The Impact of Lyrica on Writing Quality

Arielle Dortch
Beth Young
Lauren Foltz

1 Abstract

As a professional writer, Hadena James was worried that the medication she was taking was impacting her ability to produce novels. She reported feeling sluggish and "stupid," while still trying to continue her writing career. Hadena wasn't able to publish as many books in the year she was taking Lyrica and she was concerned that the quality of her writing decreased.

Our hypothesis is that when an author is on Lyrica, her writing style changes. We will use Flesch-Kincaid to determine grade level readability scores. We will apply Term Frequency to examine the vocabulary size of each book. We will also use MALLET to discover central book themes and a theme for the overall series. Using LIWC, we will explore a broad range of variables in order to investigate the author's writing patterns. Finally, we examine the Amazon star ratings and will conduct sentiment analysis on the book reviews to see if there is a general decline in the popularity of the books that were written while on Lyrica.

2 Data Set

Our primary data set consists of 14 novels written by Hadena James. Of the books, two of them were written while she was on the prescription drug, Lyrica. The "on Lyrica" books are in bold in the following chart. The author provided us with Word documents of all the novels, which we then converted to plain text documents. Additionally, Helium 10 software [1] was used to obtain Amazon reviews for the books, which were stored in .csv files, separated by book.

Books Title	**Published date**
Tortured Dreams	Feb, 2013
Elysium Dreams	Nov, 2013
Mercurial Dreams	Dec, 2013
Explosive Dreams	May, 2014
Cannibal Dreams	Aug, 2014
Butchered Dreams	Nov, 2014
Summoned Dreams	Feb, 2015
Battered Dreams	Jun, 2015
Belladonna Dreams	Sept, 2015
Mutilated Dreams	Feb, 2016
Fortified Dreams	June, 2016
Flawless Dreams	May, 2017
Demonic Dreams	Apr, 2018
Ritual Dreams	Mar, 2019

3 Methods

3.1 Flesch-Kincaid Grade Level Scoring

One of the tests we used was to view the Flesch-Kincaid Grade Level score for each book, as computed by Microsoft Word. Flesch-Kincaid scores are based on two factors; the average sentence length and the average number of syllables contained in a word. Sentences that contain a lot of words and words that contain a lot of syllables are judged harder to understand and thus require a higher reading ability or grade level. In general, the earlier books were written with easier-to-understand language, measuring an average of 4.5 grade level. As the author has progressed in her career, the reading level has been increasing. Even the two books written on Lyrica have had an increase in the Flesch-Kincaid Grade Level.

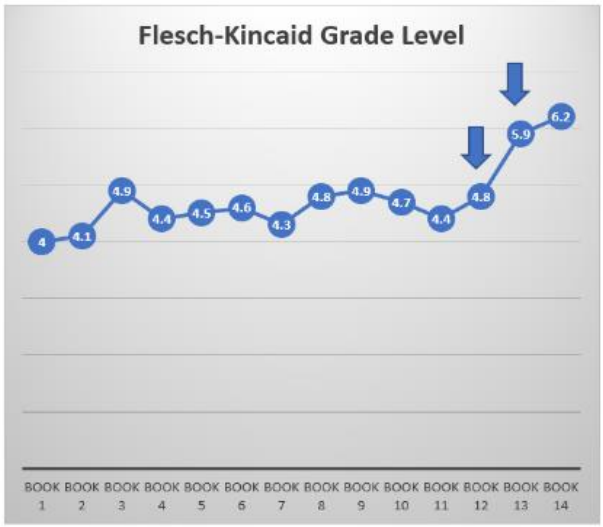

Figure 1

3.2 MALLET Topic Modeling

MALLET stands for Machine Learning for Language Toolkit. The MALLET Topic Modeling toolkit contains an implementation of Latent Dirichlet Allocation (LDA) which we used to analyze the 14 books in our corpus, looking for latent topics in each book. LDA tokenizes the words of each document and then classifies the text in each document and assigns it to a particular topic. Each topic is a cluster of words that frequently occur together [2]. We processed the files using bigrams and selected 15 topics. Each book was identifiable from the central theme of the story (Figure 2).

For example, Explosive Dreams had the main topic of explosions at carnivals and county fairs. MALLET picked out topics for that book as “fair ride plague bomber queen chapter rides fairs carnival fair_queen prairie_dog prairie explosion barb bomb funnel homeland graft cooler cake.” The topics and words associated with them are listed in Figure 3. The topic associated with column 8 in Figure 2 could be considered the overall series topics as it includes such words as “serial killer” and “body” which would appear in each book.

