 2,285 words), Syuzhet-NRC (96.59%, 4,783 words), Bing-Afinn (98.71%, 1,315 words), Bing-NRC (98.33%, 2,396 words), and Afinn-NRC (98.18%, 990 words). This means that the Syuzhet and Bing lexicons have 5,910 common words that, when given positive and negative scores, conflict 1.74% of the time. For example, the words 'avenge,' 'enough,' and 'envy' are scored 0.25, –0.25, and –0.8 by the Syuzhet lexicon, versus –1, 1, and 1 by the Bing lexicon. Looking into the comparison of the Syuzhet and NRC lexicons, the words 'absolute,' 'ancient,' and 'blush' score –0.25, 0.25, and 0.6 in the Syuzhet lexicon, versus 1, –1, and –1 in the NRC lexicon, respectively. These different decisions whether words will be assigned positive or negative can bring about different results during sentiment analysis, as shown in Figure 3.3.

Lexicons (No. of Words)	**Syuzhet-Bing (5,910)**	**Syuzhet-Afinn (2,285)**	**Syuzhet-NRC (4,783)**	**Bing-Afinn (1,315)**	**Bing-NRC (2,396)**	**Afinn-NRC (990)**
Syuzhet	54.99%	21.26%	44.50%	-	-	-
Bing	87.05%	-	-	19.37%	35.29%	-
Afinn	-	92.25%	-	53.09%	-	39.97%
NRC	-	-	73.95%	-	37.04%	15.31%

Table 3.3: Percentage of shared words between lexicons used in the Syuzhet package.

Based on Table 3.3, the percentage of words included in the Syuzhet package that are shared with any given lexicon are relatively low across the board. This is most likely due to the fact that the Syuzhet lexicon was created much later with reference to the Bing, Afinn, and NRC lexicons, and therefore includes words from all three. Because of this, Syuzhet has the most words of any lexicon (not including repeated words in the NRC lexicon) at 10,748 words, causing the disproportion between the percentages of shared words for Syuzhet and the lexicons it is being compared with. Similarly, Afinn, with the fewest words of the four lexicons, when compared with them generates higher percentages for itself.

Despite having the same tool setting conditions, depending on the lexicon, sentiment trajectories could be different due to the subjectivity of the lexicons. The inconsistent sentiment scores of the Syuzhet lexicon result in the discrediting of lexicon-based sentiment analysis. Stephen Ramsay states that literary criticism is not only “a qualitative matter” but also “an insistently subject manner of engagement.” Likewise, creating lexicons is a “subject manner of engagement” (Ramsay 2011, 8) through the subjective interpretation of emotions used in labeling words with scores. Sentiment analysis packages provide customizing functions, either through the customization of dictionaries or the use of dictionaries that are created from scratch, in order to overcome this limit. Nonetheless, it would be challenging to create a dictionary that avoids every critique of subjectivity.

Syuzhet 1.0.6 has not provided a function to use custom dictionaries yet. Syuzhet 2.0.0 is expected to provide the function, but it usually requires a considerable amount of time and effort to create sentiment dictionaries, and customized dictionaries might face the question of reliability and credibility when used in research. Instead of creating a sentiment dictionary from scratch, researchers can use pre-made sentiment dictionaries, such as the psychological Harvard-IV dictionary[7] (DictionaryGI), or customize their sentiment analysis, but they cannot change the sentiment scores from existing lexicons.

3. Syuzhet

3.1 Parsing

The goal of opinion mining is to generate relevant information from texts for analysis. To do so, parsing text is the first step. However, there can be distortions in the process of text mining if raw data are not trimmed. Therefore, well-structured text data need to be inputted for sentiment analysis to generate the correct data. In Syuzhet, there are two different ways to parse text and transform it into vector values: (1) Tokenizing the text into sentences, and then transforming the text into a numeric vector for each sentence. (2) Tokenizing the text into words, and then transforming each word into vector representations. Depending on the purpose of research, the text is tokenized into sentences or words through either the `get_sentences` function or the `get_tokens` function. For the sentiment analysis of novels, the first method, which tokenizes the text into sentences, is normally chosen, so I will focus on parsing the text into sentences using the Syuzhet package. The Syuzhet package originally (versions 1.0.1 and earlier) called upon the

OpenNLP[8] API, which is an open source, in order to implement the `get_sentences` and the `get_tokens` functions. In addition, the Syuzhet package originally required installing Oracle's Java and two R packages, namely 'openNLPdata' and 'rJava' in order to use the OpenNLP parser, which was not user-friendly. Both the `get_sentences` function and the `get_tokens` function parse sentences or tokenize words into numeric vectors of sentiment values. Parsing text is a basic query used to process natural languages, as computers cannot read characters, only numbers. Swafford points out the problems with the OpenNLP parser when grouping sentences, and Jockers responds to her by asserting that the OpenNLP parser and the Stanford CoreNLP parser are "good enough" (Jockers 2015), although he admits that these parsers have problems. In fact, the Stanford parser[9] is a well-constructed tool, which applies a Part-of-Speech (POS) tagging. The OpenNLP parser has been improved, but I found that Syuzhet no longer uses the OpenNLP parser for the `get_sentences` function, despite Jockers mentioning that it does (Jockers 2017a). Instead, Syuzhet draws upon the Textshape package developed by Tyler Rinker for parsing sentences. It seems the Syuzhet manual has not been updated yet, as this change in the parser by Jockers went undocumented. It is possible that Jockers made the change in order to acknowledge the limits of the OpenNLP parser for literary text. Syuzhet 1.0.2 was updated with the removal of the Java dependency, which means that Syuzhet users do not have to install Oracle's Java and its dependent packages, 'openNLPdata' and 'rJava,' anymore to utilize the Textshape package, in addition to parallelization of the `get_sentiment` function by Philip Bulsink on July 28, 2017.

Author | Title	Version	Syuzhet 0.2.0	Syuzhet 1.0.6	Change
Charles Dickens	*Oliver Twist*	6,887	9,128	+32.54%
	Bleak House	18,171	20,319	+11.82%
	Little Dorrit	16,241	18,110	+11.51%
	Our Mutual Friend	15,339	20,261	+32.09%
George Eliot	*Adam Bede*	8,199	8,909	+8.66%
	Mill on the Floss	7,957	8,768	+10.19%
	Middlemarch	13,540	14,415	+6.46%
Charlotte Brontë	*Jane Eyre*	8,605	9,663	+12.30%
	Villette	9,172	10,179	+10.98%
Emily Brontë	*Wuthering Heights*	5,528	6,755	+22.20%
Jane Austen	*Pride and Prejudice*	5,633	5,938	+5.41%
Wilkie Collins	*The Woman in White*	12,675	13,472	+6.29%
Elizabeth Gaskell	*North and South*	8,739	10,418	+19.21%
Mary Elizabeth Braddon	*Lady Audley's Secret*	6,670	7,288	+9.27%
Thomas Hardy	*The Return of the Native*	7,888	8,922	+13.11%
James Joyce	*A Portrait of the Artist as a Young Man*	5,146	5,347	+3.91%
Total Sentences		**156,390**	**177,892**	**+13.75%**

Table 3.4: Comparison of the parsing results from sixteen novels using Syuzhet 0.2.0 and 1.0.6.

In Table 3.4, I compared the parsing results from sixteen novels using Syuzhet 0.2.0 with the OpenNLP parser and Syuzhet 1.0.6 with the Textshape parser in order to examine the improvements of the parsing function in Syuzhet. Table 3.4 reveals the fact that the parsing function of Syuzhet was improved across the board after Syuzhet deployed the Textshape package for parsing instead of the OpenNLP parser. The parsing results from the sixteen novels between Syuzhet 0.2.0 and 1.0.6 have a 13.75% increase. For example, Table 3.5, which shows the parsing result from Charles Dickens's *Our Mutual Friend*, informs that the parsing function of Syuzhet 1.0.2 was improved by splitting sentences more correctly. The OpenNLP parser often failed to split sentences such as: "I'll take the rest of the spell.' 'No, no, father!" In addition, the OpenNLP parser did not split sentences which ended with exclamation and quotation marks. For example, Table 3.5, which is the parsing result from George Eliot's *Middlemarch,* is one of examples that proves that the OpenNLP does not process an exclamation mark as a splitter. In other words, the Textshape package parsed the text into sentences more correctly than the OpenNLP parser based on Tables 3.4, 3.5 and 3.6.

Syuzhet ≤ 1.0.1	**Syuzhet ≥ 1.0.2**
"Has Mr. Casaubon a great soul?" Celia was not without a touch of naive malice.	"Has Mr. Casaubon a great soul?"
	Celia was not without a touch of naive malice.

Table 3.5: Parsing from George Eliot's *Middlemarch* (Chapter 1).

Syuzhet ≤ 1.0.1	**Syuzhet ≥ 1.0.2**
'Here! and give me hold of the sculls.	'Here!
	and give me hold of the sculls.
I'll take the rest of the spell.' 'No, no, father!	I'll take the rest of the spell.'
	'No, no, father!
No! I can't indeed.	No!
	I can't indeed.

Table 3.6: Parsing from Charles Dickens's *Our Mutual Friend* (Book 1, Chapter 1).

Lastly, I tested the parsing function of Syuzhet with Charles Dickens's *Bleak House* to compare the part of chapter 3 where Swafford pointed out grouping errors (see Table 3.6).

Syuzhet ≤ 1.0.1	**Syuzhet ≥ 1.0.2**
Mrs. Rachael, I needn't inform you who were acquainted with the late Miss Barbary's affairs, that her means die with her and that this young lady, now her aunt is dead--" "My aunt, sir!" "It is really of no use carrying on a deception when no object is to be gained by it," said Mr. Kenge smoothly, "Aunt in fact, though not in law.	Mrs. Rachael, I needn't inform you who were acquainted with the late Miss Barbary's affairs, that her means die with her and that this young lady, now her aunt is dead--" "My aunt, sir!"
	"It is really of no use carrying on a deception when no object is to be gained by it," said Mr. Kenge smoothly, "Aunt in fact, though not in law.

Table 3.7: Parsing from Charles Dickens's *Bleak House* (Book 1, Chapter 3).

Based on the parsing result in Table 3.7, the Textshape parser split sentences after an exclamation mark, but not a dash. Syuzhet 1.0.6 with the Textshape parser sorts sentences better than Syuzhet 1.0.1 with the OpenNLP parser. The Textshape parser, however, still has room for improvement for splitting sentences. For example, the Textshape parser infrequently fails to split sentences based on a period, such as: "'My dear, I don't know it,' said I. 'You do,' she said very shortly." (*Bleak House*, Book 1, Chapter 4) in addition to the dash. Based on the parsed result of sixteen novels, I concluded that the Textshape package basically separates sentences based on a period, exclamation mark, or question mark.

3.2 Comparison of Sentiment Values

Syuzhet allocates different numeric vectors to each word/sentence based on the lexicon chosen. These transitioned numeric vectors are turned into structured data or visualization for further analysis. In Syuzhet, there are four different functions to show the emotional valence of stories throughout narrative time: `get_sentiment`, `get_percentage_values`, `get_transformed_values`, and `get_dct_transform`. The `get_percentage_values`, `get_transformed_values` and `get_dct_transform` functions are percentage-based functions, whereas the `get_sentiment` function is based on the number of sentences. The `get_sentiment` function transforms texts into accumulative numeric values for sentiment analysis by matching each word with sentiment scores in selected lexicons. The `get_percentage_values` function "divides a text into an equal number of 'chunks' and then calculates the mean sentiment valence for each." The `get_transformed_values` function uses the Fourier with a low pass filter to make the graph smooth, but Jockers recommends `get_dct_transform` in lieu of `get_transformed_values` because `get_transformed_values` is only being maintained for legacy purposes. The `get_dct_transform` function draws upon "the simpler discrete cosine transformation (DCT)," and its strength is to depict "edge values in the smoothed version of the sentiment vector" (Jockers 2017a). DCT is mostly used in digital media to efficiently process calculations and compress digital media, but it can create errors between data blocks. The fundamental idea of DCT is to compress data for efficiency by removing noise, but in doing so, DCT can distort the original data when performing sentiment analysis.

I tested Syuzhet (1.0.6), SentimentAnalysis (1.3-4), sentimentr (2.7.1), RSentiment (2.2.2), and VADER (R, 0.2.1) with seven different sentences to see how each lexicon-based sentiment analysis tool generates sentiment scores (see Table 3.8). SentimentAnalysis utilizes lexicons such as QDAP (Quantitative Discourse Analysis Package) dictionary, GI (Havard-IV) dictionary, and LM (Loughran-McDonald) dictionary. sentimentr by default uses the combination of an augmentation version of the Syuzhet and Bing lexicons. Similarly, RSentiment uses the Bing lexicon, whereas VADER deploys its own lexicon.

	Sentences	Syuzhet				SentimentAnalysis[10]			sentimentr	RSentiment	VADER[11]
		Syuzhet	Bing	Afinn	NRC	QDAP	GI	LM	Syuzhet & Bing	Bing	VADER
A	She was happy.	0.75	1	3	1	1	1	1	0.433	1	0.572
B	She was not happy.	0.75	1	3	1	1	1	1	–0.375	–1	–0.458
C	She was sad.	–0.5	–1	–2	0	–1	–1	0	–0.288	–1	–0.477
D	She was happy but she is sad now.	0.25	0	1	1	0	0	0.333	–0.397	0	–0.421
E	She was happy, and she is still happy now.	0.75	1	3	1	0.5	0.5	0.5	0.562	2	0.813
F	She was happy but she is no longer happy.	0.75	1	2	1	0.666	0.666	0.666	–0.562	0	–0.391
G	She was extremely happy.	0.75	1	3	1	0	0.5	0.5	0.675	1	0.611

Table 3.8: Experiment in Syuzhet, SentimentAnalysis, sentimentr, RSentiment, and VADER with lexicons.

The sentiment scores of each sentence created with Syuzhet are positive, aside from C. I tested B by replacing 'not' with 'never,' and I got the same result with Syuzhet. Furthermore, C produced –0.5 points, and D generated 0.25 points. The word, 'sad' was given –0.5 points. D scored 0.25 points due to the combination of ‘sad’ (–0.5) and ‘happy’ (0.75). This result indicates that Syuzhet still has issues when semantically detecting sentences, as Swafford has pointed out in the Syuzhet 0.2.0 version. The comparison between A and B shows that Syuzhet has no function to detect negators. D and F depict the lack of a detector for adversative conjunctions in Syuzhet. In addition, the fact that the sentiment score of A is the same with that of G reveals that Syuzhet does not properly detect amplifiers. Table 3.8 demonstrates how Syuzhet simply reports accumulative sentiment scores based on the words in each sentence, as does SentimentAnalysis, while VADER and sentimentr employ detectors for negators, adversative conjunctions, and amplifiers.

VADER and sentimentr provide functions for detecting negators (not, aren’t, no), amplifiers (really, absolutely, very), de-amplifiers (hardly, barely, rarely), and adversative conjunctions/transitions (nonetheless, however, although). Due to the development of machine learning algorithms, dealing with negators is no longer the challenge it used to be. Negators in sentences can be detected and processed through n-grams with high-orders based on supervised algorithms (Jung et al.). Rinker, who developed sentimentr, asserts that negators appear in conjunction with about 20% of polarized words in a sentence. Rinker created valence shifters based on n-grams with high-orders to deal with negators, amplifiers, de-amplifiers, and adversative conjunctions/transitions. Through valence shifters, the accuracy of sentiment analysis has improved, though sentimentr still creates inconsistent results based on the total number of tokens. For example, the sentiment values between the three sentences, ‘She isn’t happy’ (–0.433), ‘She is not happy’ (–0.375), and ‘Today, she is not happy’ (–0.335) are different.

