Article

Analyzing the Positive Sentiment Towards the Term "Queer" in Virginia Woolf through a Computational Approach and Close Reading

Heejoung Shin[1] (ID)

1. Library and Information Science, University of Illinois ROR, Urbana-Champaign, USA.

Abstract.   This article validates the thesis that Virginia Woolf's usage of the term "queer" is positive, and that the author is more progressive with her idea of things conceived as "queer" in the era characterized as literary Modernism and in English fiction as a whole from 1850s to 1990s. Using Word2Vec, a word embedding model, I locate the top 100 words semantically closest to "queer" in Woolf's works and in the works of other modernist authors, James Joyce, F. Scott Fitzgerald, D. H. Lawrence, Gertrude Stein, and Katherine Mansfield. I then measure the net positivity of each author's list and compare Woolf's with the individual authors', and then with words closest to "queer" in English fiction from 1850 to 2000. In demonstrating the usefulness of applying word embedding models in literary criticism, a field that has traditionally primarily relied on interpretation, this article aims to serve as a case study of how a computational approach can benefit close reading.

1. Introduction

The word "queer" appears more than 200 times in Virginia Woolf's published novels, short stories, and essays. This number may be statistically insignificant, but is nonetheless important for literary critics who aim to identify forms of repetition that do not constitute a cultural reproduction of rigid identity categories. This article thus explores how "queer" is deployed in Woolf's oeuvre against the backdrop of the history of English fiction, using Word2Vec, a powerful word embedding model (WEM) recently developed in the field of computational linguistics. This article particularly aims to look at whether Woolf's usage of the term "queer" is typical of her era characterized as literary Modernism and whether she is progressive in her treatment of queerness throughout the history of English fiction. To accomplish this goal, I compile the top 100 words semantically closest to "queer" in Woolf's works and in the works of other modernist authors, namely James Joyce, F. Scott Fitzgerald, D. H. Lawrence, Gertrude Stein, and Katherine Mansfield. Then, I compare the net positivity of each author's list of these words. I also analyze the associations around the term in English fiction as a whole from 1850 to 2000 to identify a larger usage pattern. As "queer" has a rich semantic history, having been used to indicate existing and emerging identity categories associated with what is out of sync with normativity proper, attending to the sentiment towards the term in literature can provide insight into how normativity is operative in a discursive field and how it is destabilized by its own operation. In demonstrating the

usefulness of natural language processing (NLP) using word embeddings in literary criticism, a field that has historically relied on interpretation, this article aims to serve as a case study of how a computational approach can benefit more nuanced literary analysis, beyond identifying topics that appear frequently.

I chose literary Modernism in launching this investigation, as it is a site where the term "queer" is deployed across the broadest spectrum in literary history. Originating in 16th-century England to refer to something strange, odd, eccentric, or illegitimate, "queer" began to suggest sexual practices that fall outside of the normative form of sexuality and gender in the 19th century (Barker and Scheele 2016, 24–27). By the late 19th and early 20th centuries, along with its sister terms, "fairy," "trade," and "gay," it had become a distinct identity category and a codeword within the gay male subculture in London, although its conventional usage as a term to denote "out of the ordinary" was still predominant among the British public (Houlbrook 2005, 162–163). During this period, "queer" had also gained a pejorative connotation for homosexuality and bisexuality (Houlbrook 2005, 179). The earliest known record of the usage of the term as such is from a letter written in 1894 by the Marquess of Queensberry to accuse Oscar Wilde of having an affair with his son, Alfred Douglas: "Snob queers like Rosebery" "corrupted my sons" (Barker and Scheele 2016, 27). It is also worth noting that in early 20th-century Britain, queer expressions of any sort do not necessarily correlate to a homosexual desire. Homosexuality and lesbianism themselves were more "permitted forms of sexuality" back then, although the latter was much less visible than the former (Houlbrook 2005, 10).

> In the British penal system, engaging in homosexual behaviors or importuning for homosex in public places were largely treated within a broader category of moral indecency, along with its twin problem of female prostitution.... It was only in the two decades after the Second World War that the forms of understanding that we often assume to be timeless – the organization of male [and female] sexual practices and identities around the binary opposition between homo and heterosexual.... solidified (Houlbrook 2005, 10).

Modernist authors wrote at this interesting moment where the term had not yet fully come into a rigid binary configuration of gender and there was still an overlapping assemblage of its usage. In their published works and in their often-suppressed manuscripts, letters, and diaries, "queer" is deployed in a variety of contexts, to denote homosocial/homoerotic desire, their own desired authority and authorship, and more broadly, whatever is at odds with normativity proper in terms of ethnicity, gender, nationality, etc. Yet, each writer's stance and sentiment towards what they call "queer" may radically differ. Gertrude Stein, for instance, constructs what she disavows in her characters' nationality and class around the notion of queerness in *The Autobiography of Alice B. Toklas* and in *Making of Americans*. In T. S. Eliot's suppressed poems, "queer" is almost always deployed in a self-deprecatingly comic and crude tone of voice to imagine stronger authority, often coupled with racial otherness and homosexual desire, to complement

what the poet views as a weakness in his own authority.[1] In Woolf, "queer" is usually described positively and is often associated with peculiar modes of existence, resistance, or self-expression shaped by one's moment-to-moment experience with the tyranny of the norm.[2] Demonstrating that this interpretation can be quantified will not only answer the question of whether Woolf is ahead of the curve against the backdrop of English literature and how our sense of what we consider queer has evolved across history, but also paves a new ground to frame research questions around racial and gender binaries.

A detailed discussion of the data, models, and methods used in this research follows, along with my interpretation of the modernist authors in question. Through this research, I validate the thesis that "queer" is more positive for Woolf than for her contemporaries explored in this article, and that Woolf's use of the term was ahead of her time, and possibly still is ahead of current usage of the term. Potentially, a meaningful discovery made throughout the research is that Joyce's works demonstrate the next most positive use of "queer" among this peer group. Indeed, the t-test performed for the positivity of "queer" for Woolf and Joyce cannot formally confirm that Woolf's use of "queer" is always significantly more positive than Joyce's, although the mean positivity based on the ten samplings drawn from each author's corpus is higher for Woolf than for Joyce. It is also quite noteworthy that the other two women authors' usage – Stein's and Mansfield's – exhibit the most strongly negative values. This suggests that computational approaches to literature can facilitate a more nuanced close reading and "interpretation" of gendered notions, beyond making distant reading possible.