Books	Topics														
	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15
Tortured Dreams	0.00	0.00	0.00	0.00	0.00	0.45	0.03	0.52	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Elysium Dreams	0.00	0.00	0.00	0.00	0.00	0.00	0.41	0.56	0.00	0.00	0.00	0.00	0.00	0.00	0.03
Mercurial Dreams	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.52	0.00	0.00	0.05	0.00	0.00	0.00	0.43
Explosive Dreams	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.54	0.00	0.00	0.44	0.00	0.00	0.00	0.02
Cannibal Dreams	0.42	0.00	0.01	0.00	0.00	0.00	0.00	0.55	0.02	0.00	0.00	0.00	0.00	0.00	0.00
Butchered Dreams	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.57	0.42	0.00	0.00	0.00	0.00	0.00	0.00
Summoned Dreams	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.55	0.00	0.44	0.00	0.00	0.00	0.00	0.00
Battered Dreams	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.53	0.00	0.00	0.00	0.00	0.47	0.00	0.00
Belladonna Dreams	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.55	0.00	0.00	0.00	0.45	0.00	0.00	0.00
Mutilated Dreams	0.00	0.42	0.00	0.00	0.00	0.00	0.00	0.56	0.00	0.01	0.00	0.02	0.00	0.00	0.00
Fortified Dreams	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.54	0.00	0.00	0.00	0.00	0.00	0.46	0.00
Flawless Dreams	0.00	0.00	0.42	0.01	0.00	0.00	0.00	0.57	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Demonic Dreams	0.00	0.00	0.00	0.47	0.00	0.00	0.00	0.53	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Ritual Dreams	0.00	0.00	0.00	0.00	0.51	0.00	0.00	0.49	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Figure 2

T1	0.01484	family jaguar animal lee unger cousin columbia jaguars bite tennyson joe meat grandfather animals great-aunt lila orangutan bites anderson snow
T2	0.03663	orleans voodoo magic gris contract tattoos apartment russian french_quarter balcony black_magic migraine angeles quarter aphasia los french los_angeles gym gris_gris
T3	0.05645	bones farm insect insects sword beetles apex market paso parking_lot plane wendigo momma goat super_psychopath chickens australia display claymore insect_farm
T4	0.01491	didn't bunker apex brother wasn't director phone don't helicopter chapter he'd maine leg mom i'm couldn't morphine gabriel's director_fbi lake
T5	0.00714	acid satanists personality aunt personalities cult row church religion fourth kids murders mom carr merrily religions satanist evil temple husband
T6	0.0071	torture maidens iron scavenger chapter maiden scavenger_daughter bull rats teeth spikes frowned impaling daughter apartment iron_maidens museum ten drawing drawing_quartering
T7	0.01692	snow grace gentry sheriff rope knife alaska agent pine-sol agent_gentry cigar tree wife reporter winch ash motel cameraman bell arrow
T8	0.99966	didn asked told back wasn time people serial killer looked good answered room made head found body thought make door
T9	0.02433	cane chub cattle okafor arby granddaughter columbia bubbly freon motel ufo captain kindle tape jaw bartender teacher pedophile sketch ink
T10	0.03013	detroit prostitutes bell suspect church green nipples gang dsi gangs lingon incense burned pimps campbell streets rain kayla sage gooder
T11	0.01695	fair ride plague bomber queen chapter rides fairs carnival fair_queen prairie_dog prairie explosion barb bomb funnel homeland graft cooler cake
T12	0.02284	belladonna south_dakota dakota south tall_man maya lsd tall bar drinks sioux_falls sioux crowd bottles falls tribal braun gui poison sweet
T13	0.0071	san rats texas blanks san_marcos marcos baton werewolf nails young teens san_antonio antonio austin party teen krokodil sixteen bat infected
T14	0.0071	lazar fortress marshals cafeteria killers apex prison parsons dominic ward floor demetrius inmates warden bomb secure doors timmons yuri bombs
T15	0.02411	mercury desert vegas pool salt wife summers homeless bodies las_vegas playa las valley death_valley mummies riggings painting prostitutes artist day_laborers

Figure 3

3.3 Term Frequency

Term Frequency, or the counting of unique words in a document, is a way of seeing how language used in the books may have changed over time. The author reported feeling "stupid," and as if her brain wasn't operating at full capacity. We wondered if that would show in the words she chose while writing; would her vocabulary decrease in the books written while "on Lyrica." We tokenized the books and looked at the number of unique words used in each book. The average amount of unique words seemed to be in the 6,000-7,000 range.