Current sentiment analysis tools still need to improve through alternative approaches. Lexicon-based sentiment analysis has "the inability to find opinion words with domain and context specific orientations" (Medhat et al. 1102). The layers of abstraction must be deeper to semantically and syntactically detect sentences in lexicon-based sentiment analysis tools, which simply transform sentiment words into numeric vectors based on sentiment lexicons, then create visualizations to depict the data. Likewise, Syuzhet still fails to properly deal with negators, amplifiers, de-amplifiers, and adversative conjunctions/transitions.

4. Sentiment Analysis of Charles Dickens's *Our Mutual Friend*, George Eliot's *Middlemarch*, and Charlotte Brontë's *Jane Eyre* through Syuzhet

I selected the Syuzhet lexicon to test four different functions with Charles Dickens's *Our Mutual Friend*, George Eliot's *Middlemarch*, and Charlotte Brontë's *Jane Eyre* in order to examine the compatibility, as well as the limits, of Syuzhet with literature. In Figure 3.4, each function depicts the emotional valence of *Our Mutual Friend* in different ways. Regarding the settings of the `get_transformed_values` and `get_dct_transform` functions, scale_vals=FALSE and scale_range=TRUE. The plot trajectory created by the `get_sentiment` function is complicated and condensed, showing both positive and negative emotion. Nonetheless, it is a useful function when it comes to meticulously grasping sentiment flow in a story.

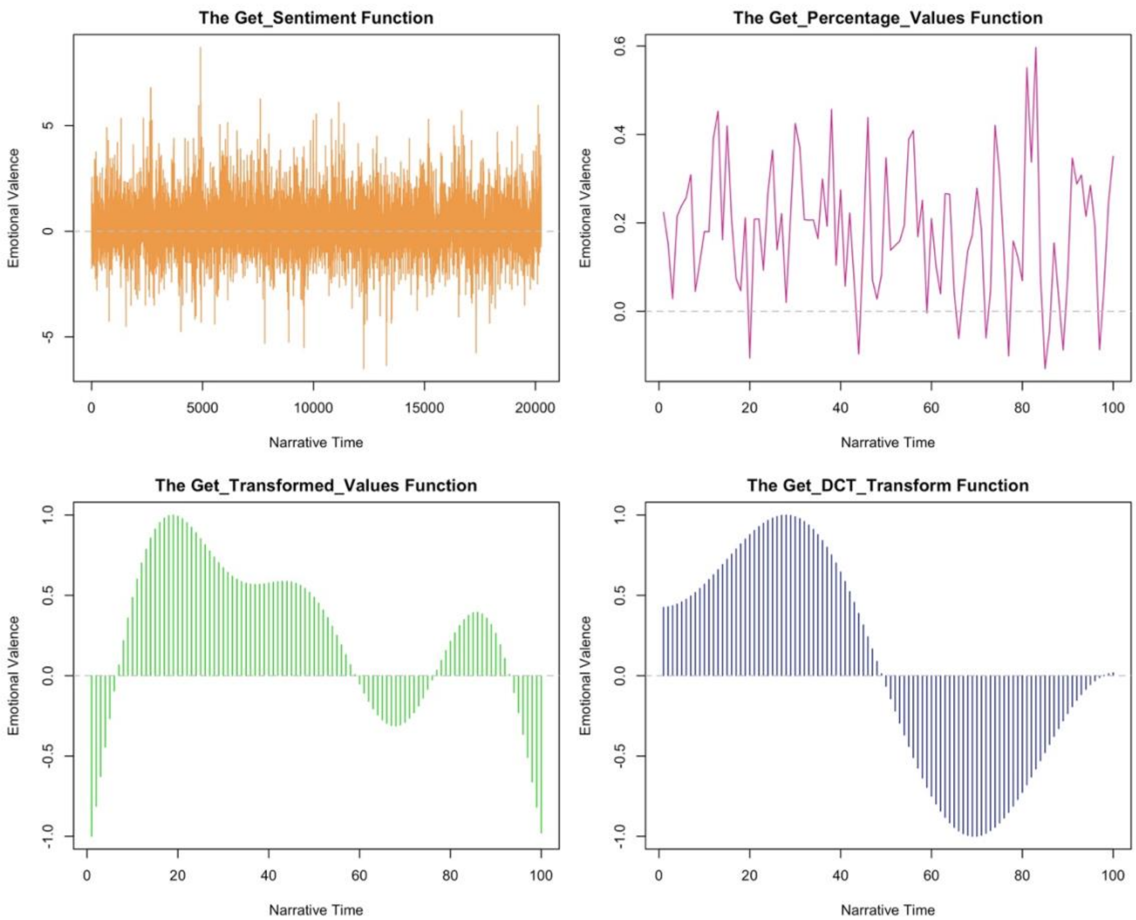


Figure 3.4: Comparison of four different functions based on the Syuzhet lexicon from Charles Dickens's *Our Mutual Friend*.

Looking into the raw file after it was processed by the `get_sentiment` function using the Syuzhet lexicon, 7,167 sentences out of 20,261 sentences scored 0 (neutral), the number of positive sentences was 8,123, and the number of negative sentences was 4,971. The positive average was 0.95, and the negative average was –0.81. Based on the emotion trajectories created by the `get_sentiment` and `get_percentage_values` functions, the whole plot of *Our Mutual Friend* is swayed by positive feelings except for eight chapters. The `get_sentiment` result shows that each chapter entails both positive and negative emotions, and that overall, positive sentiment governs over negative feelings. The `get_percentage_values` function reveals that there are more negative feelings expressed in books 3 and 4. The highest score (8.7) is found in the last chapter of book 1, x=4907: "My Dear Sir,–Having consented to preside at the forthcoming Annual Dinner of the Family Party Fund, and feeling deeply impressed with the immense usefulness of that noble Institution and the great importance of its being supported by a List of Stewards that shall prove to the public the interest taken in it by popular and distinguished men, I have undertaken to ask you to become a Steward on that occasion." The results from the `get_dct_transform` function reveal that *Our Mutual Friend* begins with slightly positive feelings, then reaches a peak of positivity in book 2, before reversing into negativity from book 3. This makes sense, as in book 2, there are a number of jocund and cheerful events, such as Mr. Headstone's and Mr. Eugene Wrayburn's wooing towards Lizzie, Mr. Veneering's luxurious life, Mr. and Mrs. Lammle's social life, Fledgeby's smooth business, Mr. Boffin's purchase of an old mansion, and Bella's taste for money. The lowest score (–6.5), on the other hand, is found in book 3 chapter 8, x=12262: "This boastful handiwork of ours, which fails in its terrors for the professional pauper, the sturdy breaker of windows and the rampant tearer of clothes, strikes with a cruel and a wicked stab at the stricken sufferer, and is a horror to the deserving and unfortunate." The `get_dct_transform` function reveals the dominance of negative feelings in the novel from the halfway point, though it becomes positive once more in the ending. Similarly, between x≈10000 and x≈15000 (Book 3) from the `get_sentiment` function, high values of negative sentiment are often found. Emotions fluctuate in book 3, but the negative atmosphere is dominant in book 3 due to an endless string of troubling plots such as Lizzie's disappearance and return, Mr. Riderhood's drowning, Bella's conflicts about money, Silas Wegg's plot, Headstone's jealousy, Mr. and Mrs. Lammle's bankruptcy, and Mr. Boffin's anger over

Rokesmith. Although chapter 4 is filled with a positive ambience surrounding Mr. and Mrs. Wilfer's wedding anniversary, the emotional flows of the plot shown by the `get_dct_transform` function are relatively correct. Still, it is impossible to assert that the `get_dct_transform` function is 100% correct due to its over-simplification of emotion flows, the inconsistent values of lexicons, and the absence of functions which detect negators, amplifiers, de-amplifiers, and adversative conjunctions/transitions. For example, at x≈20, sentiment is extremely negative in the `get_percentage_values` function, whereas both the `get_transformed_values` and `get_dct_transform` functions have positive values, which are erroneous results caused by the smoothing filter occurring in their functions.

For the first 8% of narrative time, sentiment values are opposite between the `get_percentage_values` and `get_transformed_values` functions, with positive and negative scores respectively (Figure 3.4). Here, the `get_transformed_values` function does not correctly reveal the sentiment trajectories compared to the other functions. As I mentioned above, Jockers does not recommend use of the `get_transformed_values` function, which has been preserved for legacy purposes, but it should be referenced since the `get_dct_transform` function derives from the `get_transformed_values` function. The distinctive difference between the two functions is low pass size. The `get_transformed_values` and the `get_dct_transform` functions have low pass sizes of 2 and 5 respectively, which denotes that the `get_dct_transform` function simplifies sentiment trajectory more than the `get_transformed_values` function does.

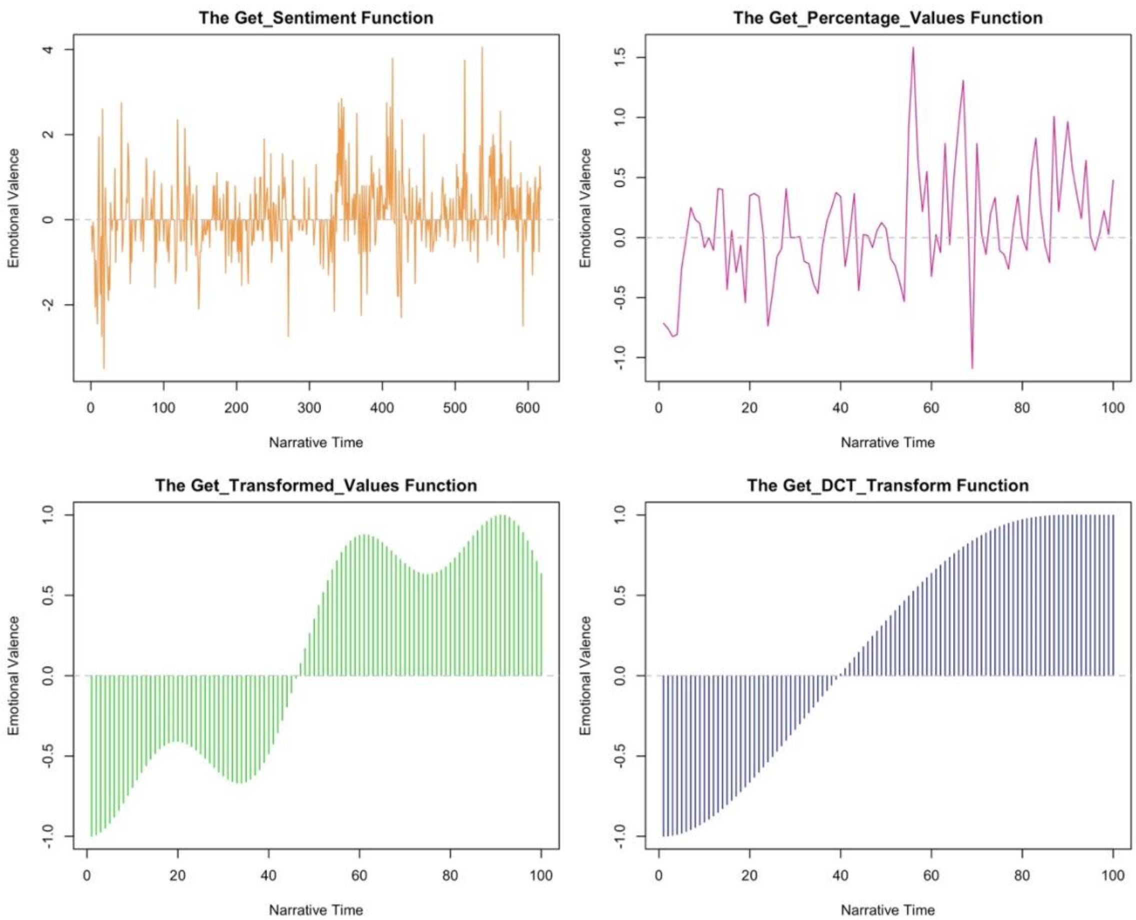


Figure 3.5: Comparison of four different functions from Book 4, Chapter 15 and 16 of *Our Mutual Friend*.

In order to specifically examine the sentiment aspect from Figure 3.4, I chose chapters 15 and 16, both from book 4, which are from x≈96% (19499) to x≈99% (20116) in Figure 3.4. After parsing, chapters 15 and 16 consist of 336 and 282 chunks, respectively. Therefore, in Figure 3.5, chapter 15 is between x=0% and x≈54%, and the rest is chapter 16. Looking into the raw file after it was processed by the `get_sentiment` function with the Syuzhet lexicon, 154 and 96 sentences in chapter 15 and 16, respectively, scored 0 (neutral), 68 and 129 sentences had positive values, and 114 and 57 sentences recorded negative values. Although the number of sentences in chapter 15 and 16 combined is less than 1000, which might bring about incorrect results, the four visualizations in Figure 3.5 appear to appropriately demonstrate the two chapters. Chapter 15 is comprised of Riderhood's blackmail towards Headstone and their subsequent death in the river. The scene which depicts Riderhood staying in Headstone's classroom is filled with tension, and the result of Syuzhet reflects this with negative sentiment values, the lowest of which is (–3.5): "But, not to be still further defrauded and overreached–which he would be, if implicated by Riderhood, and punished by the law for his abject failure, as though it had been a success–he kept close in his school during the day, ventured out warily at night, and went no more to the railway station." In addition, negative feelings are dominant due to Headstone's attempt to drown Riderhood, which results in both of their deaths, and which occurs in the last twenty sentences in chapter 15.

Nonetheless, the foundation shapes created by the `get_transformed_values` and `get_dct_transform` functions depict positive spikes, whereas the trajectories created by the `get_sentiment` and `get_percentage_values` functions at x=317 (x≈51%), to x=336 (x≈54%) correctly show negative spikes. The foundation shapes of Syuzhet, due to its smoothing feature, do not properly handle the drastic sentiment changes from the end of chapter 15, which describes drowning–"When the two were found, lying under the ooze and scum behind one of the rotting gates, Riderhood's hold had relaxed, probably in falling, and his eyes were staring upward," which is given a value of –2.15–to the number of strong positive sentiment values in the beginning of chapter 16. Jockers acknowledges the limits of transforming functions in Syuzhet by noting that "when a series of sentence values are combined into a larger chunk using a percentage based measure, extremes of emotional valence tend to get watered down." (Jockers 2017a). The limit of Syuzhet that Jockers admits to does not seem to be applied in isolation to large data, as it is also seen to affect small data.

