

Input Output: The Differences between Hysteria Experienced and Observed

Hysteria is an affliction of the self—a manifestation the self's desire to manifest its will over its immediate environs, both physically and socially. Hysteria is characterized by internalized selfishness, and the ecstasy and revulsion that selfishness elicits. Trapped within a crystalline lattice, the hysteric is encased entirely, a body entombed in glass, a perfectly clear distortion on display for all to see. A static performance—an exhibit. Exhibiting, what? – symptoms perhaps? Or herself? Display herself, perform herself? Is it for the outsiders, to those who embrace it and consume it? Those who rewrite it and dissect it, those who preserve a perfectly clear (clinical, without biases, with high DPI, color fidelity!) taste of hysteria, who view the performance, they exhibit, they perform—they absorb and emit hysteria himself.

Hysteria is an affliction of the self—a manifestation of selfishness—but to whom does it belong? The classical hysteria—a cure-all for doctors unable or unwilling to diagnose female ailments in the past—also served the only avenue of self-expression and denial of repression for many of the women who were affected by those same ailments. There is a fundamental gap between the understanding of the expression of hysteria when comparing those who characterize hysteria internally—the women with whom hysteria has been historically associated, and those who characterize hysteria externally—the people who find the former group dangerously inexplicable. Hysteria is an affliction of unfairness, a Boolean paradigm that divides those who are crystallized within the lattice of societal repression and those who find hysteria to be self-indulgent, masturbatory, and ultimately unfair to the ones who must account for the hysteric's actions. The concept of a “pure hysteria” is a flawed concept for assuming that such a deeply personal and environmentally-dependent sensation can be distilled down to a readily-consumed concept for the masses—yet it has been attempted again and again, from the wandering womb to

modern psychiatry. However, that does not mean that hysteria is entirely nebulous—after all, we oft see it translated in books, poetry, music, and other arts. By examining specific examples of hysteric literature, as well as a wider corpus of “hysterical literature” separated by those primarily categorized by internal and external hysteria, I wish to identify some of the characteristics that characterize and differentiate internal and external hysteria. Hysteria has become irrevocably linked with certain characteristics and tropes over time, and through an examination of a corpus of hysteric media, we can see the evolution of these tropes alongside cultural perceptions.

Charcot once lamented of hysteria, “Theory is good; but it doesn’t prevent things from existing.” In his pursuit of the true medical nature of hysteria, he sought to make hysteria, and in a sense, femininity itself, palatable by casting it in terms of medical theory. External perception of hysteria as a whole is characterized by the desire to cast the uncertain and inexplicable in terms of the certain. However, this desire is not marked by a hope for understanding or empathy; rather, it is a desire for control—to dominate female voices such that they are only able to echo the desires of a patriarchal society. This dynamic has been present through all of society’s history, and notably railed against in Audre Lorde’s essay “The Master’s Tools will Never Dismantle the Master’s House”. In it, Lorde decries the exclusion of minority groups from the discussion of mainstream feminism and identifies the dependence of wealthy white feminism on patriarchal structure as the cause. She argues that the desire to cast feminism in the form of gender equality as something desirable and palatable simply reinforces the patriarchy and continues to promote an external and detached view of feminism by making it about the audience. Lorde’s casting of the general populace as the external observers is indicative of the externally characterized ways in which media depicts women and hysteria. Interactions with

hysteria and hysterics when depicted externally are commonly characterized by disbelief, anger, and abandonment (often in that order). This template is apparent in several notable works, including *Wide Sargasso Sea* by Jean Rhys and *Near to the Wild Heart* by Clarice Lispector, among others. In *Wide Sargasso Sea*, the unnamed Rochester marries Antoinette and upon finding himself unable to understand or fit in with her culture summarily scorns and confines her. Antoinette's quality of being inexplicable and unable to be possessed is what so enrages Rochester; he admits as much when arguing with her about Daniel's letter:

‘Of course I will listen, of course we can talk now, if that's what you wish.’ But the feeling of something unknown and hostile was very strong. ‘I feel very much a stranger here,’ I said. ‘I feel that this place is my enemy and on your side.’ (Rhys)

Because Rochester is unable to control the land and people of Jamaica—a terrifyingly unfamiliar feeling for him--nor Antoinette while she resides there, he chooses to control her through other methods such as denial of physical gratification, choosing to call her Bertha rather than by her preferred name, and by disregarding her misgivings by claiming that she is entirely unreasonable. Ultimately, he chooses to abandon his relationship with Antoinette by outright ignoring her and physically confining her to an attic in a different country. The same story plays out in *Near to the Wild Heart*, where Otavio finds himself unable to control or even understand Joanna, ultimately choosing to pursue what is most familiar to him. Indeed, this story plays out the same way in every time period.

Internal exhibitions of hysteria are similar to those of external hysteria in many aspects. Both are characterized by a desire to exert control, as a struggle against uncertainty. However, internal hysteria is also defined by the quality of isolation. Hysteria is fundamentally an isolating condition because it manifests as rebellion against the majority. This departure from the blueprint

of the acceptable woman isolates hysterics from society by caricaturizing them as madwomen. Their hysteria is also a condition of uncertainty--unable to follow the script that society has dictated for them, these hysteric women are left to fend for themselves. When so much of their agency has been suppressed, their desire to have some measure of control over their own lives manifests itself as their hysterical symptoms. These symptoms range broadly from the hysteric rigor of the women in the Salpêtrière to the madness suffered by Antoinette, but fundamentally, all demonstrate a rejection of male control.

To explore the question of whether there are literary characteristics that distinguish hysteria, and if there are specific linguistic traits that can give insight into the difference between external and internal hysteria, I decided to conduct several data-based analyses and interpret their results. In order to conduct these analyses, the Natural Language Toolkit was used in conjunction with various corpora including the Brown Corpus, Corpus of Contemporary American English, and a self-constructed corpus consisting of various hysteric works¹. By using these tools, I hope to identify the distinguishing characteristics of the two different types of hysteric literature and explore preconceived cultural notions about hysteria.

To begin, an in-depth analysis of *Near to the Wild Heart* was accomplished in order to better understand the different elements of internal and external hysteria (Lispector, 1990). *Near to the Wild Heart* was chosen because there are switches in narrative voice throughout the novel. By taking note of the differences between these different sections and keeping cognizant that all samples are drawn from the same book, we can conclude that there is a higher chance any difference in style is not instead explained by a different factor such as time period or writing style.

¹ See Appendix 2 for details

Initially, a more qualitative analysis was carried out—first filtering out common stopwords, then generating a word frequency table. From the most frequently occurring words, I selected the twenty most frequent words and plotted them in a lexical dispersion plot.