2. Data

Sentiments, however positive or negative, are relative and exist on a spectrum. For this reason, the corpora of James Joyce, D. H. Lawrence, F. Scott Fitzgerald, Gertrude Stein, and Katherine Mansfield each are included as a comparison group to validate my thesis that Woolf's use of "queer" is more positive. Although different in nationalities, these authors all wrote in Europe. There is also a fair amount of usage of "queer" in these authors' works. Authors like T. S. Eliot, whose usage of the term is only visible in private letters, are excluded, although Eliot's works are rich with queer tensions and thus merit investigation from the perspective of queer theory. As the semantic meaning of "queer" had radically evolved in the first half of the 20th century, I also limited my selection to authors who wrote at roughly the same time as Woolf between the 1910s and the 1940s. Joyce, Fitzgerald, and Lawrence meet this condition. Stein and Mansfield are selected to

---

1. T. S. Eliot had written homoerotically-charged bawdy poems and sexual ribaldry (where he himself is imagined as femininized) and circulated them within his coterie which was exclusively comprised of his close male friends, Ezra Pound, Wyndham Lewis, and Conrad Aikon throughout his life, as a way to keep himself inspired. To see a detailed interpretation of how "queer" is figured among Eliot's coterie, see Introduction and Chapter One of my dissertation titled Granite and Rainbow: Queer Authority and Authorship in T. S. Eliot, W. B. Yeats, and Virginia Woolf. To see how "queer" is coupled with homosexual desire and Caribbean blacks, see T. S. Eliot's suppressed "Columbo and Bolo Verses," recently published in their entirety after the death of Eliot's wife, Valery Eliot. Interestingly, "queer" is nowhere to be found in Eliot's major poems that brought him fame. For more information about this, see all volumes of Letters of T. S. Eliot published by Yale University Press.
2. *Mrs. Dalloway* is representative of the positive construction of "queer" in Woolf. In the novel, "queer" often emerges in Clarissa's consciousness to describe the rainbow aspect of life. It is also employed to depict the novel characters' modes of life that falls outside of the conventional norm. In her first notes to the novel, Woolf writes, "Mrs. D. seeing the truth. SS [Septimus Warren Smith] seeing the insane truth." (Woolf and Wussow 1996, 450). Here, Woolf highlights the fact that truth is only seen by those who are categorized as queer.

verify that modernist authors' sentiments towards what was considered "queer'' may not necessarily correlate with their gender, although their creative activities spanned slightly differently from Woolf's and with Stein, "queer'' is visible mostly in *The Making of Americans*.

The very last point in the previous paragraph is particularly relevant to my choice of Joyce as part of comparison data. As sentiments around "queer'' can also vary among male authors, I hoped to select male authors that are representative of the broader spectrum of the sentiment towards "queer.'' Joyce is an ideal candidate to accomplish this goal. As Joyce scholars and biographers suggest, in real life, Joyce's stance towards homosexuality remained fairly neutral; while Joyce was not above deriving entertainment from his homosexual friends, he was neither sympathetic nor unsympathetic to homosexuality (Norris 1994, 357). Nonetheless, what is intriguing about Joyce's works is that the centrality of feminized (often racialized and satirized) male characters and masculinized female counterparts amid an intense desire for homosocial and homoerotic affiliation emerges as one of the most visible themes. Joyce was also rebellious against the norm of his time and place. He condemns three Irish norms – family, Irish nationalism, and the Catholic Church – as stifling and detrimental to the development as an artist. Regarding Fitzgerald's and Lawrence's sentiments, although one should be careful to not make a facile generalization, substantial existing research demonstrates that Lawrence writes more in a heteronormative convention while Fitzgerald views what he calls queer as an essential human condition:

> Begin with an individual, and before you know it you find that you have created a type; begin with a type, and you find that you have created-nothing. That is because we are all queer fish, queerer behind our faces and voices than we want anyone to know or than we know ourselves. When I hear a man proclaiming himself an "average, honest, open fellow,'' I feel pretty sure that he has some definite and perhaps terrible abnormality which he has agreed to conceal-and his protestation of being average and honest and open is his way of reminding himself of his misprision (Fitzgerald 1989, 317).

All modernist authors' texts utilized in this project are drawn from Project Gutenberg Australia.[3] The data on Woolf contains most of her published novels, short stories, and essays. Like the data on Woolf, data on Fitzgerald, Lawrence, and Mansfield each consists of the corresponding author's major novels, short stories, plays, and essays. For Joyce, I use three novels, *Dubliners*, *A Portrait of the Young Artist as a Young Man*, and *Ulysses*, available on Project Gutenberg Australia. Similarly, for Stein, I use *The Autobiography of Alice B. Toklas*, *The Making of Americans*, *Three Lives*, and *Geography and Plays*, available on the same site. As expected, there are differences in the size of each author's corpus. Woolf's corpus accounts for 1,760,779 words in total, Joyce, 417,765, Fitzgerald, 615,126, Lawrence, 2,371,834, Stein, 699,562, Mansfield, 239,166.

The entire English fiction dataset from 1850 to 2000 (Google N-Grams eng-fiction-all) I use for this study is from the dataset developed as part of the study titled "HistWords: Word Embeddings for Historical Text'' (Hamilton et al. 2016b, Hamilton et al. 2016a). I

---

3. For a complete list of literary works used to create the corpus, see Appendix A.

use HistWords' pre-trained word embeddings to extract the top 100 words closest to "queer" for each decade, to compare them with Woolf's list.