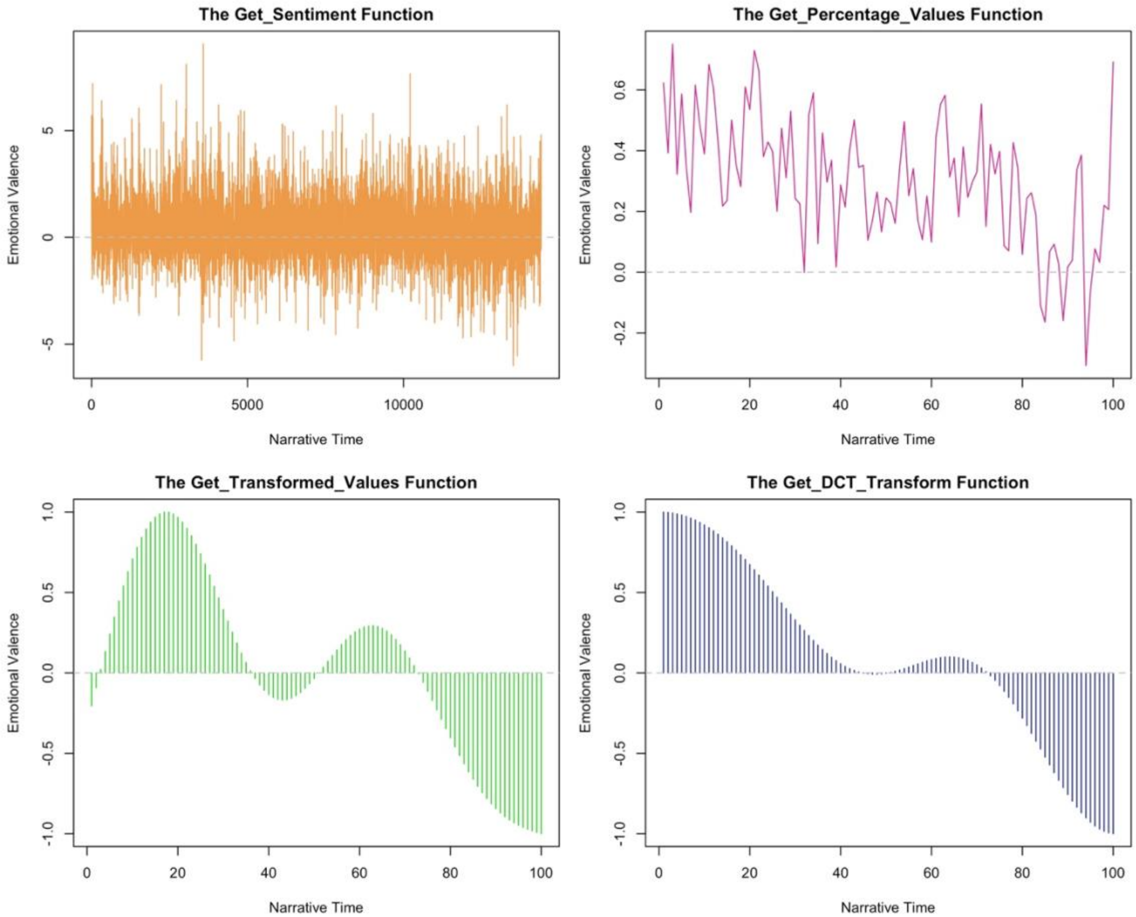


Figure 3.6: Comparison of four different functions based on the Syuzhet lexicon from George Eliot's *Middlemarch*.

Like Dicken's *Our Mutual Friend*, George Eliot's *Middlemarch* is a long Victorian novel, which includes 14,415 sentences after being processed through the `get_sentiment` function using the Syuzhet lexicon. The number of positive, neutral, and negative sentences from George Eliot's *Middlemarch* was 7,286, 3,017, and 4,112, respectively. The positive and negative averages were 1.09 and –0.88, respectively. The emotional valence from the `get_sentiment` and the `get_percentage_values` reveals the dominance of positive emotion throughout the plots, except for the last part, between x≈85 and x≈95. The emotional trajectories from the `get_sentiment` and the `get_percentage_values` precisely depict the ambience of its plots. Although *Middlemarch* has a number of conflicts during the course of the novel between Dorothea Brooke and Mr. Casaubon and between Rosamond Vincy and Lydgate, the flow of *Middlemarch* is generally filled with positive feelings with the exception of the end. With the sudden death of Mr. Casaubon and Lydgate, the last part of *Middlemarch* is dominated with negative feelings. However, *Middlemarch* still has a happy ending as Dorothea decides to get married to Will Ladislaw despite the fact that she has to give up her inheritance from Mr. Casaubon when she does so. Rosamond Vincy also remarries another man after losing Lydgate. Mary and Fred live happily together and have children. The happy ending is from x≈98 through 100 (chapter 86 to the finale). The `get_sentiment` and `get_percentage_values` functions properly catch the happy ending, whereas the `get_transformed_values` and `get_dct_transform` functions do not. In addition, looking into some chapters which have quarrels, there are some parts scored incorrectly by Syuzhet. The highest positive scored sentence is found with a score of 9.05 in chapter 20. Chapter 20 is about the first fight between Dorothea and Mr. Casaubon in Rome after their marriage, which is at x≈25 in Figure 3.6:

> These characteristics, fixed and unchangeable as bone in Mr. Casaubon, might have remained longer unfelt by Dorothea if she had been encouraged to pour forth her girlish and womanly feeling--if he would have held her hands between his and listened with the delight of tenderness and understanding to all the little histories which made up her experience, and would have given her the same sort of intimacy in return, so that the past life of each could be included in their mutual knowledge and affection--or if she could

> have fed her affection with those childlike caresses which are the bent of every sweet woman, who has begun by showering kisses on the hard pate of her bald doll, creating a happy soul within that woodenness from the wealth of her own love (*Middlemarch*, Chapter 20).

“These characteristics” signifies Mr. Casaubon’s “tenacity of occupation and … eagerness.” Looking closely into this long sentence, ‘if’ is the key word. Without ‘if’ in this sentence, it would be correct to give this sentence positive scores. In this sentence, there are 20 words which have sentiment scores out of 134 words through the `get_tokens` and the `get_sentiment` functions: unchangeable (–0.6), encouraged (0.8), womanly (–0.25), feeling (0.25), delight (1), tenderness (0.8), understanding (1), intimacy (0.8), included (0.6), mutual (0.6), knowledge (0.6), affection (1), affection (1), childlike (0.6), bent (–0.4), sweet (0.75), hard (–0.25), happy (0.75), wealth (0.5), and love (0.75). The sum of the tokens is 10.3, but the sentiment score of the sentence level through the `get_sentiment` function is 9.05. This is due to the conjunction, ‘if,’ which affects the sentence level by adding –0.25 with the `get_sentiment` function, though it does not have a sentiment score as a word. The word, ‘affection’ (1) appeared twice, so ‘affection’ (1) was only added once in the sentence level, which reveals that Syuzhet avoids summing duplicate sentiment words in sentence levels. The logic used by Syuzhet is meticulous in order to differentiate word and sentence levels. However, Syuzhet failed to semantically detect this sentence and created a faulty sentiment result. This long sentence would have been given negative scores if Syuzhet had a function to semantically detect sentences. In addition, there is another example to examine, which is the second highest scored sentence at 8.1 in chapter 16, which is at x≈21 in Figure 3.6:

> In Rosamond’s romance it was not necessary to imagine much about the inward life of the hero, or of his serious business in the world: of course, he had a profession and was clever, as well as sufficiently handsome; but the piquant fact about Lydgate was his good birth, which distinguished him from all Middlemarch admirers, and presented marriage as a prospect of rising in rank and getting a little nearer to that celestial condition on earth in which she would have nothing to do with vulgar people, and perhaps at last associate with relatives quite equal to the county people who looked down on the Middlemarchers (*Middlemarch*, Chapter 16).

As seen in the passage above, British authors such as George Eliot, Charles Dickens, and Charlotte Brontë intentionally used colons or semicolons to break long sentences into several parts. Since Syuzhet does not split sentences based on colons, the sentences in the passage above were not separated. This passage reveals Rosamond’s only reason for caring about Lydgate, which is his social rank. It would be more appropriate to consider this passage as having neutral emotion since it is based on Rosamond’s criteria in choosing her husband. In this excerpt, there are 14 words which have sentiment scores out of 108 words through the `get_tokens` and the `get_sentiment` functions: romance (0.5), hero (0.75), profession (0.25), clever (0.75), well (0.8), sufficiently (1), handsome (1), good (0.75), birth (0.6), distinguished (0.6), marriage (0.6), prospect (0.6), celestial (0.4), and vulgar (–0.5). There is no duplicates or conjunctions which would make a different sum between the bag of tokens and the bag of sentences. In addition, some words in this part which might have been considered ‘negative’ have not been scored by the Syuzhet lexicon, such as ‘piquant’ and ‘look down.’ Syuzhet simply added the sum of sentiment words, and concluded this part to be the second highest positive sentence in *Middlemarch*.

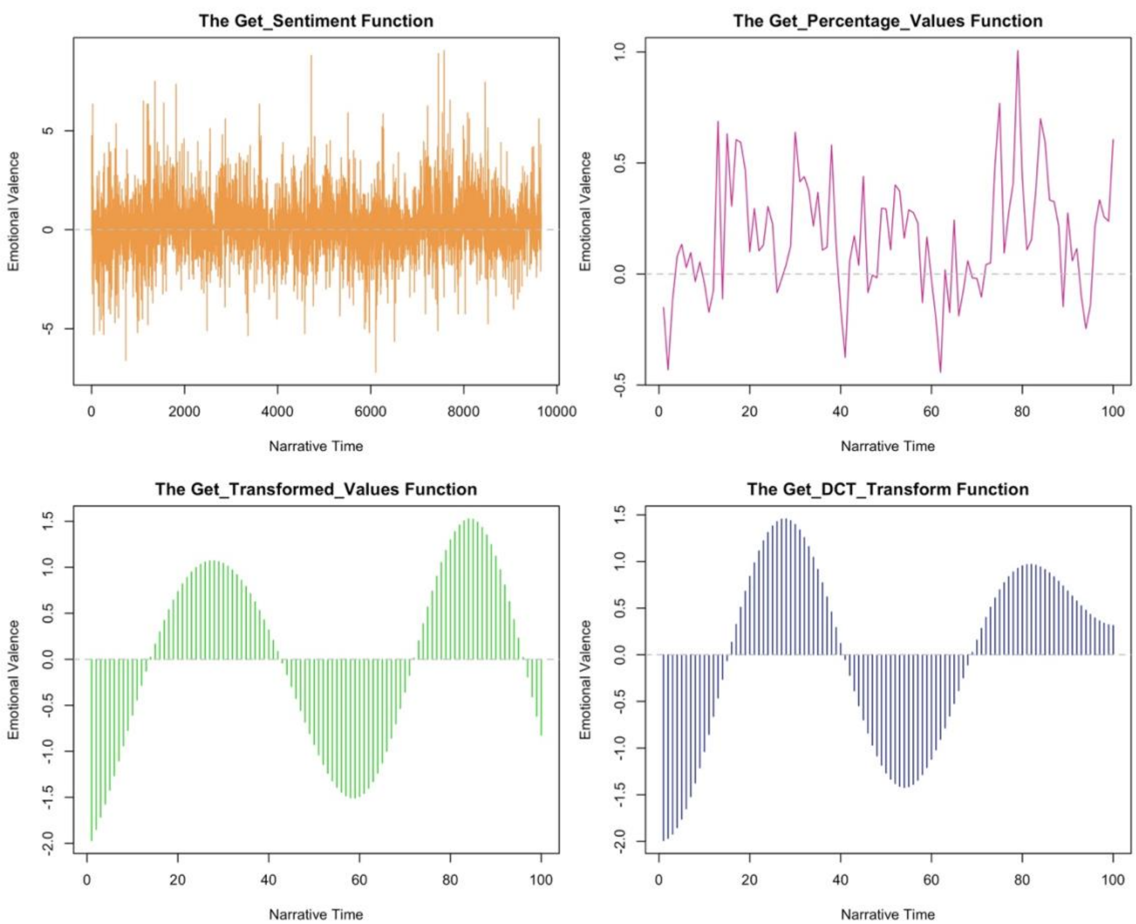


Figure 3.7: Comparison of four different functions based on the Syuzhet lexicon from Charlotte Brontë's *Jane Eyre*.

Charlotte Brontë's *Jane Eyre*, after being processed through the `get_sentiment` function using the Syuzhet lexicon, included 9,664 sentences. Out of these, 2,776 sentences scored 0 (neutral), 4,046 sentences were positive, and 2,824 sentences were negative. The positive average was 1.08, and the negative average was –0.97. Based on the emotion trajectories created by the four functions, emotions from *Jane Eyre* fluctuate between positive and negative feelings throughout the whole plot. The `get_dct_transform` result depicts the emotional flow of *Jane Eyre* as fluctuating between negative, positive, negative, and finally positive feelings, whereas the `get_percentage_values` scrupulously delineates each part with binary emotions. Jane Eyre has difficult times when staying at Gateshead and Lowood due to Mrs. Reed, John Reed, and Mr. Broklehurst, in addition to Helen's death, which occurs from x≈1 to x≈14. Once Jane moves to Thornfield, she has happier days as Adèle's governess with the slow growth of her feelings for Rochester until her wedding. Based on the `get_percentage_values` function, the flow of the emotional valence is positive between x≈15 and x≈60 except for at x≈41. Chapter 20 is full of negative feelings due to Bertha Mason's attack on Richard Mason, which occurs at x≈41 in Figure 3.7. There is a strong negative spike at x≈41: "I saw Mr. Rochester shudder: a singularly marked expression of disgust, horror, hatred, warped his countenance almost to distortion; but he only said-- 'Come, be silent, Richard, and never mind her gibberish: don't repeat it'" which is given a score of –4.5. After the chapter, the flow of the emotional valence is positive until the wedding day. The `get_percentage_values` function correctly depicts the emotional valence of this part, whereas the `get_dct_transform` does not. The wedding was canceled with Mr. Mason's disclosure of the fact that Rochester is already married. Jane reveals her severe feelings when deciding to leave Thornfield: "I wrestled with my own resolution: I wanted to be weak that I might avoid the awful passage of further suffering I saw laid out for me; and Conscience, turned tyrant, held Passion by the throat, told her tauntingly, she had yet but dipped her dainty foot in the slough, and swore that with that arm of iron he would thrust her down to unsounded depths of agony" which is given a score of –4.65 by Syuzhet at x≈61 (Chapter 27). After her marriage is canceled, Jane's hardships continue as a street beggar until she settles in at Moor House and Morton. Jane moves to a small cottage, and again experiences a positive life as a teacher at x≈76 (Chapter 31). When Jane finds Rochester in Ferndean, there are sentences which reveal negative emotions: "He [Rochester] was taken out from under the ruins, alive, but sadly hurt: a beam had fallen in such a way as to protect him partly; but one eye was knocked out, and one hand so crushed that Mr. Carter, the surgeon, had to amputate it directly" which is given a score of –3.25 by Syuzhet at x≈93 (Chapter 36), and which the `get_percentage_values` function detects precisely. The ending of *Jane Eyre* arouses positive feelings with the successful marriage of Jane and Rochester.

The most negative sentence from *Jane Eyre* has a score of –7.2 in chapter 27, which occurs at x≈63 in Figure 3.7, where Rochester explains about Bertha Mason after the cancellation of their wedding.

> "These were vile discoveries; but except for the treachery of concealment, I should have made them no subject of reproach to my wife, even when I found her nature wholly alien to mine, her tastes obnoxious to me, her cast of mind common, low, narrow, and singularly incapable of being led to anything higher, expanded to anything larger--when I found that I could not pass a single evening, nor even a single hour of the day with her in comfort; that kindly conversation could not be sustained between us, because whatever topic I started, immediately received from her a turn at once coarse and trite, perverse and imbecile--when I perceived that I should never have a quiet or settled household, because no servant would bear the continued outbreaks of her violent and unreasonable temper, or the vexations of her absurd, contradictory, exacting orders--even then I restrained myself: I eschewed upbraiding, I curtailed remonstrance; I tried to devour my repentance and disgust in secret; I repressed the deep antipathy I felt" (*Jane Eyre*, Chapter 27).

This passage reveals that Syuzhet does not split sentences based on dashes and semicolons. In this excerpt, there are 27 words which have sentiment scores out of 173 words through the `get_tokens` and the `get_sentiment` functions: vile (–0.75), treachery (–0.5), concealment (–0.8), reproach (–0.5), found (0.6), alien (–0.6), obnoxious (–0.75), incapable (–0.75), led (0.4), found (0.6), comfort (0.75), kindly (0.5), received (0.6), coarse (–0.6), perverse (–0.5), imbecile (–0.75), quiet (0.25), household (0.6), violent (–0.75), unreasonable (–0.5), temper (–0.5), absurd (–0.75), contradictory (–0.5), exacting (–0.25), devour (–0.4), disgust (–1), and antipathy (–0.5). The sum of the word tokens is –7.35. After excluding the duplicated word, 'found,' the sum should be –7.95, but the Syuzhet score is –7.2. This is because Syuzhet perceives words with dashes as being together. In this part, 'imbecile' should have been counted as –0.75, but 'imbecile' was processed as 'imbecile--when' which is considered null by Syuzhet. Although Syuzhet successfully labeled this part as negative, it shows the limits of the Syuzhet functions.