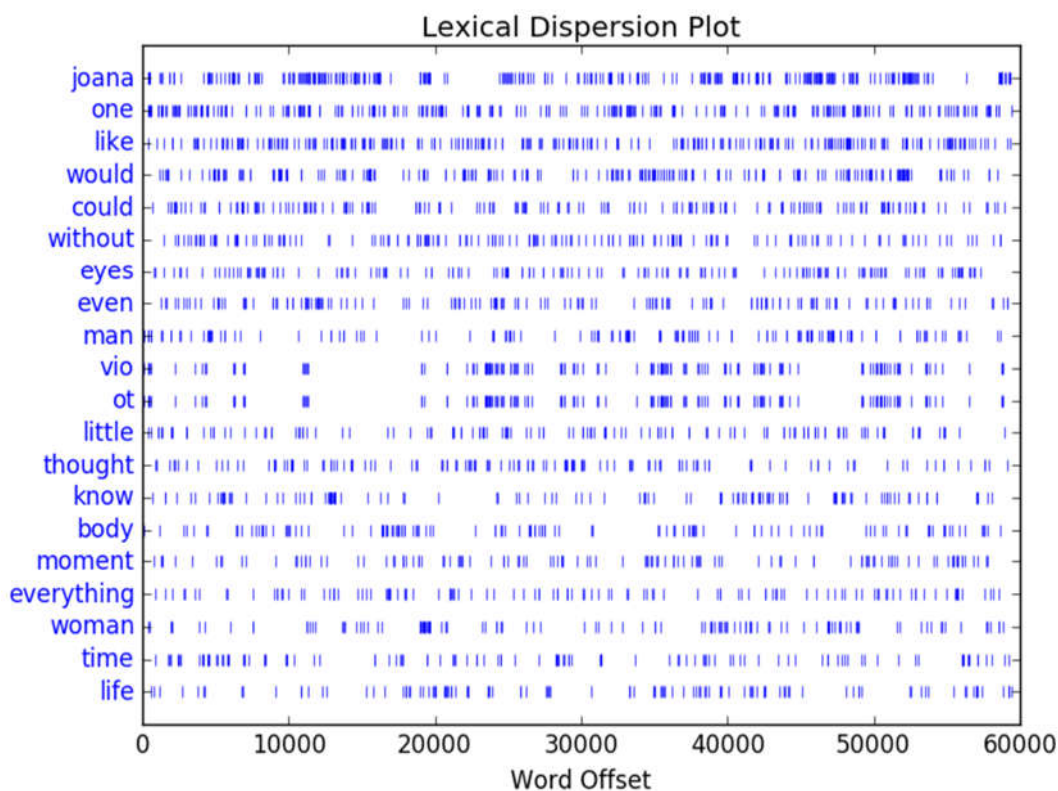


Figure 1: *Near to the Wild Heart* Lexical Dispersion²

From the figure, several interesting observations can be made. “Joana” is the most frequently occurring word. Considering that the chapters from Joana’s perspective still refer to her in the third-person, it makes sense for her name to be the most common word. Therefore, it is implied that the areas with the greatest density of “Joana” are also ones which deal with her narrative—an assumption which I physically corroborated with the novel. Among the other words, “know,” “body,” and “time” are most closely correlated with “Joana,” each yielding a

² Due to splitting on non-alphanumerics, “ot” and “vio” represent instances of “otávio”.

Pearson correlation coefficient of 0.178, 0.323, and 0.275 respectively³. These three terms all seem quite introspective, though it is difficult to draw conclusions from the analysis of one text. However, the results seem to support the idea that internal hysteria is often characterized by introspection and questioning.

We can further explore this idea by analysing the character relationships within the book. On a qualitative level, it does seem as if Joana is isolated to a large degree—she is unable to empathize with any of the characters in the book except for, arguably, Lidia. A character coappearance chart was constructed according to the examples in *The Stanford GraphBase* (Knuth, 1994)⁴. In addition to the standard values in the character array, an additional parameter was added to represent whether the coappearance was negative (-1), positive (1), or neutral (0). Negative values were defined by interactions that caused the distance between characters to increase, while positive interactions were attached to encounters that brought the two characters closer together. Ultimately, the average of the relationship variable was -0.3911, demonstrating that an overwhelming portion of Joana's interactions with other characters distanced herself from them. This supports the hypothesis that a recurring theme of isolation appears in hysteric works, but again, the results from just one book cannot be extrapolated to the entirety of hysteric literature.

In order to better understand the differences between the different corpora, a variety of parameters were tested. The calculated lexical diversity for *Wide Sargasso Sea* is 0.0997197845529, compared to 0.1358194 for the entirety of the fiction genre in the Brown Corpus. However, one sample work is insufficient to draw a statistically significant conclusion,

³ In order to generate relevant results for the correlation test, words were binned into 3000 word offset bins

⁴ I attempted the creation of a force-directed graph, but found that the results were not representative of the source material because only coappearances were taken into account.

so I compared the entire internal and external hysteria corpora. The average lexical diversity for the internal hysteria corpus is 0.1522793 and 0.1197564 for the external hysteria corpus.