3. Models and Methods

3.1 Associations around "Queer" in Woolf and in Joyce, Fitzgerald, Lawrence, Stein, and Mansfield

One way to measure Woolf's and others' sentiments towards the term "queer" is to compile a list of the top 100 words semantically closest to "queer" in the texts of each author and compare their net positivity. Word embedding models (WEM) are optimized for this task. Unlike topic models that map a text as a network of words based on co-occurrences, word embedding models map a text as relationships between words so that they enable "searching for spatial relations embedded in words", a framework, I would argue, essential to close reading highlighting the particular, effected by close attention to the relationship between words (Schmidt 2015).

To develop and train word embeddings specific to each author, each author's oeuvre was combined into a separate single text file. While it is widely known to be effective to adapt word embeddings trained on large collections of texts for predictive purposes, it is worth highlighting again that it is each author's individual sentiment to a certain word that emerges within the works of his or her creation that is being analyzed in this research, and that in literature as a peculiar genre, plethora of figurative words and styles are employed and destruction of normative usage of language, experimented.[4] If, for example, an author consistently uses "queer," "miracle," and "loving" to describe, say, "pebbles," these four words are closer in meaning and are thus placed closer within the space of the particular author's corpus.[5] For other authors, however, "pebbles" may not likely be queer at all; they may likely be ordinary objects. This implies that, as Laura Burdick, Jonathan K. Kummerfeld, and Rada Mihalcea also aptly point out, word embeddings change if different authors' texts or different collections of texts are used as input, as words have different connotations when employed to discuss different topics (Wendlandt (Burdick) et al. 2018). This further suggests that using word embeddings trained from a large number of texts that have nothing to do with each author might be risky, no matter how precise or sophisticated they are. For this reason, I took the path of developing and training word embeddings specific to each author, although this choice inevitably raises a question about the relatively small size of individual authors' corpus and methodology.

In developing and training word embeddings for each author's corpus, I chose Word2vec, using Gensim, a Python library, which implements many variants of word embeddings (Řehůřek 2022). Specifically, Gensim's Word2vec is well maintained and takes a single text file containing each author's corpus as input. My use of Gensim's Word2vec was primarily to transform the authors' corpora into semantic spatial vectors, so I could extract "queer"'s semantic vector and its top 100 closest words.

---

4. With profusion of styles and the quantity of allusions, modernist authors' works are, in general, experimental and difficult to interpret, with Joyce's *Ulysses* being one of the most appropriate examples.
5. Here, I use this rather strange example to remind the reader that words are essentially signs.

As I was working with limited amounts of texts, there may also be a dispute about the choice of Word2vec, which generally requires more text input. To offset this concern, I followed the best practice recommended by Ben Schmidt for those working with relatively smaller corpora on Word2vec: "Run many iterations. A hundred, maybe. If your model trains in less than a minute, it's probably no good" (Schmidt 2017). Experimenting with the size of vector dimensionality was also useful in getting meaningful embeddings.[6] Additionally, as it was uncertain how much the Word2vec training runs across sentence boundaries, sentence triplets were used instead of single sentences to minimize information loss. I ran 100 iterations in developing and training the models for all authors except Mansfield, for whom I ran 200 iterations given the small corpus. The training time for all models vary due to the corpus size. The longest training time was 8 minutes 19 seconds for Lawrence. The shortest training time was 1 minute 23 seconds for Mansfield. Stop words were not removed from the compiled text files because in the case of Word2Vec models, they can provide contextual information. The model can also indirectly learn the sentence representation while feeding the context as the output or input (Arindam 2019).

By the time the models for each author had been created and trained, I was able to extract the top 100 words most similar to "queer" from the corpus of each author. Yet, before measuring the net positivity of each word, it was necessary to ensure that the words identified as most similar to "queer" were not dependent on one or two instances. I thus ran ten models on different subsamples (sentence triplets) of the authors' corpora. As each word is given its own vector (position) in the space of the corpus specific to a certain author, I measured the distance from "queer" to positive words, and to negative words, and ultimately, the difference between the distances. Since we were measuring distance rather than similarity, the positivity of "queer" (net positivity) could be assessed as such:

> Positivity of "queer" (net positivity) = negative distance (distance from negative) – positive distance (distance from positive)

In terms of the positive and negative words, I used a list created by Bing Liu in 2005, which contain roughly 5,000 positive and negative words respectively (Liu et al. 2005). I performed a t-test for the positivity of "queer" for Woolf and individual authors respectively to formally confirm the stability of the pattern I observed.

3.2 Associations around "Queer" in Woolf and in English Fiction from the 1850s to the 1990s

To situate Woolf's use of "queer" in the broader context of English fiction beyond literary Modernism and to see whether Woolf was progressive with her ideas of queerness, I measured how different Woolf's associations are from other authors' associations across the collective history of English fiction from the 1850s to the 1990s, using "English Fiction (1800s-1990s) (from Google N-Grams eng-fiction-all)," one of the pre-trained word embeddings developed by William L. Hamilton, Jure Leskovec, and Dan Jurafsky for

---

6. According to one entry from stackoverflow, in general, smaller vector dimensionality works better for smaller corpora. For smaller corpora, vector-dimensionality should be no more than the square-root of the count of unique words. See https://stackoverflow.com/questions/66267818/minimum-number-of-words -in-the-vocabulary-for-word2vec-models.

their project titled HistWords.[7] As the vector of "queer" itself is missing in HistWords' dataset between the 1800s and the 1840s, this period was excluded.[8] I took the path of extracting the top 100 words closest to "queer" from each decade from the 1850s to the 1990s and measured the sentiment of those the same way I did for my selected authors. In other words, for each decade, I measured the distance from "queer" to positive words, and to negative words, and calculated the difference between the distances, using Liu's lists. The positivity of "queer" (net positivity) was similarly assessed as (negative distance - positive distance.)