There was a decline for books 12 and 13. For those 2 books, the average was just shy of 6,000 words. The amount of unique words went back to the normal range for book 14.

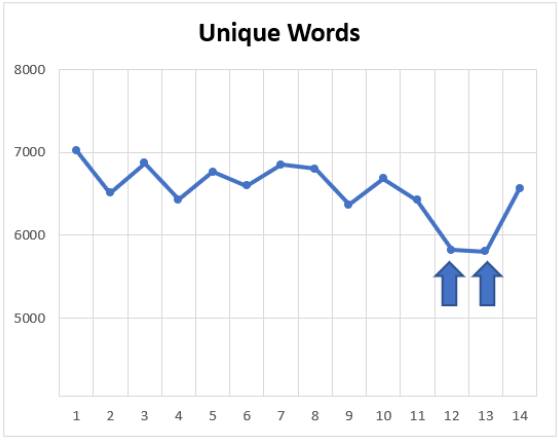

Figure 4

3.4 LIWC Analysis on Books

Linguistic Inquiry and Word Count, or LIWC, is used for text analysis to provide a broad range of social and psychological insights [3]. A LIWC analysis was run on all 14 books, and all 93 variables were examined for trends. We looked for trends for books 12 & 13, to see if those books showed a different writing style.

We found that books 12 & 13 were low in Analytical Thinking but the values bounced back up in book 14. The Analytical thinking variable captures the degree to which people use words that suggest formal, logical, and hierarchical thinking patterns. Those high in Analytical Thinking perform better in college and have higher college board scores [4].

We also found that books 12 & 13 were low in article usage, but the values bounced back up in book 14. Additionally, these books were high in usage of two cognitive processes: discrepancy (should, would) and tentative (maybe, perhaps), and this value bounced back down in book 14.