Considering the small size of each of the collections of works pertaining to hysteria, the variation observed can be explained as a function of randomness.

Selfishness is often attributed to hysteria and similar conditions—but is this a fabrication caused by external interpretation or an actual quality of hysteria? To answer this question, I compared the number of personal pronouns “I, me, etc.” to the number of verbs. The reasoning behind such a methodology is that selfishness can be correlated to a large degree with self-centeredness, measured as a function of personal pronouns to verb ratio. In natural language processing, verb-noun ratio is considered one of the four most important classifiers for genre identification, thus, the belief is that personal pronoun-verb ratio will also be representative for a collection of works (Su et al., p. 311). The results were striking—0.1453846 for internal, and 0.2961953 for external⁵. This massive difference in ratios suggests that rather than internal hysteria being characterized by selfishness, external observers may actually be more self-centered in their writing. If so, then the selfishness so often attributed to a hysteric can instead be attributed to the egoism of the external observer—absolving her of a long-standing charge that she was never guilty of in the first place.

Developing classifiers to identify hysteric literature allows us to create models that can also be used to leverage our understanding of culturally depicted hysteria--serving as a foundation for discussion on our perception of both the condition and the word itself. Indeed, it allows us to address the question: “Is there such a thing as hysteric literature?” Three machine

⁵ However, this number must be taken with a few caveats. The total dataset is small, relative to the complete body of works. Additionally, there may be genre factors in play (ie. Freud uses a large amount of personal pronouns relative to verbs because he is giving his opinion).

learning mechanisms were employed: stochastic gradient descent for n-dimensional regression, a Naïve Bayes classifier, and randomized forest. The training data set was the corpus of hysteric works. The training accuracy and cross-validation score for 100 cross-validations were as follows:

Table 1: Training/Cross-Validation Accuracy for Classifiers

	SGD	Naïve Bayes	Random Forest
Training Accuracy	65%	67%	73%
Cross Validation Accuracy	55%	55%	64%

The randomized forest classifier was by far the most successful out of the machine learning algorithms, probably due in large part to not needing input parameters. Yielding an accuracy of 64% is already fairly impressive, and suggests that there is some quality about hysteric literature that distinguishes it from other works⁶.

Constructing case-insensitive 2-grams from the Corpus of Contemporary American English (COCA), we can generalize our findings to American literature as a whole. The selected databases for the purposes of this project are all case-insensitive and part-of-speech tagged and range from 2-grams to 5-grams. These sources were selected to address the following concerns: excess noise, over-specificity, researcher bias, and runtime concerns. In order to prevent excess noise from crowding out signal in the data, the datasets are filtered for terms which are too sparse in the raw COCA texts. Additionally, part-of-speech tagging allows n-grams to be filtered for specific sequences of parts of speech, allowing for a fairer comparison between corresponding search terms (ie. Male, female). The selection of low-length n-grams from two to five words in length prevents highly specific terms from impacting the model. Additionally, n-grams can be

⁶ However, due to the time constraints of the project, and the “black-box” quality of the randomized tree algorithm, it is extremely difficult to ascertain what those qualities are.

construed as basic linguistic constructions and therefore should be less influenced by the literary conventions of any certain time or author. Because these expressions are of such short length, the relative frequency of different n-grams should better express the popularity and concept of the idea the expression is rooted upon. Controlling for n-gram length also limits researcher bias when selecting interesting or representative results from the n-gram dataset while simultaneously preventing the computational demands of generating and sorting n-grams with higher length.

With respect to hysteria, n-grams are only able to express the popularity of certain word sequences within specific corpora. However, they are still able to offer some insight into the prevalence of ideas that may correspond with phrases—by noting the most common n-grams, we can grasp the most common relations and descriptors attached to a given word.