4. Results

4.1 Stability Test and p-Values from t-Tests

As can be seen from Figure 1, the stability tests for Woolf, Joyce and Fitzgerald returned positive values, the ones for Lawrence, Stein and Mansfield returned negative values. Notably indeed, for Woolf, all 10 runs returned positive numbers. It is not a big difference, but there is usually a lean toward the positive (90-100 percent), even when we run the model multiple times and compare all runs. This output shows that for Woolf, "queer" is always more positive than negative. For Joyce, the test outcome is consistently positive although it varies in degree. For Fitzgerald, it is mostly positive, although it is less positive than for Woolf. For Lawrence, Stein, and Mansfield, it is consistently negative and, like Joyce's data, there is a large variance.
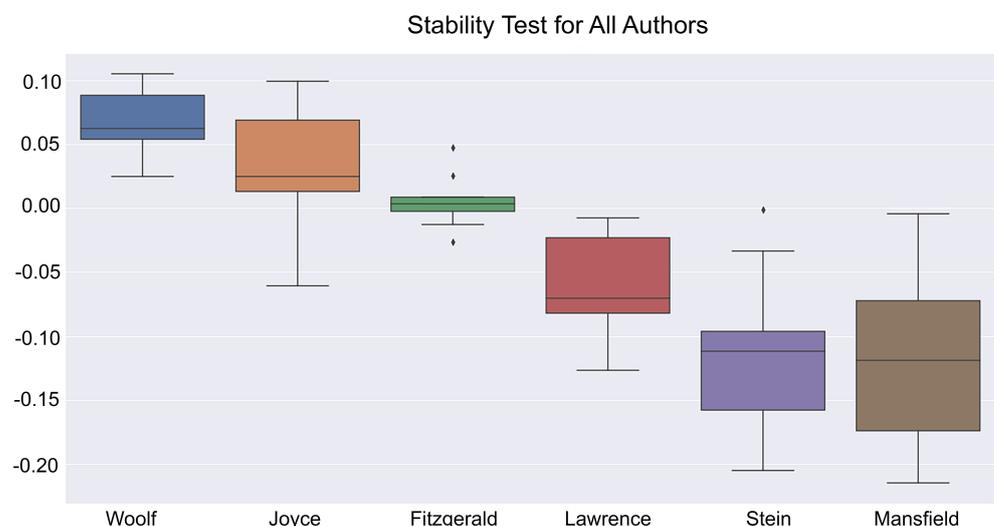
Figure 1: Box plot showing net positivity of the term "queer" for Woolf, Joyce, Fitzgerald, Lawrence, Stein, and Mansfield based on ten tests.

---

7. To borrow Hamilton's description, the goal of the HistWords project is to facilitate quantitative research in diachronic linguistics, history, and the digital humanities. They release pre-trained historical word embeddings spanning from 1800 to 2000 for multiple languages - English, French, German, and Chinese. Embeddings constructed from many different corpora and using different embedding approaches are also included. To read more about this project or access their tools and datasets, visit their site titled HistWords: Word Embeddings for Historical Text on `https://nlp.stanford.edu/projects/histwords/`.

8. This is what I see as a limitation of using pre-trained word embeddings in research aiming to identify the particular. Both topic models and word embedding models tend to suppress low-frequency data, the very data that the close readers may want to explore. The fact that the token "queer" is entirely missing in the dataset of the first half of the 19th century reveals how the norm has operated in a discursive field to oppress those not considered to be the norm. Apparently, the term "queer" was in existence and in use in the early 19th-century.

| Woolf compared to...	p-value
Joyce	0.0882
Fitzgerald	0.0008
Lawrence	1.208e-07
Stein	5.961e-07
Mansfield	9.993e-07

Table 1: Test for statistical significance regarding the difference in positivity values between Woolf and the other authors.

The p values from each of the t-tests for the net positivity in Woolf compared to the other authors are shown in Table 1. Except for the case of Woolf and Joyce, the p values are much smaller than 0.05. This shows us that in the cases of Woolf and Fitzgerald, Woolf and Lawrence, Woolf and Stein, and Woolf and Mansfield, the difference of means between these samples would not be likely to occur by chance if these samples were drawn from populations that actually had the same mean value. In short, we can claim with statistical confidence that "queer'' is more positive in Woolf than it is in Fitzgerald, Lawrence, Stein, and Mansfield. However, we cannot claim with assurance that Woolf's usage of queer is always more positive than Joyce's, because the p-value is above the conventional value of 0.05.

4.2 Associations around "Queer'' from the 1850s to the 1990s from Histwords' Word Embeddings

Figure 2 reveals some interesting patterns about the associations around "queer'' in the history of English fiction.