The most positive sentence from *Jane Eyre* scored a 9.05 in chapter 32, which is at x≈78 in Figure 3.7:

> "She was hasty, but good-humoured; vain (she could not help it, when every glance in the glass showed her such a flush of loveliness), but not affected; liberal-handed; innocent of the pride of wealth; ingenuous; sufficiently intelligent; gay, lively, and unthinking: she was very charming, in short, even to a cool observer of her own sex like me; but she was not profoundly interesting or thoroughly impressive" (*Jane Eyre*, Chapter 32).

This is Jane's positive description of Rosamond Oliver. In this part, there are 19 words which have sentiment scores out of 69 words through the `get_tokens` and the `get_sentiment` functions: hasty (–0.5), good (0.75), vain (–1), flush (–0.4), loveliness (1), innocent (0.8), pride (0.25), wealth (0.5), ingenuous (1), sufficiently (1), intelligent (1), lively (0.75), charming (1), cool (0.75), sex (0.1), like (0.5), profoundly (0.8), interesting (0.75), and impressive (0.75). Syuzhet seems to successfully detect this part as positive. The original score should be 9.8 instead of 9.05 since 'good-humoured' was not separately detected in the sentence level due to the dash, which means 'good' (0.75) was not counted towards the sentiment score sum in this part. However, in the last sentence, "but she was not profoundly interesting or thoroughly impressive," Syuzhet failed to detect the negation 'not' and simply added scores from the words, profoundly (0.8), interesting (0.75), and impressive (0.75) without reversing them, which brought about incorrect results.

5. Conclusion

Through the sentiment analysis of the three novels, the `get_transformed_values` and `get_dct_transform` functions do not indicate sophistication of emotion, since their purpose is to grasp the whole emotional flow of plots by simplifying the emotional valence with a smoothing filter, whereas the `get_sentiment` and `get_percentage_values` functions create more detailed results of the emotional valence, which is more appropriate for micro sentiment analysis. Syuzhet reveals its limits through the lack of functions to detect dashes, negators, and adversative conjunctions/transitions, which brings about faulty results. Syuzhet does not detect the syntactical and semantic information of each sentence, but simply transforms each word found in the lexicons into numerical sentiment vectors. In addition, the application of DCT for sentiment analysis of literary texts is still questionable as the graphs of sentiment analysis generated with DCT are over-simplified and often incorrect. Syuzhet has been the most popular sentiment analysis tool for R despite its limits. However, it will continue to be questionable as a research tool in the digital humanities without overcoming the limits mentioned above.

Sentiment analysis has been developed with the implementation of machine learning and deep learning approaches, which attempt to solve the issues it faces. Deep learning in natural language process has shown a shining future for sentiment analysis. For example, convolutional neural networks (CNNs) which include the convolution stage, detector stage, and pooling stage can improve the accuracy of sentiment analysis by detecting locality and negativity of words. Bing Liu notes that opinion words have different meanings depending on the context (Liu 16). For example, the sentences, 'I am not happy to work out' and 'I am happy not to work out,' have different meanings. The locality of 'not' can be processed in pooling layers, which are usually applied after the convolutional and detector stages. For example, MALLET (MAchine Learning for LanguagE Toolkit),[12] a text mining toolkit, employs conditional random fields (CRF), including the Naïve Bayes classifier and decision trees. CRF is an efficient method of natural language processing that fixes the issues of two previous models, namely HMM [Hidden Markov Model] and MEMM [Maximum Entropy Markov Model]. Using deep neural network (DNN) models with word embeddings, which are "typically pre-trained," made it possible for the learned word vectors to "capture general syntactical and semantic information" (Do et al. 276). Similarly, the BERT model, which was created and released by Google AI researchers in 2018, possesses the possibility of application for sentiment analysis with literary texts. BERT, a bidirectional language model, performs a variety of natural language tasks based on a pretrained model with a deep bidirectional transformer that achieves "state-of-the-art performance on a large suite of sentence-level *and* token-level tasks" (Devlin et al. 4172). The recent experiment by Haider et al. (Haider et al. 2020) revealed the inconsistent results of fine-tuning the BERT-Base model for the sentiment analysis of poems, due to the lack of vocabulary in poems. While deep learning cannot achieve perfect results, current research shows that deep-learning based sentiment analysis has higher accuracy than lexicon-based sentiment analysis. Stephen Ramsay mentions that "the real failure would not be a result that is deemed incorrect" but "the decision to banish" computational literary analysis entirely (Ramsay 2016, 529). Although it would require the collaborative creation of literary datasets for deep learning-based sentiment analysis, we should strive to implement deep learning models for sentiment analysis in the digital humanities.

It is painstaking to improve the precision, accuracy, and efficiency of digital tools, and the process entails a great deal of effort, emotion, time, and money, which are also needed to maintain tools after development. Some scholars show disdain for and misunderstanding of the funding necessities for DH projects by stating that "almost all of the works" can be recreated with only one laptop (Da 603). As a mobile/web developer, whenever I had meetings with clients interested in making apps without in-depth knowledge in the IT field, there was always a common qualm about costs to develop apps, before they even thought about the cost of future maintenance. To create a simple app that contains only a few functions requires a project manager, iOS/Android developers, back-end developers, and an UI/UX designer. DH projects are no different: Amy Earhart and Toniesha Taylor shared their experiences facing institutional funding issues while collaborating on a DH project. Due to insufficient funding in the humanities field, it will be challenging to develop new functions and maintain Syuzhet. Syuzhet is a free digital tool that will continue to be developed even though, like any other existing computer program, it is not perfect. I believe that the necessary improvements will be made to Syuzhet for semantically and syntactically detecting sentences, so long as digital humanists support Syuzhet. Improving sentiment analysis as well as digital tools should not remain only as the duty of developers or labs, but as a responsibility of all digital humanists who employ digital tools by participating in making improvements through the provision of feedback, such as that in Swafford's blog post. We need to keep testing and providing feedback to improve tools like Syuzhet for the affluence, development, and application of sentiment analysis in literature.

Notes

[1] Laura Mandell expressed her qualms about sentiment analysis after reading Swafford's post, "Problems with the Syuzhet Package." Jonathan Goodwin shared the incorrect results of Syuzhet on Twitter (see https://twitter.com/joncgoodwin/status/563734388484354048/photo/1).

[2] Check the update notes for Syuzhet at https://github.com/mjockers/syuzhet/blob/master/NEWS. Although some versions of Syuzhet, including Syuzhet 1.0.5, were annotated in the note, some of them were not released to the public.

[3] VADER is also a popular sentiment analysis tool in Python, but the number of VADER downloads in R is low since it was only recently released, on May 22, 2020.

[4] The Bing lexicon was created by Bing Liu and collaborators.

[5] The Afinn lexicon was created by Finn Årup Nielsen.

[6] The NRC lexicon was created by Saif M. Mohammad and Peter D. Turney.

[7] See descriptions of inquirer categories and use of inquirer dictionaries: https://www.wjh.harvard.edu/~inquirer/homecat.htm.

[8] See Apache OpenNLP developer documentation: https://opennlp.apache.org/docs/1.9.1/manual/opennlp.html.

[9] See Stanford parser: http://nlp.stanford.edu:8080/parser/index.jsp.

[10] I excluded the test results of Henry's finance-specific dictionary (HE) since they are all zero.

[11] The sentence "She was happy but she is no longer happy" created different sentiment values between the VADER R and Python packages, with –0.391 in VADER 0.2.1 in R released on September 7, 2020 and –0.665 in VADER 3.3.2 in Python released on July 27, 2018, respectively.

[12] See MAchine Learning for LanguagE Toolkit: http://mallet.cs.umass.edu.

Bibliography

Binder, Jefrrey. M. “Alien Reading: Text Mining, Language Standardization, and the Humanities.” *Debates in the Digital Humanities 2016*, pp. 201–17.

Brontë, Charlotte. *Jane Eyre*. 1847. Introduction by Margaret Smith. Oxford UP, 1973.

Clement, Tanya E. “The Ground Truth of DH Text Mining.” *Debates in the Digital Humanities 2016*, pp. 534–5.

Da, Nan Z. “The Computational Case against Computational Literary Studies.” *Critical Inquiry*, vol. 45, no. 3, 2019, pp. 601–39. doi.org/10.1086/702594.

Devlin, Jacob, et al. “BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding.” *The 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, vol. 1, 2019, pp. 4171–86.

Dickens, Charles. *Our Mutual Friend*. 1864–65. Introduction by Salter Davies. Oxford UP. 1952.

——. *Bleak House*. 1852–53. Introduction by Sir Osbert Sitwell. Oxford UP. 1948.

Do, Hai Ha, et al. “Deep Learning for Aspect-Based Sentiment Analysis: A Comparative Review.” *Expert Systems with Applications*, vol. 118, 2019, pp. 272–99. doi.org/10.1016/j.eswa.2018.10.003.

Earhart, Amy, and Toniesha Taylor. “Pedagogies of Race: Digital Humanities in the Age of Ferguson.” *Debates in the Digital Humanities 2016*, 2016, pp. 251–264.

Eliot, George. *Middlemarch*. 1871–72. Introduction by Barbara Hardy. Oxford UP. 1967.

Habimana, Olivier, et al. “Sentiment Analysis Using Deep Learning Approaches: An Overview.” *Science China Information Sciences*, vol. 63, no. 1, 2020, pp. 1–36.

Haider, Thomas, et al. “PO-EMO: Conceptualization, Annotation, and Modeling of Aesthetic Emotions in German and English Poetry.” *International Conference on Language Resources and Evaluation (LREC) 2020*, Marseille. 2020.

Jockers, Matthew. “Re: Problems with the Syuzhet Package.” [Blog] Anglophile in Academia: Annie Swafford’s Blog, 2015. https://annieswafford.wordpress.com/2015/03/02/syuzhet/.

——. “Introduction to the Syuzhet Package.” 2017. cran.r-project.org/web/packages/syuzhet/vignettes/syuzhet-vignette.html.

——. “Syuzhet Package v1.0.4 | R Documentation.” 2017. rdocumentation.org/packages/syuzhet/versions/1.0.4.

Jung, Yu-Chul, et al. “A Study on Negation Handling and Term Weighting Schemes and Their Effects on Mood-Based Text Classification.” *Korean Journal of Cognitive Science*, vol. 19, no. 4, 2008, pp. 477–97.

Kim, Evgeny, and Roman Klinger. “A Survey on Sentiment and Emotion Analysis for Computational Literary Studies.” 2018. arXiv preprint arXiv:1808.03137.

Kwartler, Ted. *Text Mining in Practice with R*. John Wiley and Sons. 2017.

Liu, Bing. “Sentiment Analysis and Subjectivity.” *Handbook of Natural Language Processing*, vol. 2, 2010, pp. 627–66. cs.uic.edu/~liub/FBS/NLP-handbook-sentiment-analysis.pdf.

Medhat, Walaa, et al. “Sentiment Analysis Algorithms and Applications: A Survey.” *Ain Shams Engineering Journal*, vol. 5, no. 4, 2014, pp. 1093–113. doi.org/10.1016/j.asej.2014.04.011.

Misuraca, Michelangelo, et al. “Sentiment Analysis for Education with R: Packages, Methods and Practical Applications.” 2020. arXiv:2005.12840.

Mohammad, Saif M. “From Once Upon a Time to Happily Ever After: Tracking Emotions in Mail and Books.” *Decision Support Systems*, vol. 53, no. 4, 2012, pp. 730–41.

Ramsay, Stephen. *Reading Machines: Toward an Algorithmic Criticism*. Champaign: U of Illinois P, 2011.

–—. “Humane Computation.” *Debates in the Digital Humanities 2016*, pp. 527–9.

Reagan, Andrew J. et al. “The Emotional Arcs of Stories are Dominated by Six Basic Shapes.” *EPJ Data Science*, vol. 5, no. 1, 2016, pp. 1–12.

Rhody, Lisa. “Re: Problems with the Syuzhet Package.” [Blog] Anglophile in Academia: Annie Swafford’s Blog. 2015. annieswafford.wordpress.com/2015/03/02/syuzhet/.

Sinclair, Stéfan, et al. “Information Visualization for Humanities Scholars.” *Literary Studies in the Digital Age-An Evolving Anthology*, 2013. DOI: 10.1632/lsda.2013.6.

Swafford, Annie. “Problems with the Syuzhet Package.” [Blog] Anglophile in Academia: Annie Swafford’s Blog. 2015. annieswafford.wordpress.com/2015/03/02/syuzhet/.

Underwood, Ted. “Theorizing Research Practices We Forgot to Theorize Twenty Years Ago.” *Representations*, vol. 127, no. 1, 2014, pp. 64–72.

Welch, Charles, and Rada Mihalcea. “Targeted Sentiment to Understand Student Comments.” *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, 2016, pp. 2471–81.

CHAPTER IV

Dickensian Sentiment and Sentiment Analysis of Victorian Novels

1. Introduction

A recent article about Dickensian sentiment by Richard Bonfiglio begins by stating that the dismissive and disparaging reviews of Dickens's sentimental novels by nineteenth-century scholars such as James Fitzjames Stephen, George Henry Lewis, and Walter Bagehot, who considered Dickensian sentiment to be "superficial, effeminate, and déclassé," have made an impact on "modern approaches to sentiment" in Dickens's fiction (Bonfiglio 264; Carney). Under the influence of the pejorative criticism, scholars such as Camilla Cassidy, Richard Menke, and Jonathan Grossman have delved into the ways in which Dickens's novels within the materialist conception of history (Bonfiglio 264), while critics such as Emma Mason, Bethan Carney, and Richard Bonfiglio have considered Dickensian sentiment as a necessary literary device in terms of the exposition of Dickensian characters and plots. Similarly, computational humanities researchers have applied sentiment analysis to literary texts with the development of research approaches while facing resistance from traditional humanists who have prevented sentiment analysis in literature from growing, leading to the marginalization of humanists in debates of sentiment analysis in literature.[1]

Computational humanities researchers have explored Dickens's texts using a variety of digital research methods such as topic modeling, stylometry, and random forests, due to a combination of his being a representative author of the nineteenth century, and the era being free from copyright. Dickens was successful as a commercial writer and reflected Victorian society in his works by purposefully employing sentimentalism. “An understanding of sentimentalist discourse” is, according to Mary Lenard, “clearly crucial for Dickens critics” (106). Charles Dickens was intent on encouraging readers to “interpret the world through its emotional content” (Mason), and he believed that readers could be enlightened through texts by sympathizing with his characters based on moral sentiments. Can sentiment analysis reveal Victorian sentimentality and Dickensian sentiment? In this article, I first look into the definition of sentimental, sentimentality, and sentimentalism, which all stem from sentiment. After that, I introduce Dickens’s sentimentality based on critics’ sentimentalist discourse, examine Dickensian moral sentiments, and see whether or not sentiment analysis helps us understand sentiments in Victorian novels. As a research method for sentiment analysis in this paper, I draw upon lexicons when examining sentences for sentiment analysis and fine-tune the BERT-Base model to complete literary sentiment analysis tasks which predict the sentiments of each sentence based on binary oppositions,[2] using the VictorianLit dataset and deploying the simple discrete cosine transformation to visualize sentiment plots.