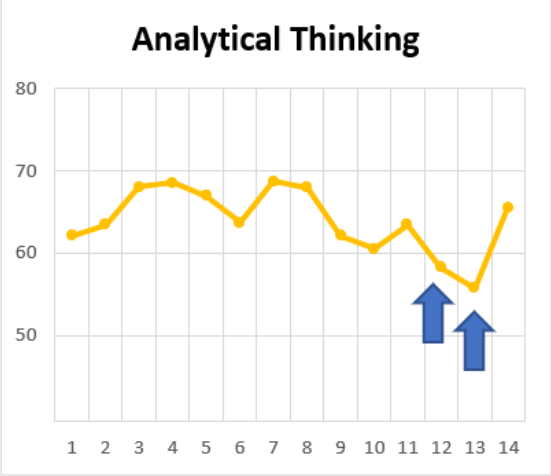

Figure 5

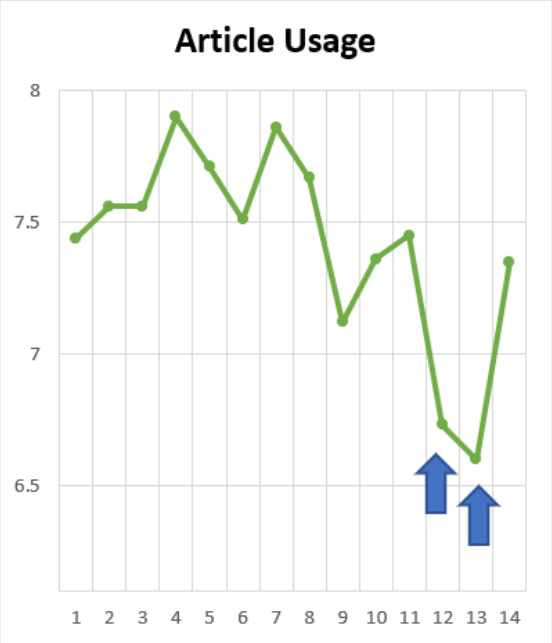

Figure 6

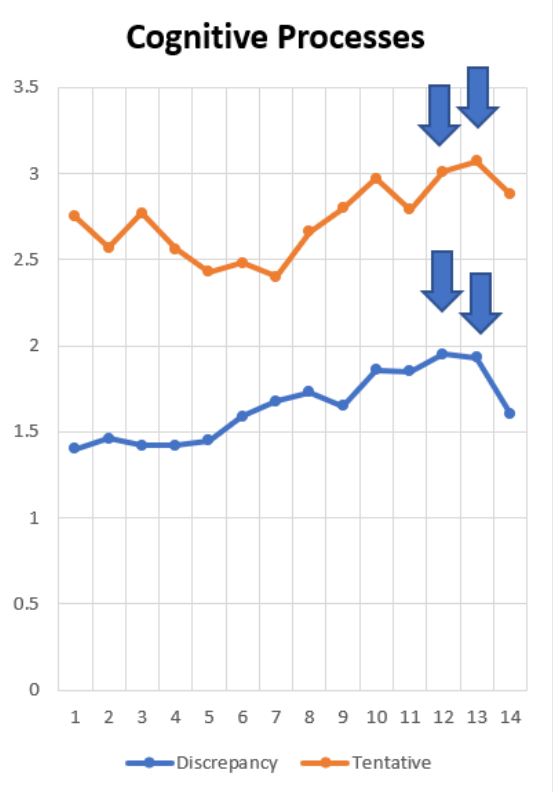

Figure 7

3.5 Amazon Star Ratings

Amazon Star ratings were documented and plotted. Interestingly, the overall star rating is not straightforward, as Amazon calculates the product's star ratings based on a machine learning model instead of a raw data average. The model takes into account factors such as age of the review, whether it is associated with a verified purchase, and other undisclosed factors. Out of curiosity, we calculated the raw overall star ratings, and found that the weighting that Amazon performed did not skew the raw data significantly.

After plotting the overall Amazon Star ratings for this series, we found that they increased over time in general, but books 12 & 13 have ratings that are below the average rating of 4.8 stars.

Figure 8

3.6 Sentiment Analysis on Reviews

LIWC was run on the Amazon Reviews, separated by book, to discover the review tone. Most books were high in positive tone, in the range of 83 to 91 (book 1 and 5 were outliers). Book 12 reviews were mid-range (87) and book 13 had the highest positive tone (91).

It was curious to see that reviews for book 5 had such a drop in positive tone, since it received only 4- and 5-star ratings. We followed up by doing a close reading of the reviews for book 5, "Cannibal Dreams," as well as running the reviews through a word cloud program [5] to see a list of word frequencies. It appears that the readers had an enjoyable experience with the book, but also described the content with words such as "gruesome" and "disturbing," which impacted LIWC tone results.

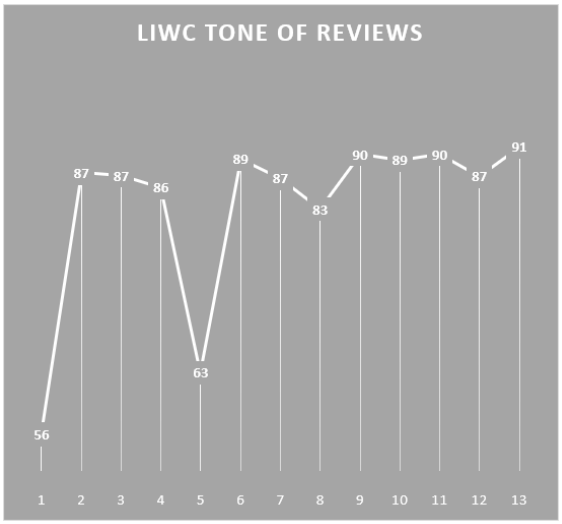

Figure 9

4 Conclusion

The LIWC analysis of books suggests that while on Lyrica, this author had a change in writing style, as evidenced by the decrease in analytical thinking, decrease in article use, and increase in the cognitive processes of discrepancy and tentative. The Term Frequency analysis suggests that vocabulary size decreased, yet the Flesch-Kincaid analysis suggested that writing complexity increased. Therefore, although it appears that Lyrica influenced this author's writing style, we can't definitively say that it led to a decrease in quality.

The Amazon star ratings, LIWC tone analysis of reviews, and Flesch-Kincaid analysis all indicate that this author's writing has improved over time.

The MALLET LDA analysis suggests that while all books share the topic of serial killers, each book does have a distinct topic.

5 Limitations and Further Work

The number of reviews per book went down over time, and book 14 has not yet been released, resulting in no reviews for this book. It might be interesting to analyze the book reviews again once book 14 has received some reviews. Some reviews for books 12 and 13 mentioned the presence of a large number of errors, but it is difficult to know if this is due to a change in editor or a drop in writing quality. Perhaps an analysis of pre-edited books could answer this question.

Ideas for future work include running K-means on books to see if the documents cluster into 2 clusters ("on Lyrica" and "off Lyrica") and running Naive Bayes on reviews to see if we can predict if a review goes with an "on Lyrica" or "off Lyrica" book.

6 References

[1] https://www.helium10.com/
[2] http://mallet.cs.umass.edu/topics.php
[3] https://liwc.wpengine.com/
[4] https://liwc.wpengine.com/interpreting-liwc-output/
[5] https://www.wordclouds.com/