First, historically, the term "queer'' consistently had negative connotations, indicated by the negative net positivity numbers. Interestingly, there was a big shift towards the positive in the 1860s. After that, until the 1890s, it consistently moved further negative. We observe a consistent movement towards positive from the 1930s to the 1980s, although the general sentiment towards the term was still negative. Intriguingly, however, there was a move back towards negative in the 1990s. Viewed together with both Google Books Ngram Viewer's and Bookworm: HathiTrust's data (Figure 3) in regard to the frequency of "queer'' across English fiction between 1930s and 1990s, this movement merits investigation.
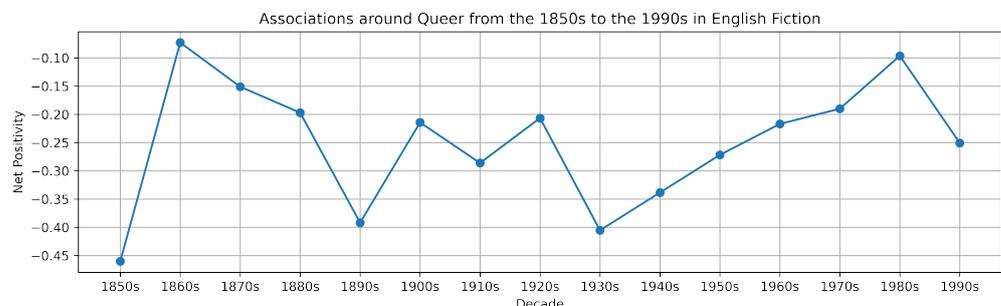
Figure 2: Visualization showing associations around the term "queer'' from the 1850s to the 1990s in English fiction, which had always been negative.
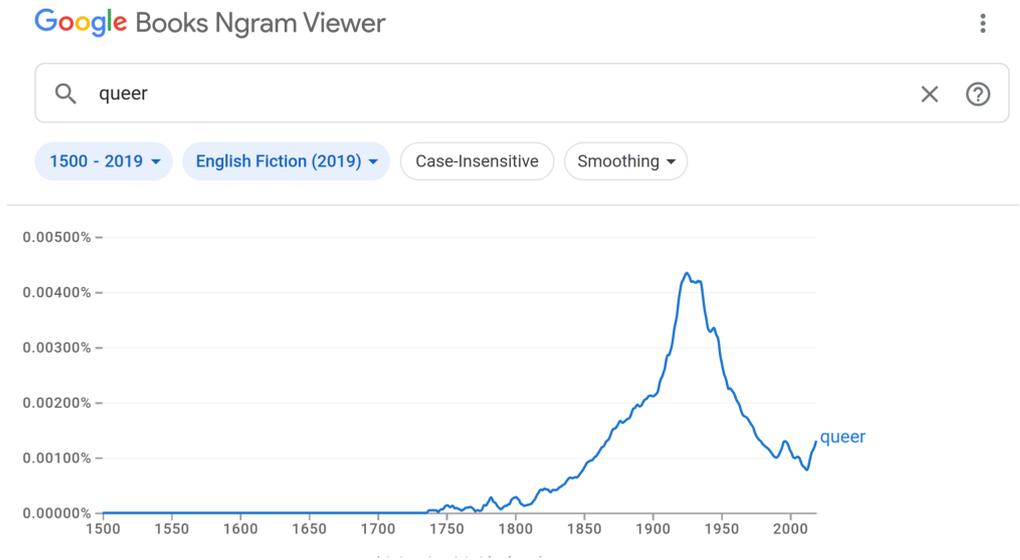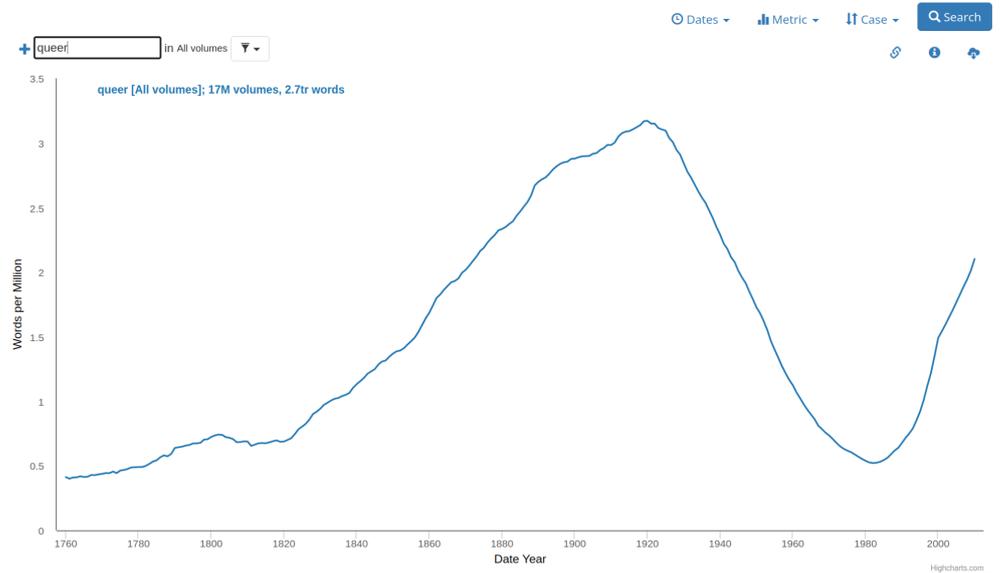
Figure 3: Google Books Ngram Viewer (top) and Bookwork HathiTrust (bottom) data in regards to the frequency of "queer" across English fiction from 1800 and 1760, respectively, to 2000 and beyond ("Queer" 2022, "Queer" 2022).

Going back to the discussion of Figure 2, "queer'' became less and less frequently represented in English fiction from the 1930s until its frequency increased back again in the 1990s. That is to say, during this period, the frequency of "queer" and the net positivity of "queer'' moved in opposite directions. Without data on the 2000s and the 2010s, it is difficult to determine whether the move further negative in the 1990s was part of a larger trend. It might be due to a conservative backlash against the LGBT rights movements that became increasingly visible following the Stonewall riots of 1969, which requires a separate investigation (Boag 2021).[9] One claim I can still confidently make, though, is that Woolf was more positive about the things that were viewed as "out of the ordinary,'' and that her use of the term was progressive compared to its use in English literary history from the 1850s to 1990s.

4.3 Words associated with "queer" in different author corpora

Table 2 shows the top 100 words closest to "queer'' and their corresponding similarity score for Woolf from one model. Strikingly, the words identified as closest to "queer'' are not simply adjectives but include nouns and proper nouns. For example, Maisie and Walsh are characters from *Mrs. Dalloway*, and Richard indicates Richard Dalloway, who appears both in *Voyage Out* and *Mrs. Dalloway*. The relative proportion of positive, neutral, and negative words varies by model.

In Figure 4, words carrying a positive sense are plotted in green, words with negative connotation, in red, and words that are neutral, that is, words not present in Liu's positive or negative words lists, in grey. It is worth noting that on Liu's lists of positive and negative words, "queer'' is categorized as negative. As that is unlikely to be the case for Woolf, it is marked as a separate category on the graph in purple. The X and Y axes are used to represent semantic vectors specific for each word. Thus, words plotted closer to "queer" on the graph indicate their closer proximity to "queer" in meaning in Woolf. Principal Component Analysis (PCA) was used to reduce the dimensions for the plot.

We can see that, while most words are categorized as neutral, there are slightly more positive words than negative ones: 12 vs. 10. This appears to be a small difference. Yet, it is important to remember that what we measured earlier is the net positivity of the words closest to "queer.'' This means that in the ten samplings drawn from Woolf's corpus, positive words always outnumber negative words among the top 100 words identified as closest to "queer,'' regardless of the proportion of neutral words. Another potentially important discovery we can make is that, as I mentioned earlier, the model identifies a significant number of proper nouns and nouns as words close to "queer.'' Proper nouns and nouns are extremely important in literary analysis, as they are the locus in which interpretation is anchored, whether it is about themes, tropes, characters, or sentence structures.