2. Definition of Sentimental, Sentimentality, and Sentimentalism

Sentimentalism in literature began in the eighteenth century. The term ‘sentimentality’ began as a “pejorative term in the 1770’s.” By the year 1800, the adjective form, ‘sentimental’ became more common, the negative connotation remaining (qtd. in Howard 71). Sentiment defined by Samuel Johnson in *A Dictionary of the English Language* stands for “rational thought” and “judgment.” However, ‘sentimental’ denotes “affectation ‘in a contemptuous sense’” in the modern world (Banfield 2). According to the 1827 *Dictionary of the English Language* compiled from Samuel Johnson and the 1864 *Comprehensive English Dictionary,* ‘sentimental’ includes three different annotations: thought, feeling and affectation. In addition, sentimentality is defined as exquisite feeling, while sentimentalism contains moral sentiment. The definition of 'sentiment' deployed in eighteenth-century texts played a significant role in terms of morality (Howard 70). In late-Victorian and early-twentieth-century literature, 'sentimentality' was often considered a dishonor and disparagement. Howard, however, argues on the discussion of sentimentality that "[n]or can any account of the form end discussion and produce a consensus for a single definition of sentimentality" (76). The definition of sentimentality differed from its usage by 1800, since Victorian authors used sentiments on purpose in fiction: Dickens, for instance, combined sentiments with morality, known as 'moral sentiments.' Although George Eliot did not deploy Dickens's moral sentiments, she also could not avoid drawing upon sentiments within her moral realism. Based on the European moral philosophy in the eighteenth century, sentimentalism was born again as moral sentiment.

3. Charles Dickens's Sentimentality: Moral Sentiments

Charles Dickens's sentimentality was influenced by eighteenth-century moral philosophers, such as Fielding, who believed that there are "innate differences from birth in the degree of moral sentiment human beings possess" between characters (Kaplan 29), and Goldsmith, author of *The Vicar of Wakefield* (1766), which was considered the bible of moral sentiment in the Victorian age. Sentimentality was "central to the attempt of British literature and philosophy in the first half of the nineteenth century to defend the value of the ideal against the increasingly powerful forces of philosophical realism" (37).

Charles Dickens deployed sentimentalism in his novels in order to "arouse his readers' innate moral sentiments" (50), believing that readers should "experience the sensation of sentimentality in literature" (Mason). Dickens mainly drew upon "pathos and emotionalism" in his novels (Lenard 78). Dickens intentionally evoked his readers pathos through characters who suffer from illness, death, or poverty. Through sentimental events, such as birth, death, marriage, separation, and violence, Dickens stimulated his readers' emotions. Dickens believed that there are moral sentiments in human nature, although some Dickensian characters, such as Silas Wegg and Ralph Nickleby, are void of "moral sentiments to such a negligible degree," that these characters "challenge the assumption that moral sentiment is an innate human quality, a basic constituent of human nature" (Kaplan 63). Dickensian characters, on the other hand, such as Pickwick, Oliver, Little Nell, Nicholas, Fanny Dombey, Esther Summerson, and Little Dorrit demonstrate "human nature defined fully in terms of the moral sentiments" (62). In addition, there are characters in his novels whose lives changed with the effort to learn through other characters as role models, such as in the case of Arthur Clennam with Little Dorrit.

In fact, Dickens has "a much more precise sense" in terms of sentimentality (Howard 72). The purpose of the sentimentality in Dickens's novels was to teach "his audience their social duties" as a "part of sentimentalist discourse of social reform" (Lenard 105-6). Karen Sánchez-Eppler's description of reading sentimental fictions as "a bodily act" reveals the reactions of feelings and bodies to the narratives of stories (110). Dickensian readers unconsciously absorb moral sentiments by reading Dickens's sentimental fiction. Moral sentimentality found in his novels reminds his readers that "the more emotionally sensitive they are to death, the more morally attentive they will be to the values of life," although some of his readers might not feel comfortable with the dramatic sentimentality in his novels (Kaplan 50). For example, Lenard mentions that Dickens deployed "the sentimentalist understanding of death, especially (but not limited to) the deaths of children" in many of his works, such as "the deaths of Little Nell in *The*

Old Curiosity Shop, Little Paul Dombey in *Dombey and Son*, Smike in *Nicholas Nickleby*, William and Frederick Dorrit in *Little Dorrit*, and baby Johnny in *Our Mutual Friend*" (87). Through the death of children, sympathy on the characters is generated and naturally transmitted to readers as moral sentiment, which Dickens expected his readers to associate his texts with.

Adam Smith concluded that there is a significant division in human morality between positive and negative virtues, although the "normativity of Smith's theory is very different for these two categories of moral virtue" due to the impossibility of formulating "a universal idea of the highest good or , more generally, the good life" (Haakonssen viii-ix). Similarly, Dickens, in terms of moral sentiments, created two different kinds of characters in his novels: good characters who have moral sentiments and villains who lack moral sentiments. Little Dorrit is the best example of the good character with moral sentiments that Dickens desired to depict. It might be difficult to find a character who is as full of virtue as Little Dorrit in other Dickens's novels. In terms of moral sentiment, Little Dorrit is contrasted with Bella Wilfer in *Our Mutual Friend*. Little Dorrit's moral sentiments are revealed through her poor environment in the Marshalsea, whereas Bella Wilfer escapes from her poor environment by choosing to live with the Boffins. Kaplan considers Little Dorrit an embodiment of "Dickens' belief in the moral sentiments, part of his effort to depict absolute ideals in a mixed world" (40). Dickens aimed to reveal idealistic moral sentiments through his characters.

Dickens correlated sentiments with morality, then combined them as moral sentiments, a result of his being deeply influenced by British Victorian sentimentality in eighteenth-century moral philosophy. The purpose of Dickens's moral sentiments was to help his readers realize they possess moral sentiments by sympathizing with characters' situations in novels. For example, death evokes sad feelings, in general, and marriage generates happiness. Although there is an unhappy marriage in *Our Mutual Friend* between Sophronia Akershem and Alfred Lammle, Sophronia reveals the fact that she has moral sentiments through the scene: “She bursts into tears, declaring herself the wretchedest, the most deceived, the worst-used, of women. Then she says that if she had the courage to kill herself, she would do it. … Then she cries again. Then she is enraged again, and makes some mention of swindlers” (*OMF* 125; bk. 1, ch. 10). Sophronia and Alfred Lammle deceived one another for a profitable marriage, then Sophronia expressed a variety of emotions, such as rage, sadness, and hopeless, when she found it was a fraudulent marriage. When Alfred Lammle mockingly asked Sophronia if she was being sentimental while attempting to swindle the Boffins, Mr. Boffin interposed their conversation to defend her by narrating, “it’s a very good thing to think well of another person, and it’s a very good thing to be thought well of by another person. Mrs. Lammle will be none the worse for it, if she is” (*OMF* 649; bk. 4, ch. 2). For Mr. Boffin, being sentimental is to reveal that they are considering other people and being cared for by other people, assuming that having sentiments is a desired value in society. Kaplan considers Mr. Boffin as the most representative character for the definition of sentimentality: “Neither the *Oxford English Dictionary* nor any other authority has done better than Nicodemus Boffin in defining sentimentality” (56). Conversely, Rigaud in *Little Dorrit*, who has three names (Lagnier/Blandois), and Roger Riderhood in *Our Mutual Friend* are representative villains, void of moral sentiments. Dickens seems to have admitted that there might be characters who do not possess moral sentiments at all (69). Dickens believed moral sentiment to be the innate ability to learn, although there are some characters who show moral sentiment in latter plots without having them initially, such as Mrs. Clennam and Miss Havisham.

Dickens's fiction pursues encouraging good and punishing evil through moral sentiment. Through the theme of good and evil, Dickensian readers have satisfactory feelings, perceiving social justice as morality. According to Lenard, "the same moralistic sense of social purpose, and reliance on pathos and emotionalism" is often found in Dickens's novels (78). With the usage of sentimentality, Dickens's fiction intended to help his readers have moralistic values. However, using sentimentality is not limited to Dickens's fiction. The word "sentiment" including its derivatives, is found 29 times in *Our Mutual Friend* and 27 times in *Little Dorrit*, and similarly 30 times in Charlotte Brontë's *Villette* and 28 times in George Eliot's *Middlemarch*. Although the appearance of the word "sentiment" does not directly support the idea that the writers used sentimentalism in their novels, it denotes that they were consciously or unconsciously aware of sentimentalism. Victorian authors could not avoid sentimentalism, as literary texts reflect contemporary values and social streams. For example, George Eliot had "the voice of a higher culture, learned, self-reflexive, tormented by her own aesthetic and moral aspirations" (Levine 1). As Dickens emphasized moral sentiments through his characters, George Eliot also deployed moral sentiments to make a decision under harsh circumstances. Josephine McDonagh asserts that the end of *The Mill on the Floss* emphasizes an "emotional conflict, which focuses attention on questions of individual human development" by comparing the relationship between Maggie and Philip Wakem, who is the crippled son of her father's enemy, as a "conflict of duty versus feeling or compassions" and between Maggie and Stephen, duty versus sexual attraction (54; 52).

The Victorians continuously drew "directly and deeply on moral philosophy" (Howard 72). Victorian authors who deployed sentiment in fiction drew upon the moral philosophy that was dominant in the eighteenth British society. 'Sentimentality,' however, arose "in the reactions against the elevation of emotional sensitivity to the status of a moral touchstone" (71). Furthermore, 'sentimental' had commonly a pejorative connotation by 1800 (Todd 9). However, due to the fact that Dickens combined sentiment with morality as moral sentiment in his fiction, the negative connotations of sentiment began to fade away. For example, Kaplan asserts that Dickens affirmed "the Victorian belief that a wet face is not an embarrassment" in *Old Curiosity Shop*, and revealed that tears are "effective expressions and communication of moral feeling" (45). Dickens's moral sentiments not only changed the perspective on feelings but also helped the Victorians form an emotionally healthy society based on moral sentiment.

4. Application of Sentiment Analysis to Literary Texts: Charles Dickens's *Our Mutual Friend*, George Eliot's *Middlemarch*, and Charlotte Brontë's *Jane Eyre*

Sentiment analysis, also known as opinion mining or emotion AI, is the processing of deploying natural language processing to classify opinions toward entities such as movies, food, and products or to identify human emotions or affects through conversations or descriptions. Sentiment analysis began with the aim to employ binary classification on tweets and product reviews and expanded to news articles and blog posts. Recently, sentiment analysis has expanded its usage to improve human lives, especially for those who live alone, with interactions between humans and AI assistants. Digital humanities scholars and computer scientists have explored sentiment analysis as a subfield of computational literary research by applying it to literary texts. Sentiment analysis research, however, has not been acknowledged in the digital humanities as much as in the computer science field, due to resistance from humanists. In spite of this resistance against deploying sentiment analysis in literature, sentiment analysis in literature has the potential to be a useful research approach for humanists by revealing the latent structures of emotion in literary texts when distant reading.

Sentiment analysis in literature has made steady progress with the study of theories, methods, and lexicons. Saif Mohammad (2012) applied the NRC lexicon to the sentiment analysis of literary texts, mainly Shakespeare's works, based on the basic emotion model (Ekman 1992; Plutchik 1980, 1991), which consists of eight different emotions in addition to positive and negative labels. Reagan et al. claimed that fictional stories basically contain "six core emotional arcs" (rise, fall, fall-rise, rise-fall, rise-fall-rise, and fall-rise-fall) through three methods, namely "matrix decomposition by singular value decomposition (SVD), supervised learning by agglomerative (hierarchical) clustering with Ward's method, and unsupervised learning by a self-organizing map." Furthermore, they assert that "The emotional arc of a story does not give us direct information about the plot or the intended meaning of the story, but rather exists as part of the whole narrative (e.g., an emotional arc showing a fall in sentiment throughout a story may arise from very different plot and structure combinations)" (Reagan et al. 2016). Gao et al. (2016) suggested that the Hurst parameter for sentiment analysis of literary texts should be "larger than ½ but cannot be too close to 1" as a "sentiment time series always possesses long-range correlations" when the Hurst parameter is set to more than ½. In addition, they emphasized that "the spikes, troughs, and zeros of the smooth trend signals of sentiment" make each novel distinct as a proxy of plot development. Lexicon-based sentiment analysis tools can generate the emotional valence of a narrative based on the sum of emotional values from all words in each sentence found in a given lexicon, such as the Bing, Afinn, or NRC lexicons. Lexicon-based sentiment analysis classifies each sentence based on the sum of positive or negative values of words using lexicons, whereas machine learning-based sentiment analysis decides if each sentence or word is positive, neutral, or negative. The accuracy of lexicon-based sentiment analysis in literature is usually around 70% depending on the genre, whereas deep learning models with large datasets, such as the BERT model, can create state-of-the-art results for sentiment analysis in literature.

Recently, for the sentiment analysis of literary texts, an increase in the use of machine learning-based approaches has been triggered by the success of the BERT model, which is based on a transfer learning method, in order to overcome limits of lexicon-based sentiment analysis. Haider et al. (2020) show the questionable results of fine-tuning the BERT-Base model for the sentiment analysis of poems due to the limited vocabulary. However, they reveal the possibility of employing multiple sentiments for the sentiment analysis of literary texts by conducting experiments with a poetry dataset annotated with multiple labels per line, using the BERT model. In machine learning-based sentiment analysis, binary sentiment analysis with positive and negative labels has been commonly used, but these two categories cannot cover the entire range of emotions that exist within literary texts. Due to challenges with processing the variety of possible emotions in computing, I performed a multi-class sentiment analysis task with five labels (very negative, negative, neutral, positive, and very positive) for Victorian literary texts.

Sentiment analysis is fundamentally different from sentimentalism and sentimentality. Sentiment analysis is the classification of words or sentences based on basic emotion models (Ekman 1992; Plutchik 1980, 1991), binary (positive and negative) classification, or lexicons, whereas sentimentalism and sentimentality are combined with social and moral values. For example, Bella hates the poor environment of her family: “I hate to be poor, and we are degradingly poor, offensively poor, miserably poor, beastly poor” (*OMF* 37; bk. 1, ch. 4). Based on the Bing and Afinn lexicons, the word ‘poor’ is a negative word, and the words ‘degradingly,’ ‘offensively,’ and ‘miserably’ are negative amplifiers as well. Should speaking about being poor be considered an expression of negative feelings? Being poor does not equate to being immoral, but the word ‘poor’ has a negative connotation according to sentiment analysis. Bing and Afinn lexicons label ‘poor’ with negative points, –1 and –2 respectively. Meanwhile, the word ‘money’ has a positive score in the NRC lexicon of 1. In *Our Mutual Friend*, Bella expresses that she would marry a man for money if she could: “I have made up my mind that I must have money […] I hate and detest being poor, and I won’t be poor if I can marry money” (320-1; bk. 2, ch. 8). These sentences are considered negative sentiments based on the Bing and Afinn lexicons. In sentiment analysis, the words ‘money’ and ‘poor’ have positive and negative labels, respectively. However, in Dickens’s fiction, pursuing money, also known as mammonism, often has the connotation of immorality. Poor characters are described as good people who need to be sympathized with. Sentiment analysis is to perform tasks to categorize words or sentences based on sentiments, whereas sentimentalism is closely attached to social and cultural values such as moral sentiments. That is, sentiment analysis cannot simply be applied using Dickens’s moral sentiments without reflecting the complexity of sentiments in sentiment analysis models.