From the 2-gram dataset, I selected the most common word pairs with gender-identifying nouns as the second word (woman, women, man, men), then sorted resulting data by number of instances within the corpora (Appendix 1). From the sorted 2-grams divided by gender, the most common male descriptors are “young”, “old”, “black”, and “white”, in that order (Appendix 1a). In contrast, the most common female descriptors are “young”, “old”, “other”, and “another” (Appendix 1b). Though the first two descriptors are the same between the two sets of n-grams, the third and fourth most prevalent descriptors differ greatly. For women, these descriptors quickly become relational—rather than adjectives describing the woman herself, they define her in relation to another entity. Within the hundred most common descriptors for each group, the number of 2-grams with an appearance or attractiveness based descriptor also reflects the degree to which such traits are inseparable from the character’s identity—traits that differ greatly by sex. For the most popular descriptor pertaining to attractiveness for each sex, “beautiful woman” appears within the data 1037 times and “beautiful women” appears 498 times while “handsome

man” appears only 436 times and “handsome men” appears only 42 times. This observation alone tells us that measures of attractiveness are both stated and observed far more for women in contemporary literature than for men. Indeed, attractiveness also inherently implies a relational aspect—if one is described as attractive, it must be by or to another entity.

The idea that women are measured by their relationships with other characters far more than by their individual identity is further corroborated by the prevalence of relational descriptors for women and men. In order to measure this statistic, the percentage of relational descriptors among the total number of descriptors with a number of recorded instances above a hundred was used as a representative number. The statistic in men was recorded at 9.27% (28 out of 302), while the same statistic in women was recorded at 21.4% (53 out of 248). When extended to the percentage of instances recorded, the above statistic is even more striking: 13.1% for men compared to 30.7% for women. Out of all the instances of one-word descriptors for women in the corpus, nearly a third of them are describing women in relation to another entity. This aspect of being “owned”—of being cast in a supporting role in her own life—is the type of societal repression that seeks to police female bodies and female minds.

The same analysis for possessives yielded a reversed result. “my woman” and “his woman” appeared 176 and 93 times respectively, whereas “my man” and “her man” appeared a total of 634 and 439 times respectively. These results are indicative of the paradigm under which women are always a needy party—she must assert a relationship with a man in order to gain social approval. In contrast, the relatively low number of possessives used on women (primarily by men) indicates that the male characters do not feel that they need to verbally assert their relationship with a woman when there is already an unspoken dynamic of control.

Imagine if someone today claimed cerebral palsy or diabetes in women is caused by a wandering womb that floats throughout the body. Such an idea seems beyond preposterous. Even Charcot, who practiced less than two hundred years ago, now seems backwards in comparison to the ideas that we hold today. Today, with modern medical knowledge and technology, we scoff at many of the medical practices that past peoples performed. However, while medicine and technology has leapt forwards, the remnant social attitudes from those past eras still remain common today. Tracing the history of hysteria, we can see a legacy of systematic repression of women. For many women, hysterical symptoms have manifested as a physical and sexual rebellion against this repressive system; however, these women are quickly isolated and socially blacklisted by the patriarchal society that fears what it cannot dominate. In turn, cultural perceptions are reflected and canonized through literature, often perpetuating harmful tropes and stereotypes simply through repetition. This cycle of negative self-reinforcement continues to widen the gap between public perception of hysteria and the internal expression of hysteria, a phenomenon mirrored by contemporary literature. Through analysis of these hysterical texts, it is possible to identify the defining characteristics of hysteric literature and gain insight into the nature of hysteria itself. Detailed scrutiny of literary trends can reveal underlying truths, dispel the myth of the selfish hysteric—just who is unfair in the case of hysteria? Hysteria is an affliction of the self, a crystallization of desire manifest. Hysteria is distinct. Selfish. Unfair. It is repressed yet unyielding. It, like its literary reflection, says what needs to be said.

References

- Davies, M. (2011). N-grams data from the Corpus of Contemporary American English (COCA).
- Keh-Yih Su, J. T.-H. (2004). *Natural Language Processing – IJCNLP 2004*. Springer.