---

9. Several studies were conducted on the conservative backlash against the LGBTQ movements in the late 1890s and the 1990s, among which Peter Boag's "Gay and Lesbian Rights Movement'' is one of the most representative. This phenomenon was universal across the globe.

Ranks 1-33	Ranks 34-66	Ranks 67-100
awfully, 0.386	masculine, 0.280	vacancy, 0.260
sized, 0.371	stogdon, 0.279	innocence, 0.259
young, 0.368	exploded, 0.279	seeming, 0.259
suspected, 0.365	comparison, 0.278	hovering, 0.259
absorption, 0.355	deleterious, 0.277	smiles, 0.255
posing, 0.351	slang, 0.277	hives, 0.255
nice, 0.350	squirrels, 0.276	suits, 0.254
maisie, 0.344	this, 0.276	roused, 0.254
oblivion, 0.339	plans, 0.276	transferred, 0.253
horrors, 0.326	significant, 0.274	falsehood, 0.253
speeches, 0.323	asquith, 0.274	accomplishment, 0.253
evanescent, 0.311	persian, 0.273	hideous, 0.252
reputed, 0.306	negligently, 0.271	anyhow, 0.252
just, 0.301	tirade, 0.271	dog, 0.251
blotted, 0.300	armenians, 0.271	different, 0.251
buzzing, 0.300	invalids, 0.270	albanians, 0.250
dreaded, 0.296	omitting, 0.270	craftsman, 0.249
basins, 0.296	proof, 0.269	escaped, 0.248
perennial, 0.295	immovable, 0.267	cheerless, 0.248
assuring, 0.293	game, 0.267	ascertained, 0.248
booming, 0.292	richard, 0.267	solicitous, 0.247
bent, 0.291	convict, 0.266	judd, 0.247
hailed, 0.290	porous, 0.266	crabs, 0.246
tender, 0.290	fountains, 0.266	elms, 0.246
twice, 0.290	that, 0.266	mingling, 0.245
lampsher, 0.288	affinity, 0.266	dangled, 0.245
walsh, 0.288	sucked, 0.263	incompatible, 0.245
heavens, 0.286	cleanliness, 0.263	ceremonial, 0.245
kissing, 0.284	contamination, 0.263	withheld, 0.245
caen, 0.282	about, 0.263	groom, 0.244
pockets, 0.282	happened, 0.262	hitching, 0.244
painters, 0.280	equitable, 0.261	diction, 0.244
cocking, 0.280	toy, 0.260	mentioned, 0.243
		tidy, 0.243

Table 2: The top 100 words closest to "queer" and their corresponding similarity score for Woolf from one model.

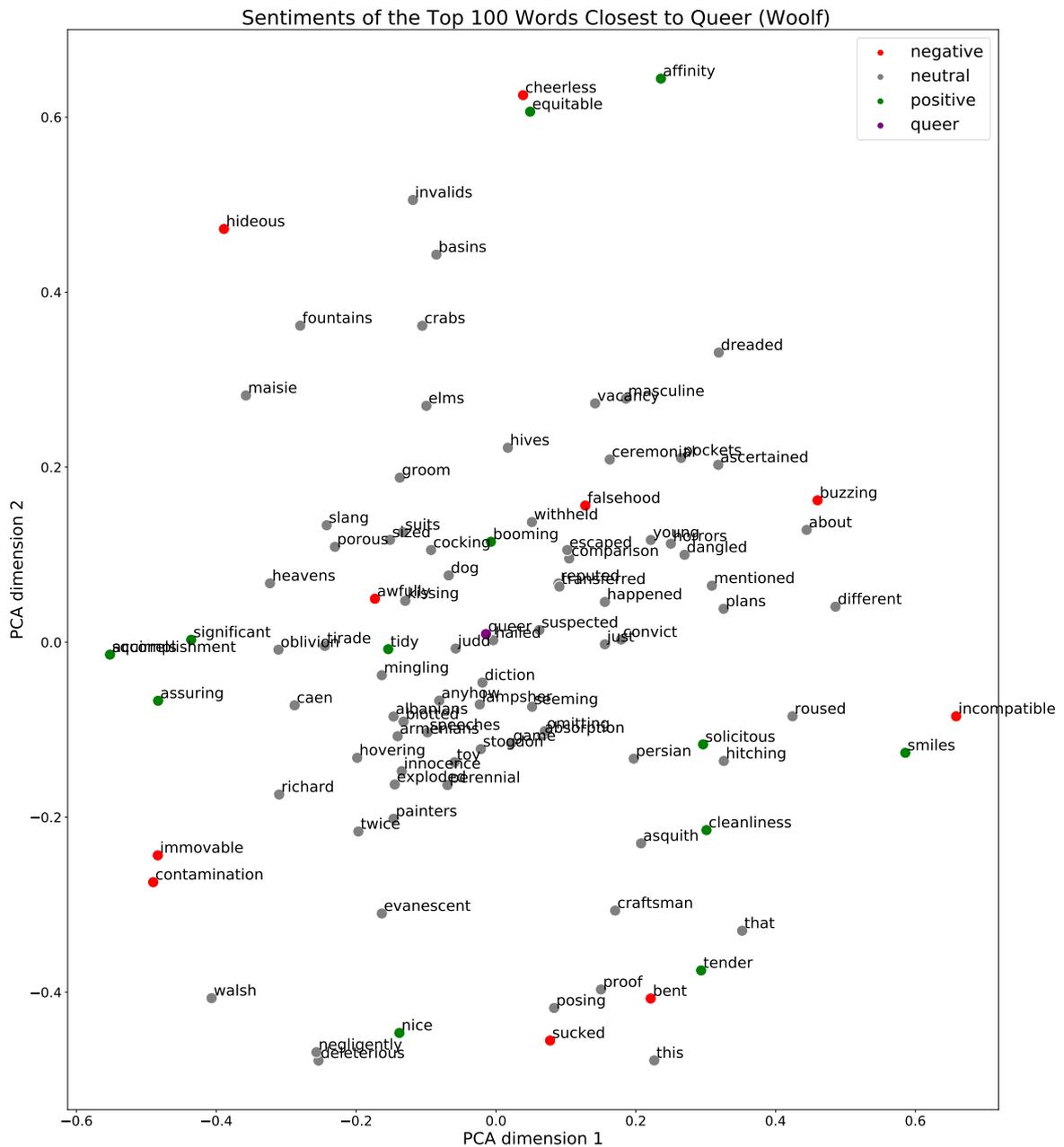Sentiments of the Top 100 Words Closest to Queer (Woolf)

Figure 4: Plot of sentiments of the top 100 words closest to queer in Woolf's text.