Despite the fundamental difference between sentiment analysis and sentimentalism, it is still helpful to use sentiment analysis when trying to grasp the sentimental flow of plots in literary texts, as Gao et al. claimed that “sentiment is a good proxy of plot development.” Sentiment analysis has three different classification approaches: a machine learning approach; a hybrid approach; and a lexicon-based approach. For the sentiment analysis of literary texts, computational literary scholars originally deployed the lexicon-based approach, which draws upon “a collection of known and precompiled sentiment terms” (Medhat et al. 1098), but have recently begun to apply machine learning-based sentiment analysis to literary texts (Reagan et al.). Lexicon-based sentiment analysis commonly has issues detecting a variety of different sentence structures such as sarcasm, irony, oxymoron, and puns, in addition to subjectivity. Machine learning-based sentiment analysis such as BERT has shown improved results for sentiment analysis tasks. For this reason, I deployed the BERT-Base model for a sentiment analysis task in literature to achieve high accuracy, the VictorianLit dataset to fine-tune the model, and a simple discrete cosine transformation in order to achieve higher accuracy and have distinct visualizations with emotion arcs. The outputs of sentiment analysis would not be distinguishable without the smoothing filter based on cosine transforms, though the smoothing filter experiences distortion due to transformed values with sine and cosine formulas.[3]

The BERT-Base model (12-layer, 768-hidden, 12-heads, 100M parameters) with the VictorianLit dataset proved to be 93% accurate, along with having high numbers for precision, recall, and F1 scores for each label. The BERT model, created and released by Google AI researchers in 2018, is a bidirectional language model which performs a variety of natural language tasks based on a pretrained model with a deep bidirectional Transformer that achieves "state-of-the-art performance on a large suite of sentence-level *and* token-level tasks" (Devlin et al.). BERT deploys Transformer, which generates attention marks that reveal contextual relationships between words in a sentence. Since BERT is a bidirectional language model, it is possible for it to learn the context of a word in a sentence. This helps the BERT model achieve high accuracy when performing language predictions in a text.

The VictorianLit dataset is for the machine learning-based sentiment analysis of Victorian novels and consists of five Victorian novels: Charles Dickens's *Little Dorrit* and *Oliver Twist*, Elizabeth Gaskell's *North and South*, George Eliot's *Adam Bede*, and Mary Elizabeth Braddon's *Lady Audley's Secret*. There are two columns (sentences and labels) in the dataset. The labels consist of five different sentimental values: 0 (very negative), 1 (negative), 2 (neutral), 3 (positive), 4 (very positive). The number of sentences in the dataset with a score of 0 (very negative) and 4 (very positive) are 1,848 (3.43%) and 1,595 (2.96%), respectively. The number of sentences with a score of 1 (negative) and 3 (positive) are 11,860 (22.03%) and 11,888 (22.09), respectively. Neutral sentences account for the largest portion, at 26,635 (49.48%). It is common for the neutral label to make up the largest percent of sentences in sentiment datasets.

Label	Precision	Recall	F1	Support
very negative	0.8177	0.8315	0.8245	178
negative	0.8998	0.9193	0.9095	1153
neutral	0.9554	0.9493	0.9523	2663
positive	0.9345	0.9238	0.9292	1221
very positive	0.8294	0.8443	0.8368	167
accuracy			0.9300	5382
macro avg	0.8874	0.8937	0.8905	5382
weighted avg	0.9303	0.9300	0.9301	5382

Table 4.1: Precision, recall, and F1 scores of the BERT-Base model with VictorianLit for each sentiment label on the test set.

Figure 4.1: Sentiment co-occurrence matrix for a test set from VictorianLit.

I chose three Victorian novels—Charles Dickens's *Our Mutual Friend*, George Eliot's *Middlemarch*, and Charlotte Brontë's *Jane Eyre*—in order to test the effectiveness of sentiment analysis in literature as a distant reading method, based on the BERT-Base model with the VictorianLit dataset. To predict sentiment values for each sentence from the selected novels, the BERT-Base model was trained for 4 epochs with a batch size of 16 and 1e-5 learning rates. The VictorianLit dataset was randomized and split into training (80%), test (10%), and validation (10%) sets. The accuracy with the validation set was 93%. The maximum length of a sentence in the VictorianLit dataset is 374, but I set the parameter to 400 to account for attention marks. The average training loss for each epoch was 0.59, 0.25, 0.17, and 0.12. The average training loss reveals how well the model works in training and validation sets. The accuracy indicates how accurately the trained model predicts a validation set. This revealed that the model was successfully trained with VictorianLit as the average training loss decreased in each epoch. In order to get visualizations for the results of sentiment analysis, I deployed the simple discrete cosine transformation to reduce noise and to catch the trends of sentiment signals as "smoothing is a key issue in sentiment analysis" (Gao et al.).

Table 4.1 is a detailed breakdown of each sentiment label as predicted by the model. Precision, recall, and F1 scores are high overall, although the scores of the 'very negative' and 'very positive' labels are relatively lower than other labels. Figure 4.1 is a sentiment co-occurrence matrix for a test set from VictorianLit which shows the predicted sentiment for each label compared to its true sentiment. The mismatch ratio for the predicted results is very low. The number of sentiments predicted as positive or very positive sentiment that actually had negative or very negative sentiment (or vice versa) was only 6 out of 5382 (0.11%). The high accuracy of this test result reveals that the BERT-Base model is properly trained with the VictorianLit dataset for a multi-class sentiment analysis task.

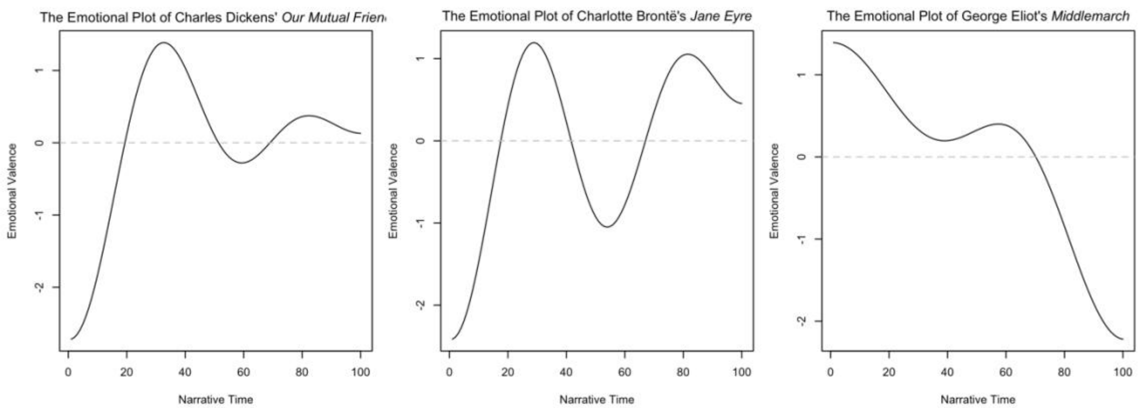


Figure 4.2: Sentiment analysis using the BERT-Base model with the VictorianLit dataset.

In Figure 4.2, the emotional valence of *Our Mutual Friend* is very similar to that of *Jane Eyre*, which has several transition points where emotions are reversed. According to Reagan et al., there are "six core emotional arcs" (rise, fall, fall-rise, rise-fall, rise-fall-rise, and fall-rise-fall) in fiction. *Our Mutual Friend* and *Jane Eyre* each have a rise-fall-rise plot. *Our Mutual Friend* begins with the dark scene of Lizzie and Gaffer Haxam fishing the body of John Harmon out of the river, while in book 2, a number of cheerful events, such as Mr. Boffin's purchase of an old mansion, Mr. Headstone's and Mr. Eugene Wrayburn's wooing towards Lizzie, Mr. Veneering's fancy life, the Lammles's social life, and Fledgeby's successful business, occur. Book 3 of *Our Mutual Friend* is full of primary conflict events, such as Lizzie's disappearance and return, Mr. Riderhood's drowning, Silas Wegg's secret plot, Headstone's jealousy, the Lammles's bankruptcy, Mr. Boffin's fury over Rokesmith, and Bella's inner conflicts about money. Book 4 of *Our Mutual Friend* is called "A Turning," which alludes to how the emotional valence turns from negative to positive once again. Similarly, *Jane Eyre* begins with flashbacks of Jane's childhood at Gateshead Hall and later, her move to Lowood. This period between x≈0 and x≈16 covers Jane's depressed childhood as an orphan up until she leaves for Thornfield to work as a governess for Adèle. Jane's life at Thornfield is peaceful, joyful, and blissful, but in chapter 20 (x≈41), when Bertha Mason attacks Richard Mason, the emotional valence moves towards negative. The negative emotion arc spikes at x≈58 when Jane discovers that Rochester is already married to Bertha Mason. Then, the emotional valence turns positive once again when Jane becomes a teacher at the village school in Morton, inherits money from her uncle, and marries a now blind Rochester in the end of the book.

The shape of the emotional valences in *Middlemarch* are different from *Our Mutual Friend* or *Jane Eyre* since the overall trajectory of the emotions in *Middlemarch* falls towards the end of its plot, similar to the 'Oedipus' emotional plot category coined by Reagan et al. As opposed to the emotional valence of *Our Mutual Friend* and *Jane Eyre*, *Middlemarch* begins with a positive emotional valence and stays positive, dipping closer to 0 between x≈25 and x≈65, despite the fact that there are several conflicts between Dorothea and Mr. Casaubon and between Rosamond and Lydgate. The emotional valence, however, drastically reverses to negative with the sudden death of Mr. Casaubon and Lydgate. *Middlemarch* seems to end somewhat happily, with marriages between Dorothea and Will Ladislaw, Rosamond's marriage with another man, and the contented lives of Mary and Fred with their children, but overall, the conclusion is complicated and realistic. In the Finale of *Middlemarch*, Eliot implies that individuals cannot be unchained from cultural constraints and social pressures because "Dorothea's second marriage" was perceived by "a younger generation" as a marriage to "a sickly clergyman" (Eliot 837; Finale). Eliot realistically concludes the fiction by noting that "[c]ertainly those determining acts of her life were not ideally beautiful. They were the mixed result of young and noble impulse struggling amidst the conditions of an imperfect social state, in which great feelings will often take the aspect of error, and great faith the aspect of illusion." Eliot refuses to finish the story with simply a happy ending, instead leaving space for readers in the form of "unvisited tombs" (Eliot 838; Finale).

5. Conclusion

Sentiment analysis in literature has been developed with advanced research methods, but sentiment analysis is fundamentally different from sentimentalism and sentimentality, and cannot properly reflect Dickensian sentiment. Sentiment analysis, however, can be helpful for distant readers in grasping the flow of emotions in Victorian novels despite its imperfection. Just as Thomas Carlyle was "hostile to sentimentality mainly because he associated it with sensuality, particularly in fiction," due to his misunderstanding of Dickens's sentimentality (Kaplan 8), sentimentality should not be understood to be identical with sentiment analysis. Dickens's sentimentality is principally moral sentiment, which evokes in readers moral and empathetic feelings toward characters. To understand Dickensian sentiment and sentimentalism in Victorian novels, it is required to have varied literary, historical, and social knowledge. Sentimentalism, in literary texts, however, cannot be detected at a sentence level even in contextual deep learning models. Sentiment analysis cannot always reflect the contemporary social values, morality, and stylistics of texts, but it can provide meaningful results for predicting sentiment plots and measuring sentiments within sentences. Deep learning-based sentiment analysis with literary texts is the next phase of sentiment analysis in the digital humanities. Advanced deep learning models and vast lexicons that reflect contemporary social and moral values need to be developed for specific tasks such as detecting Dickensian sentiment.

Notes

[1] In the digital humanities field, debates over Syuzhet led humanists to discord over sentiment analysis in literature. Syuzhet was developed by Matthew Jockers and his collaborators and first released on February 22, 2015 for the sentiment analysis of literary texts.

[2] Stoic emotional terms are varied between scholars, but emotions are generally categorized into binary opposition: “good (joy, desire, wish) and bad (fear, lust, caution)” (Bonfiglio 273). Sentiment analysis is also generally based on the binary opposition.

[3] It is important to set fixed parameters to avoid random walks.

Bibliography

Banfield, Marie. "From Sentiment to Sentimentality: A Nineteenth-Century Lexicographical Search." *19: Interdisciplinary Studies in the Long Nineteenth Century*, vol. 4, 2007. doi:10.16995/ntn.459.

Bonfiglio, Richard. "Sentimental Transport and Stoic Sacrifice in A Tale of Two Cities." *Dickens Quarterly*, vol. 37, no. 3, 2020, pp. 264-284.

Carney Bethan, "Introduction: 'Mr Popular Sentiment': Dickens and Feeling." *19: Interdisciplinary Studies in the Long Nineteenth Century*, no. 14, 2012. doi:10.16995/ntn.644.

Devlin, Jacob, et al. "BERT: Pre-training of deep bidirectional transformers for language understanding." In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, vol. 1, pp. 4171-86.

Dickens, Charles. *Our Mutual Friend*. 1865. Oxford UP, 2008.

Ekman, Paul. "An argument for basic emotions." *Cognition & Emotion*, vol. 6, no. 3-4, 1992, pp. 169-200.

Eliot, George. *Middlemarch*. 1871. Penguin Classics, 2003.

Gao, Jianbo, et al. "A multiscale theory for the dynamical evolution of sentiment in novels." *International Conference on Behavioral, Economic and Socio-cultural Computing (BESC)*, 2016.

Haider, Thomas, et al. "PO-EMO: Conceptualization, Annotation, and Modeling of Aesthetic Emotions in German and English Poetry." *International Conference on Language Resources and Evaluation (LREC) 2020*, Marseille.

Haakonssen, Knud. “Introduction.” *Adam Smith: The Theory of Moral Sentiments*. 1759. Cambridge UP, 2002.

Howard, June. “What is Sentimentality?” *American Literary History*, vol. 11, no. 1, 1999, pp. 63-81. doi:10.1093/alh/11.1.63.

Kaplan, Fred. *Sacred Tears: Sentimentality in Victorian Literature*. Princeton UP, 2013.

Kwartler, T. *Text Mining in Practice with R*. John Wiley and Sons, 2017.

Lenard, Mary. “‘Mr. Popular Sentiment’: Dickens and the Gender Politics of Sentimentalism and Social Reform Literature.” *Dickens Studies Annual*, vol. 27, 1998, pp. 45-68.

Levine, George. “Introduction: George Eliot and the Art of Realism.” *The Cambridge Companion to George Eliot*, edited by George Levine. Cambridge UP, 2001, pp. 1-19.

Liu, Bing. *Sentiment analysis: Mining opinions, Sentiments, and Emotions*. Cambridge UP, 2015.

Mason, Emma, “Feeling Dickensian Feeling.” *19: Interdisciplinary Studies in the Long Nineteenth Century*, no. 4, 2007. doi:10.16995/ntn.454

McDonagh, Josephine. “The Early Novels.” *The Cambridge Companion to George Eliot*, edited by George Levine. Cambridge UP, 2001, pp. 38-56.

Medhat, W., Hassan, A. and Korashy, H. Sentiment Analysis Algorithms and Applications: A Survey. *Ain Shams Engineering Journal* vol. 5, no. 4, 2014, pp. 1093-1113. doi:10.1016/j.asej.2014.04.011.

Mohammad, Saif M. “From once upon a time to happily ever after: Tracking emotions in mail and books.” *Decision Support Systems*, vol. 53, no. 4, 2012, pp. 730-41.

Plutchik, Robert. “A general psychoevolutionary theory of emotion.” *Theories of emotion*, Academic Press, 1980, pp. 3-33.

---. *The emotions*. UP of America, 1991.

Reagan, Andrew J., et al. “The emotional arcs of stories are dominated by six basic shapes.” *EPJ Data Science*, vol. 5, no. 1, 2016, pp. 1-12.

Sánchez-Eppler, Karen. *Touching Liberty: Abolition, Feminism, and the Politics of the Body.* U of California P, 1993.

Todd, Janet M. *Sensibility: An Introduction*. London: Methuen, 1986.

CONCLUSIONS

In this article-based dissertation, I have discussed colorization and sentiment analysis by deploying computational methods in the humanities, with a focus on the deep learning fields of computer vision and natural language processing. The first chapter "*Victorian400*: Colorizing Victorian Illustrations" was published in the *International Journal of Humanities and Arts Computing* by Edinburgh University Press in October 2021. The chapter "Sentiment Analysis: Limits and Progress of the Syuzhet Package and Its Lexicons" is to be published by *Digital Humanities Quarterly*. I presented the chapter "Dickensian Sentiment and Sentiment Analysis of Victorian Novels" at the Dickens Society Symposium 2021. The chapter "Case Study: Using Machine-Colored Illustrations of Charles Dickens's Fiction in the Classroom" was accepted to the DH2020 conference with the title "Colorization of Illustrations in Charles Dickens's Novels Using Deep Learning," but the conference was canceled due to COVID-19. I believe all my dissertation chapters have contributed to the development of the digital humanities and data science by being published and/or presented at conferences, in addition to my datasets being shared with the public.