Knuth, D. E. (1994). *The Stanford GraphBase: A Platform for Combinatorial Computing*. New York: ACM Press.

Lispector, C. (1990). *Near to the Wild Heart*. Manchester, England: Carcanet Press Limited.

Rhys, J. (1966). *Wide Sargasso Sea*.

Steven Bird, E. L. (2009). *Natural Language Processing with Python*. O'Reilly Media Inc.

7

⁷ Due to using ebooks for *Wide Sargasso Sea* and *Near the Wild Heart*, I was unable to provide in-text citations with page numbers for these books.

Appendix 1

Table 1a: Male 2-Grams

48976	a	man
37643	the	man
18010	the	men
11987	young	man
11100	The	man
10705	old	man
8415	of	men
7204	this	man
6337	young	men
6256	two	men
4642	A	man
3748	The	men
3264	and	men
3225	one	man
3193	for	men
3093	of	man
2681	black	man
2566	these	men
2510	that	men
2447	white	man
2419	other	men
2312	his	men
2196	that	man
2113	another	man
2011	than	men
1860	both	men
1787	with	men
1716	black	men
1714	to	men
1657	three	men
1588	This	man
1518	gay	men
1475	good	man
1417	white	men
1410	by	men
1325	big	man
1233	other	man
1207	dead	man
1205	in	men

Table 1b: Female 2-Grams

33964	a	woman
19491	the	woman
16902	and	women
16566	of	women
15923	the	women
7739	for	women
6661	young	woman
6178	The	woman
5794	that	women
4001	this	woman
3926	young	women
3859	to	women
3315	A	woman
3286	old	woman
3152	other	women
2957	these	women
2880	The	women
2757	two	women
2455	many	women
2290	in	women
2048	with	women
1896	one	woman
1770	on	women
1688	by	women
1627	another	woman
1532	American	women
1398	Black	women
1384	are	women
1347	about	women
1337	some	women
1330	that	woman
1315	pregnant	women
1260	more	women
1223	black	women
1218	all	women
1193	first	woman
1137	than	women
1118	black	woman
1103	among	women

Appendix 2

The construction of a corpus containing hysteric literature and media was important for the various applications outlined in the paper. This corpus was constructed from the following works (by author):

Internal

Clarice Lispector- *Agua Viva, A Breath of Life, The Hour of the Star, Near to the Wild Heart, The Passion According to G.H.*

Jean Rhys– *Wide Sargasso Sea*

Sylvia Plath – all poems⁸, *The Bell Jar*

Audre Lorde- *Sister Outsider*

External

Georges Didi-Huberman- *The Invention of Hysteria*

Sigmund Freud- complete works⁹

The process of constructing the dataset proved to be accomplishable manually, given the small number of works. However, should the corpus expand to include a larger body of works, then at some point, it would become infeasible to construct such a dataset in this manner. By leveraging the various resources, it is possible to greatly expedite the speed of the operation while simultaneously increasing the efficiency of the script.

⁸ (found at http://www.internal.org/Sylvia_Plath)

⁹ I chose to include Freud's complete works because I felt that personally picking out which of his works should be included would introduce too much bias to the dataset

The standard method of parsing every file would be to read the contents of the file into memory, then split the string and store each individual value in some data structure, presumably an array or list. However, the employed method iterates through the files in the target folder¹⁰, passing each file to a MapReduce architecture¹¹. By leveraging the parallel computing capabilities of MapReduce, we are able to greatly increase the speed of the process when run on a parallel computing cluster. Construction of this sort will also allow for the script to be run on cloud-computing services (ie. Amazon Spark) to utilize an online cluster. Though the current script does not utilize stemming, it is easy to add a stemmer for future use simply by calling the stemming algorithm in another mapper step before reducing.

¹⁰ Assumes .txt format. Though PDF and EPUB libraries exist for Python, the method in which they read data differs greatly and makes it difficult to entirely automate the process

¹¹ Several of the MapReduce functions were written by the TAs of the CS1951A class