Undeniably, for Joyce as well, "queer'' is consistently used positively. For Fitzgerald, 6 models return positive outcomes. The plots of Joyce's and Fitzgerald's top 100 words closest to "queer'' are provided in Figure 5 and Figure 6. Similar to Woolf's list, we see nouns and pronouns present in Joyce's and Fitzgerald's lists. One can notice, however, that the proportion of positive and negative words decreases in both Joyce and Fitzgerald, compared to Woolf. How 100 individual terms are deployed around "queer'' in the texts of Joyce and Fitzgerald used in this research, along with their idea of (hetero)normativity proper, will require a separate in-depth exploration. A point that should be noted here is that in *Ulysses*, "queer'' is often deployed within the male protagonist Leopold Bloom's stream of consciousness as a reference to the intricacies of life, which resist a facile, binary categorization. Above all, in the case of Joyce, that all ten models return positive outcomes testifies to Norris' depiction of Joyce as unbiased with the matter of homosexuality to a certain degree. Norris argues that, not being one of his own personal predilections, homosexuality is an aspect of human behavior to which Joyce did not devote a great deal of attention (Norris 1994, 357). Indeed, Joyce views homosexuality as a product of the social system, rather than as a personal trait that should be abhorred. In his essay "Oscar Wilde: The Poet of Salome,'' written approximately around the same time as *A Portrait of the Artist as a Young Man*, Joyce describes Wilde's homosexuality as the "logical and inevitable product'' of sexual "secrecy and restrictions'' and "unhappy mania'' endemic to British public schools (Valente 2004, 215). Similarly, Colleen Lamos views the matricidal fantasies that often emerge throughout *Ulysses* as the author's defensive gestures that attest to the violent consequences of the modern disavowal of same-sex desire (Lamos 1998, 15).

On the other hand, the Lawrence plot from one model, seen in Figure 7, shows us that compared to Woolf and Joyce, there are a significantly greater number of negative associations around "queer.''

Intriguingly, while running multiple iterations of the model, I could see "savage" and "barbaric'' appear several times as one of the top 100 terms closest to "queer'' for Lawrence. This is a meaningful discovery given that Lawrence is notorious for having written in the heteronormative convention and for associating whatever is at odds with conventional femininity with the primitive. Indeed, this discovery verifies that what Gayle Rubin terms as "traffic in women'' strongly operates in Lawrence's narrative strategy. In other words, in Lawrence, the primitive feminine trope is deployed only to strengthen the bond between males or celebrate conventional ideas of masculinity and femininity, with *Women in Love*, *Sons and Lovers*, and "The Fox'' being only a handful of examples (Rubin 1975, 180). In *Women in Love*, for example, the sisters, Ursula and Gudrun, particularly, Gudrun's unruly sexuality and rebellious personality – are deployed to ultimately strengthen the bond between Birkin and Gerald. After all, in the novel, Gudrun is constructed as an artist known for her primitive, savage art. The novel ends with Birkin's mourning over the loss of Gerald who freezes to death after his violent fight with Gudrun.