In the chapter "*Victorian400*: Colorizing Victorian Illustrations," I discussed how I created and curated the Victorian400 dataset, which consists of 400 illustrations from the nineteenth century. I also validated the Victorian400 dataset by using the pix2pix model, which is based on cGANs (conditional Generative Adversarial Networks). Since the introduction of the GAN model by Goodfellow et al. in 2014, GAN-derived models such as cGAN, StyleGAN, and SRGAN have been developed further and have influenced a variety of domains. By using the cGAN model with the Victorian400 dataset, I revealed the possible usage of deep learning models in the digital humanities. As a digital humanist, when curating the Victorian400 dataset, I chose images from the Victorian era to reflect the color palette at the time so that my deep learning models could be trained with colors from the Victorian era. I also removed images which resulted in biased outputs in terms of color. In addition, I emphasized the importance of creating humanities datasets for deep learning, since most of the current humanities datasets for deep learning have been created, curated, and shared by non-humanists. Ultimately, to create humanities datasets for deep learning, digital humanists need to deal with deep learning models, which can be a barrier to the field.

With the Victorian400 dataset, it is possible to colorize black-and-white images from the nineteenth century. Although the Victorian400 dataset is small, my experiments proved that it produces reasonable results. In addition, through data augmentation, the Victorian400 dataset can be more efficiently used for deep learning tasks. The Victorian400 dataset was created, curated, and shared for data scientists and digital humanists who create, train and test deep learning models.

In the future, the Victorian400 dataset will be tested for tasks other than colorization such as style-transfer and data augmentation. The Victorian400 dataset was originally created to colorize black-and-white illustrations from the nineteenth century, so it needs to be validated for other tasks to show that the usage of the dataset can be expanded. In addition, since new GAN-derived models have been developed by researchers, the Victorian400 dataset will continuously need to be tested with them. Through the exploratory data analysis, I showed how to create, curate, and deploy a humanities dataset, with hope that more scholars can perform similar tasks for the development of the digital humanities and data science.

In the following chapter, “Case Study: Using Machine-Colored Illustrations of Charles Dickens’s Fiction in the Classroom,” I performed a case study to see how colorization would be helpful in the classroom and how students perceive machine-colored and hand-colored illustrations when reading Victorian fiction. The results of this survey imply that most students prefer to read Victorian fiction with illustrations. Throughout the survey, contrary to our expectation that students would show resistance against machine-colored illustrations, most students had positive reactions toward them. In addition, the survey shows that students tend to prefer machine-colored and hand-colored illustrations over black-and-white illustrations. Through my case study, I found promising possibilities for using machine-colored illustrations for Victorian fiction in the classroom. However, machine-colored illustrations still have room for improvement in recognizing fine details. In addition, colorizing black-and-white illustrations might distort the original illustration. Nonetheless, there are more benefits than drawbacks when it comes to reading Victorian fiction with machine-colored illustrations. For example, machine-colored illustrations provide entertainment for modern readers, along with a better understanding of the text. Machine-coloring is cost-effective as well, unlike hand-coloring. These positive aspects reveal a promising future for machine-coloring.

Most participants in my case study expressed that they would like to see a wider variety of colors in the machine-colored illustrations, along with more details. To satisfy contemporary readers with varied and vibrant colors, I will need larger, well-curated datasets and advanced deep learning models for a colorizing task. In the future, I plan to extend the Victorian400 dataset and create a web-based tool for the colorization project so that anyone can colorize black-and-white images for free. The tool will not only be used pedagogically to help students engage with texts and provide enjoyment when reading, but also for research purposes, such as to colorize and restore black-and-white images. I believe that machine-colored illustrations will bring more enjoyment and imagination for contemporary readers, a usage which is not limited to the classroom.

The last two chapters were about sentiment analysis. In the chapter "Sentiment Analysis: Limits and Progress of the Syuzhet Package and Its Lexicons," I examined Syuzhet, which is a sentiment analysis tool for literary texts in R, in order to see its limits and progress. Due to debates over the limits of Syuzhet in the digital humanities, digital humanists have had qualms about using sentiment analysis as a research method. Through my experiments with three Victorian novels: Charles Dickens's *Our Mutual Friend*, George Eliot's *Middlemarch,* and Charlotte Brontë's *Jane Eyre*, I revealed that Syuzhet fails to detect negators, dashes, and adversative conjunctions/transitions, simply transforming tokens found in the lexicons into numerical vectors. In addition, depending on the lexicon used, the results of sentiment analysis with the same text differ due to the subjectivity of each lexicon. I also argued that the application of DCT (Discrete Cosine Transformation) when performing sentiment analysis of literary texts with Syuzhet is questionable, as the graphs generated with DCT were over-simplified, resulting in distorted results. In this chapter, through the comparison of sentiment analysis tools and lexicons and the microanalysis of literary texts with Syuzhet, I suggested that Syuzhet will continue to be questionable as a research tool in the digital humanities without improvements.

Since current research shows that deep-learning based sentiment analysis has higher accuracy than lexicon-based sentiment analysis, I performed sentiment analysis with literary texts using the BERT-Base model (12-layer, 768-hidden, 12-heads, 100M parameters), which was created and released by Google AI research in 2018. In the chapter "Dickensian Sentiment and Sentiment Analysis of Victorian Novels," I examined whether sentiment analysis could reveal Victorian sentimentality and Dickensian sentiment, whether sentiment analysis helps us understand sentiments in Victorian novels, and the definition of sentimental, sentimentality, and sentimentalism. To perform sentiment analysis with BERT, I created the VictorianLit dataset for the machine learning-based sentiment analysis of Victorian novels. The VictorianLit dataset consists of five Victorian novels: Charles Dickens's *Little Dorrit* and *Oliver Twist*, Elizabeth Gaskell's *North and South*, George Eliot's *Adam Bede*, and Mary Elizabeth Braddon's *Lady Audley's Secret*. The dataset uses five different sentimental values: 0 (very negative), 1 (negative), 2 (neutral), 3 (positive), 4 (very positive). For sentiment analysis tasks, the BERT-Base model with the VictorianLit dataset proved to be 93% accurate, along with having high precision, recall, and F1 scores for each label.

For my case study, I chose three Victorian novels—Charles Dickens's *Our Mutual Friend*, George Eliot's *Middlemarch*, and Charlotte Brontë's *Jane Eyre*. The result of sentiment analysis shows that *Our Mutual Friend* and *Jane Eyre* each have a rise-fall-rise plot, whereas the shape of the emotional valences in *Middlemarch*, which are similar to the 'Oedipus' emotional plot category, are different from *Our Mutual Friend* or *Jane Eyre*. However, while sentiment analysis helps readers understand the flow of emotions in Victorian novels, it cannot properly reflect Dickensian sentiment. Through performing sentiment analysis with BERT, I proved the possible usage of deep learning-based sentiment analysis with literary texts and revealed the role of sentiment analysis for Victorian novels.

In this dissertation, I used English datasets for each chapter. During my experiments with English datasets, I questioned if the results of computational analysis would be identical if the computational analysis was performed in languages other than English. Digital humanists have pointed out that the digital humanities are English-centric; digital tools, datasets, and tutorials are mainly based in English. This creates a barrier for those who have received all of their academic training in English; they often face difficulties when conducting computational literary analysis with languages other than English, such as having to manually create corpora and navigate varied methods of text preprocessing, along with facing differing word frequencies, lexical dispersion, and keywords. If the results are different between two languages when performing computational literary analysis, we should consider the contributing factors.

During my research, I created visualizations of word frequencies and lexical dispersion to see if computational literary analysis test results differed in languages other than English. I chose to compare the English and Korean versions of Charlotte Brontë's *Jane Eyre* (1847) and Charles Dickens's *David Copperfield* (1849-50). To perform computational experiments with the texts, I firstly needed to create text datasets in both English and Korean. I was able to download the English versions of Charles Dickens's *David Copperfield* and Charlotte Brontë's *Jane Eyre* from Gutenberg, but the Korean versions were not available publicly online due to translation copyrights. Therefore, I created the Korean datasets via OCR (Optical Character Recognition), a time-consuming method that is commonly used when performing computational literary analysis of translated texts.

Methods	**Texts**
OCR	당신의 마음에 가장 큰 부분을 차지하는 정서 나 취힝은 다 어쩌겠소 7.
Manually Cleaned	당신의 마음에 가장 큰 부분을 차지하는 정서 나 취향은 다 어쩌겠소?
PyKoSpacing	당신의 마음에 가장 큰 부분을 차지하는 정서나 취향은 다 어쩌겠소?

Table 5.1: The process of cleaning Korean texts. The Korean texts were preprocessed in order from top row to bottom row.

I first performed word frequencies with the Korean versions of each text to see how data curation influences the results of computational analysis. It revealed that the most frequent words were the book titles and '7,' since in Korean books, the book title is printed on the bottom of every (other) page, and question marks are often perceived as '7' instead of '?' by OCR. There were a number of typos due to character recognition failures. In addition, spacing between morphemes or words was often ignored or added through OCR. In order to fix the typos created through OCR, I went through every single word, comparing the OCR processed texts with the original texts, which was a repetitive and weary task. An example of this process is shown in Table 1, where "취향" (type, preference, taste) and a question mark were recognized as "취힝" and "7". After manually fixing typos such as these, I used a PyKoSpacing package to fix spacing issues generated in the process of OCR. However, texts processed with PyKoSpacing need to be double-checked. For example, the word "구레나룻" was transformed into the word "구레나 룻". To fix these issues, I added rules to the PyKoSpacing package for more accurate text processing. After completing this process, I was able to perform word frequency analyses, lexical dispersion, and topic modeling. For the English datasets, the only steps for curation were to remove unnecessary parts such as copyrights, the table of contents, and author introductions.

I created word frequency and lexical dispersion plots for the English and Korean versions of Charles Dickens's *David Copperfield* and Charlotte Brontë's *Jane Eyre*. I included stop words when analyzing word frequencies and lexical dispersion, as stop words are important in revealing how the languages bring about different results.

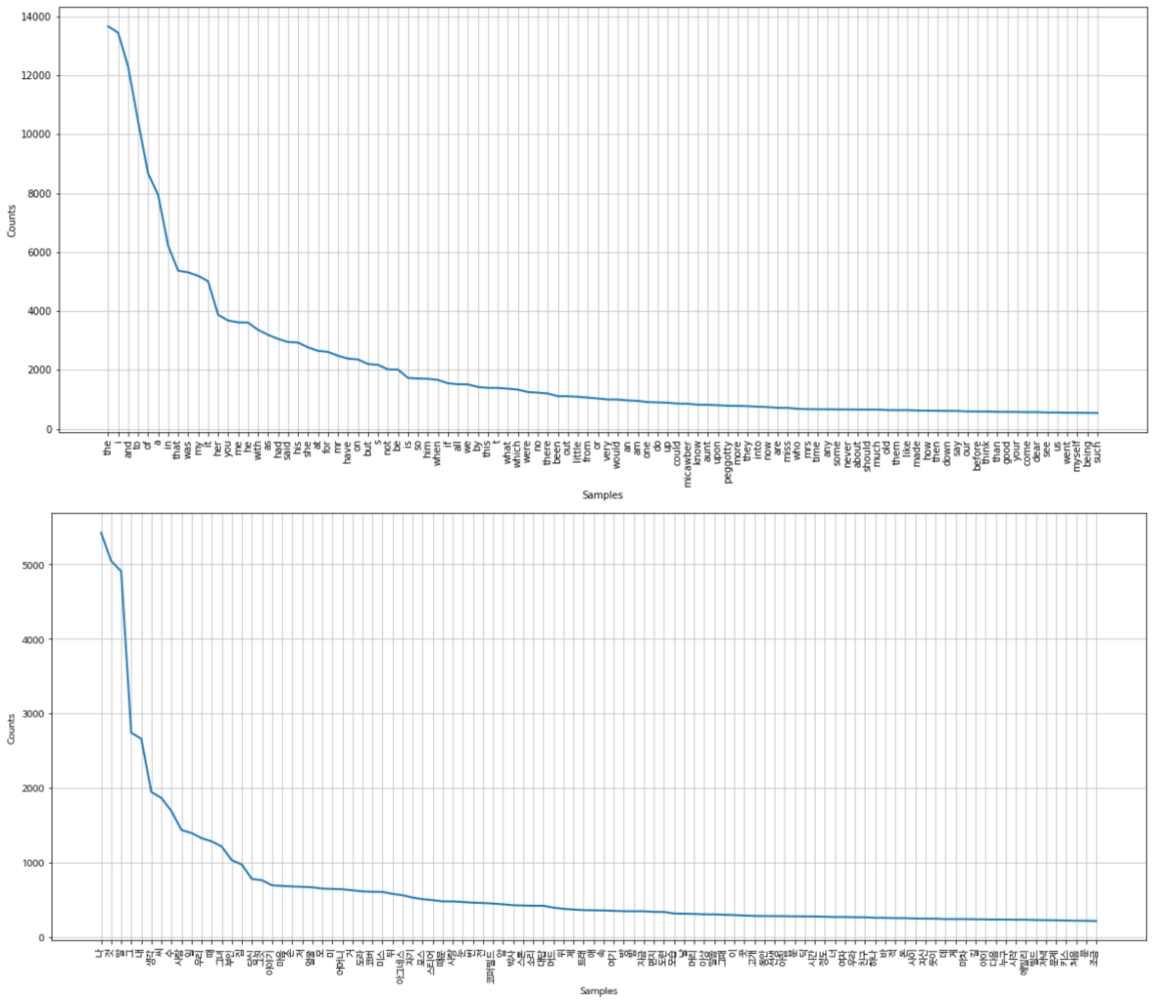
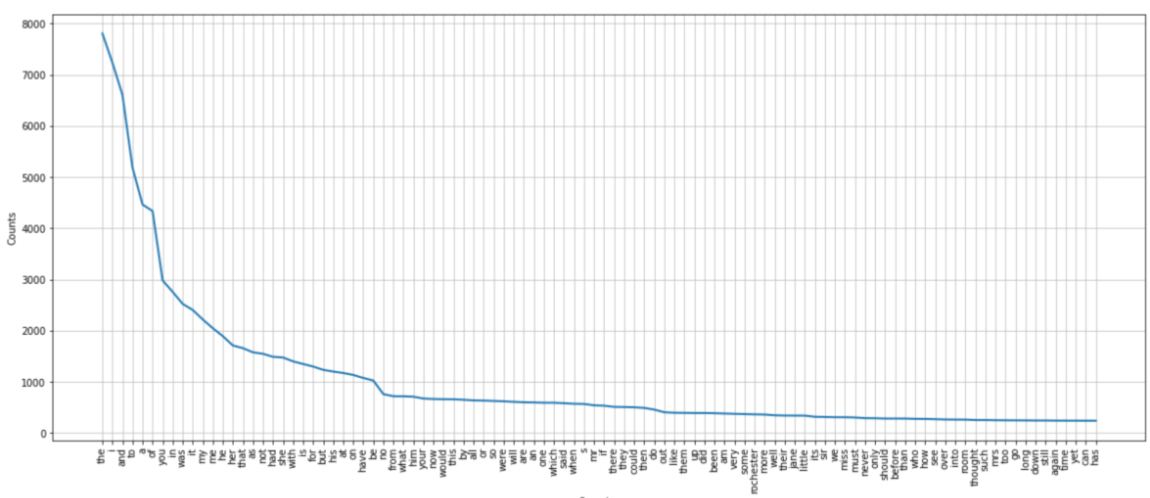


Figure 5.1: Word frequencies of the English and Korean versions of Charles Dickens's *David Copperfield*.



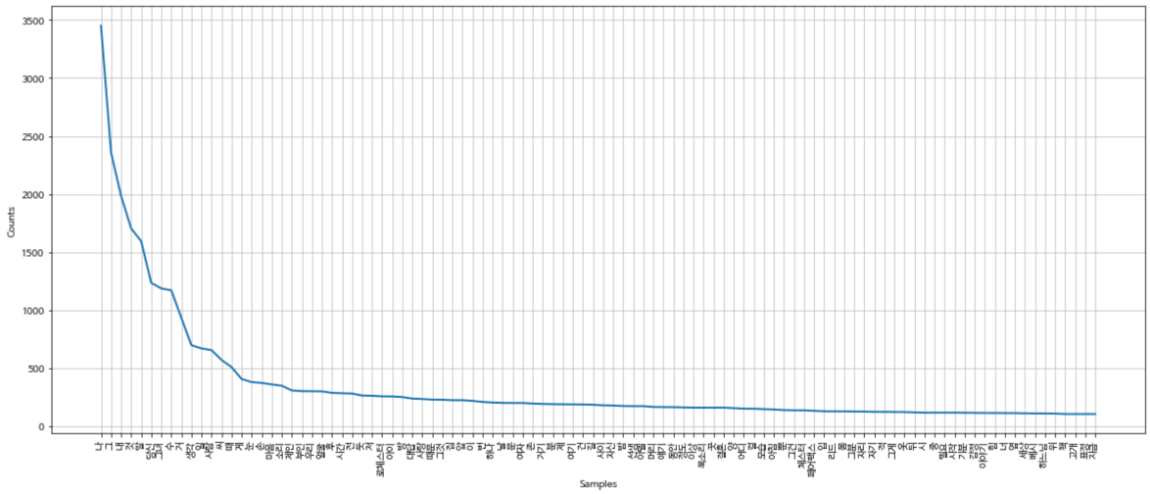
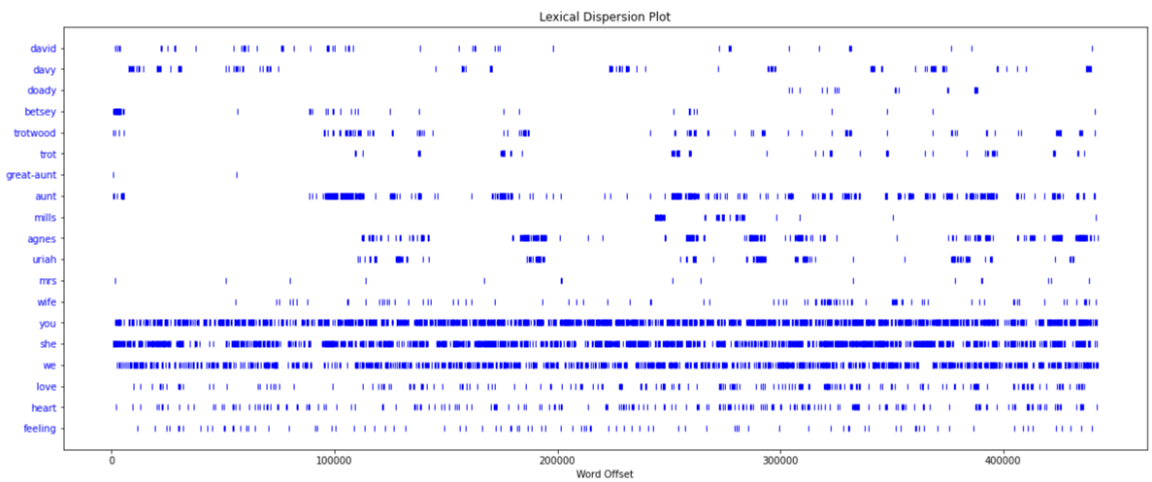


Figure 5.2: Word frequencies of the English and Korean versions of Charlotte Brontë's *Jane Eyre*.

In *David Copperfield*, the main character, David Copperfield, is called three different names depending on who calls his name: Copperfield, Davy, and Doady. In the English and Korean versions, Copperfield (코퍼필드), Davy (데이비), and Doady (도디) appeared 529 (452), 173 (129), and 26 (2) times, respectively. In the Korean version of *David Copperfield*, names are often omitted through translation. For example, in the sentence, "Peggotty's love is a great deal better than mine, Davy" (Chapter 2), the Korean version was translated without Davy. This is because, in the Korean language, it is unnatural to call someone's name at the end of a sentence. Similarly, in *Jane Eyre*, Jane (제인), Rochester (로체스터), and Fairfax (페어팩스) appeared 342 (293), 366 (359), and 137 (124) times, respectively. The translator seemed to leave out names if they would sound unnatural in Korean. However, not all names occurred less in the Korean versions. For instance, Uriah (우라이아), Steerforth (스티어포스), and Em'ly (에밀리) appeared 233 (259), 398 (500), and 203 (425) times, respectively. This is because pronouns in the English version were translated into names for clarification in Korean. Betsey Trotwood, who is one of the most important characters in *David Copperfield*, is called 'aunt' by David, but in the Korean version, Betsey is mainly called 'great-aunt'. Betsey Trotwood is mostly called 'great-aunt'

(대고모) in the Korean version of *David Copperfield*. The word ‘great-aunt’ (대고모) appears 1,092 times in Korean, while appearing only 2 times in English. Instead, Betsey or Trotwood appeared more in the English version: Betsey (88) compared to “뱃시” (74) and Trotwood (153) compared to “트롯우드” (136). Similarly, the word ‘master’ (도련님) appeared 215 (500) times, since the translator intentionally added ‘master’ for Copperfield in the Korean version, considering the culture; The words great-aunt (대고모) and ‘master’ (도련님) appeared much more in the Korean version since it is common to call people by their titles in Korean, rather than or in conjunction with their name.



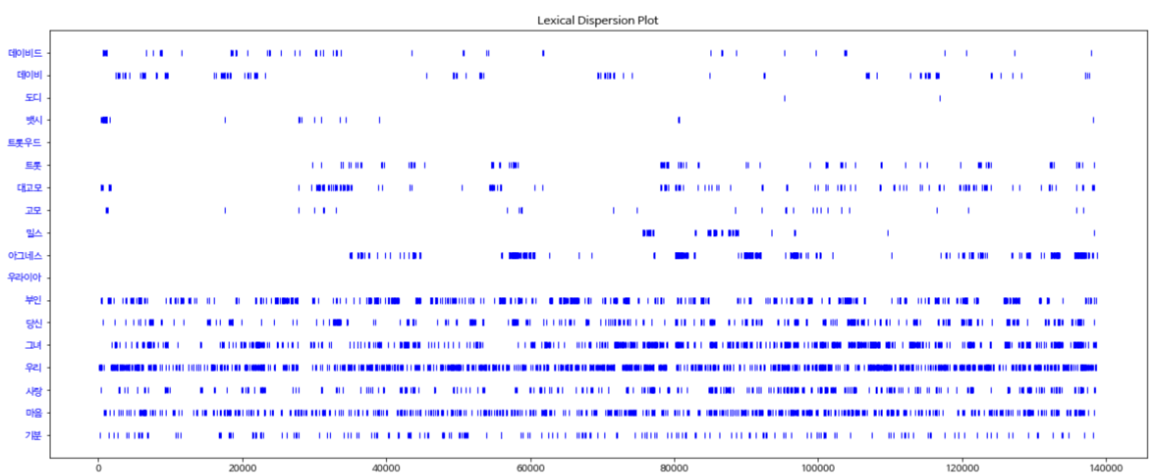
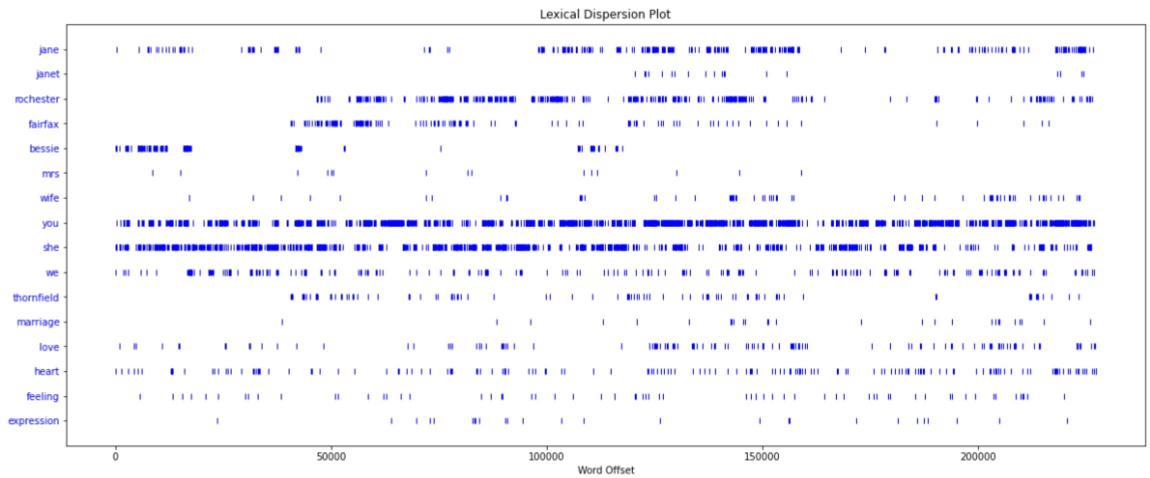


Figure 5.3: Lexical dispersion plots of the English and Korean versions of Charles Dickens's *David Copperfield*.



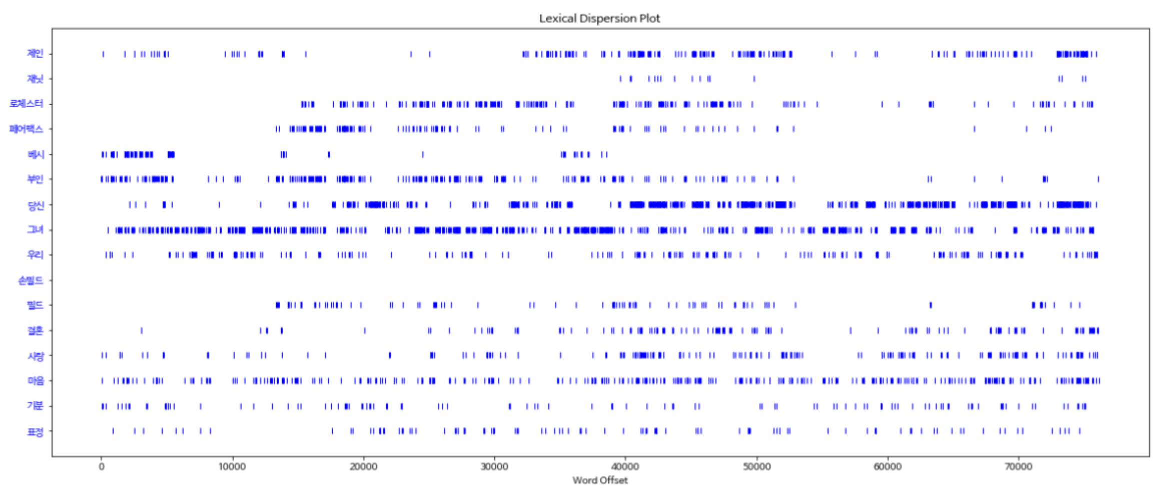


Figure 5.4: Lexical dispersion plots of the English and Korean versions of Charlotte Brontё's *Jane Eyre*.

In the lexical dispersion plots for both *David Copperfield* and *Jane Eyre*, pronouns in English such as 'you' (당신), 'she' (그녀), and 'we' (우리) appear more than in Korean. In English, sentences usually follow the subject, verb, object (SVO) order, whereas in Korean, sentences follow the subject, object, verb (SOV) order. In both languages, it is possible to omit objects, and in Korean, subjects often are left out, too, especially in spoken language. In the Korean version, those pronouns were abridged through translation. However, the word frequencies of emotional words such as ‘love’ (사랑), ‘heart’ (마음), and ‘feeling’ (기분) were greater in Korean than in English, due to word choices made by the translator. From the English version of *David Copperfield* and *Jane Eyre*, synonyms for love such as ‘cherish’ and ‘affection’ were sometimes translated into ‘love’ in Korean. Occasionally, indirect translation does not reflect the primary definition of words. For instance, the word ‘cherish’ in the sentences from *Jane Eyre*, “I cherished towards Mrs. Fairfax a thankfulness for her kindness” (chapter 12) and “he had once cherished what he called a ‘*grande passion*.’” (chapter 15), were translated into ‘appreciate’ and ‘bear’, respectively, in the Korean version. The word ‘cherish’ contains those meanings in English, but through the translation, the word ‘cherish’ was changed into three different words to convey the meaning in Korean, which resulted in different word frequencies. Translation sometimes changes the original meaning of a text to reflect cultural, social, and linguistic differences. These factors make it difficult to perform computational literary analysis in different languages, in addition to creating inconsistent results.

Cultural and linguistic differences not only make it difficult to translate between the two languages correctly, but are also a major factor in the generation of differing test results when performing computational literary analysis on translated texts. In addition, there are difficulties when performing computational literary analysis in languages other than English. When determining word frequencies, for instance, performing name entity recognition (NER) is important, but a number of English names were not recognized by Korean NLP tools. For example, Trotwood (트롯우드) was recognized as ‘Trot’ (트롯) and ‘wood’ (우드), and Murdstone (머드스톤) was processed as ‘Murd’ (머드) and ‘stone’ (스톤) in Korean, due to the nouns ‘wood’ and ‘stone’. Boyd-Graber et al. mention that personal names which overlap with general nouns such as the word ‘daisy’ can make text mining complicated, and suggest avoiding the issue by using upper cases for personal names (72). Because Korean does not have letter cases, this solution could not be implemented in my case study. Although customized code can improve the recognition of personal names, as Boyd-Graber et al. mentioned, “there is no known way to avoid careful consideration of the meaning of words in context” (72-4). Through my case studies, I introduced the challenges of performing computational literary analysis in languages other than English. When I presented this part of the work at the ACH2021 conference, many DH scholars were surprised, having never considered the potential challenges and differing test results when performing sentiment analysis in different languages. By presenting my case studies at the DH conference, I believe I raised awareness of diverse languages in the digital humanities.

These works could not be achieved without other researchers’ contributions to the digital humanities and data science fields. I referenced a variety of works such as academic articles, tutorials, and blog posts to successfully complete research tasks. For my colorization project, I deployed the pix2pix model built based on cGANs by Isola et al. to validate the Victorian400 dataset. The pix2pix model, which is a GAN-derived model, would not have been invented without generative adversarial networks, which were first introduced by Goodfellow et al. in 2014. For my experiment with deep learning-based sentiment analysis, I used BERT, which was first introduced by Google AI research in 2018. The introduction and development of advanced deep learning models such as GANs and BERT opened new avenues for a variety of generative tasks such as colorization, style-transfer, and sentiment analysis in broad fields such as the digital humanities, data science, and computer science. I believe this dissertation contributes to the development of the digital humanities and data science fields through exploratory data analysis, pedagogical case studies, and experiments with humanities datasets. I finish off the last chapter of this dissertation with hope to see more humanities datasets with exploratory data analysis for the continual development of the digital humanities and data science.

Bibliography

Goodfellow, Ian, et al. "Generative Adversarial Nets." *Advances in Neural Information Processing Systems*, 2014. papers.nips.cc/paper/5423-generative-adversarial-nets.pdf.

Isola, Phillip, et al. "Image-to-Image Translation with Conditional Adversarial Networks." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017. openaccess.thecvf.com/content_cvpr_2017/papers/Isola_Image-To-Image_Translation_With_CVPR_2017_paper.pdf.