The discovery made possible by the model can also be potentially useful when used to complement or modify existing interpretations. For example, in *Gone Primitive*, Marianna Torgovnick points out that in Lawrence, there are two versions of the primitive (Torgovnick 1991, 159). The first is a feminine version: the primitive as "dangerous,''
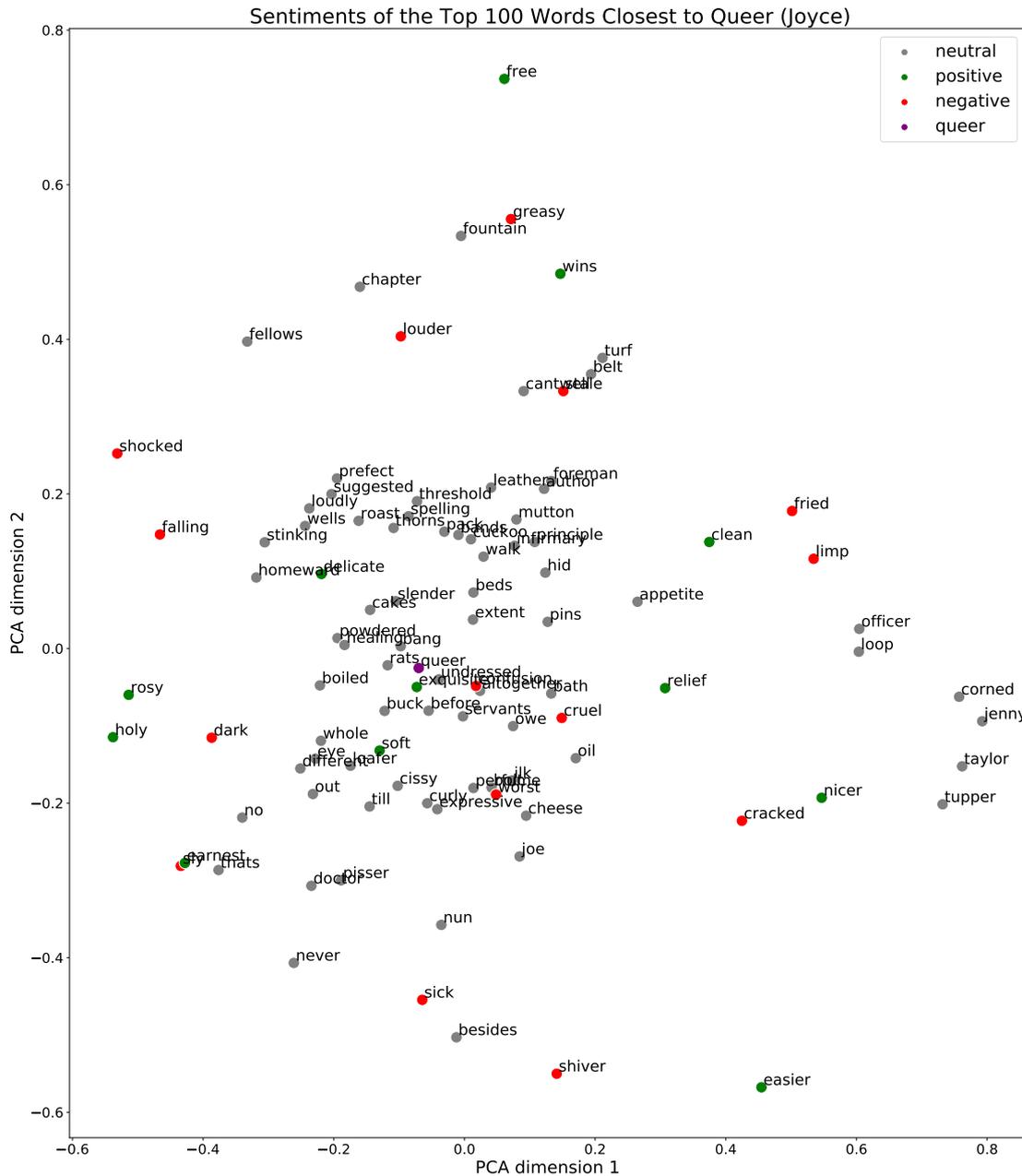
Figure 5: Plot of sentiments of the top 100 words closest to queer in Joyce's text.
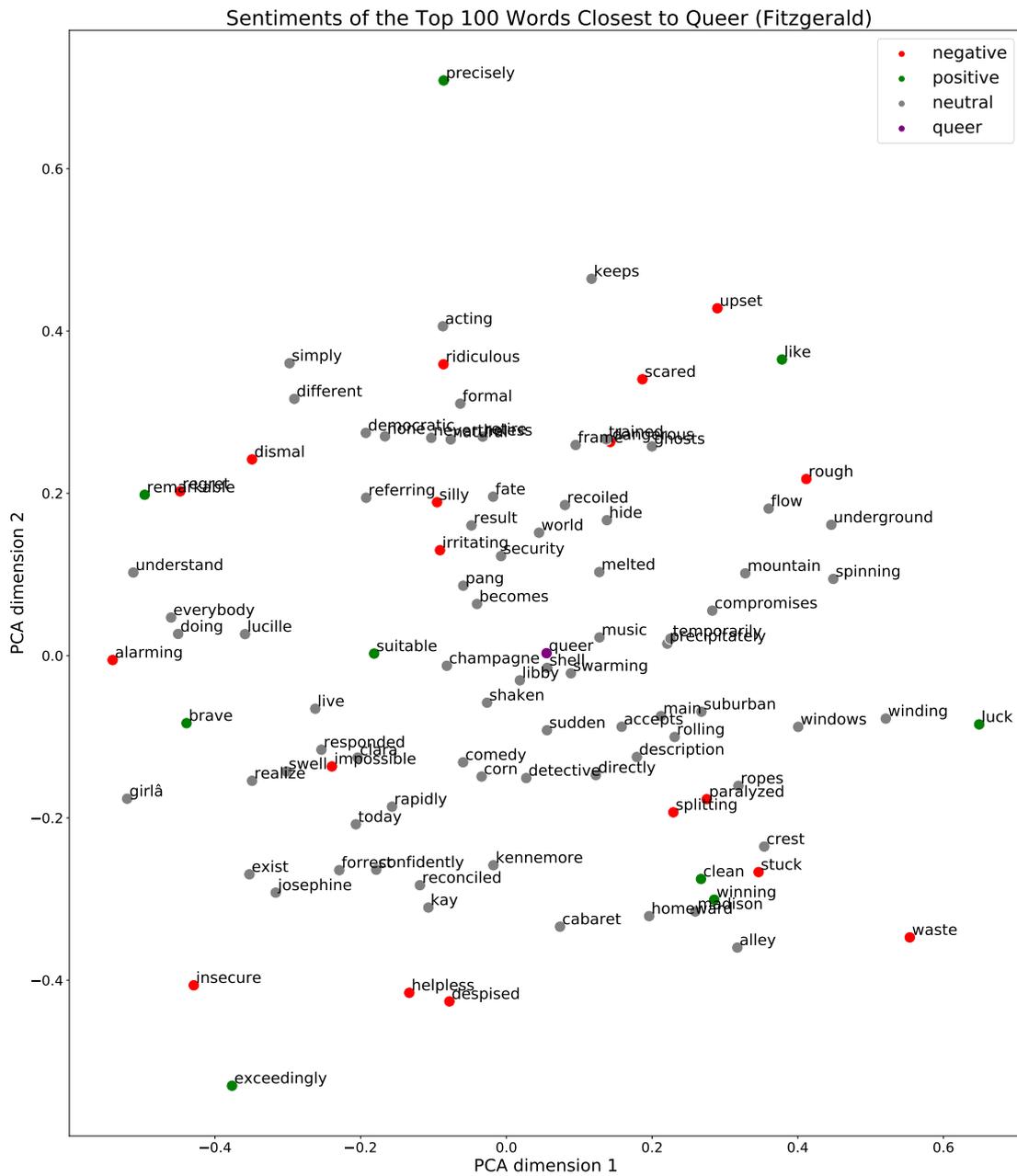
Sentiments of the Top 100 Words Closest to Queer (Fitzgerald)

Figure 6: Plot of sentiments of the top 100 words closest to queer in Fitzgerald's text.
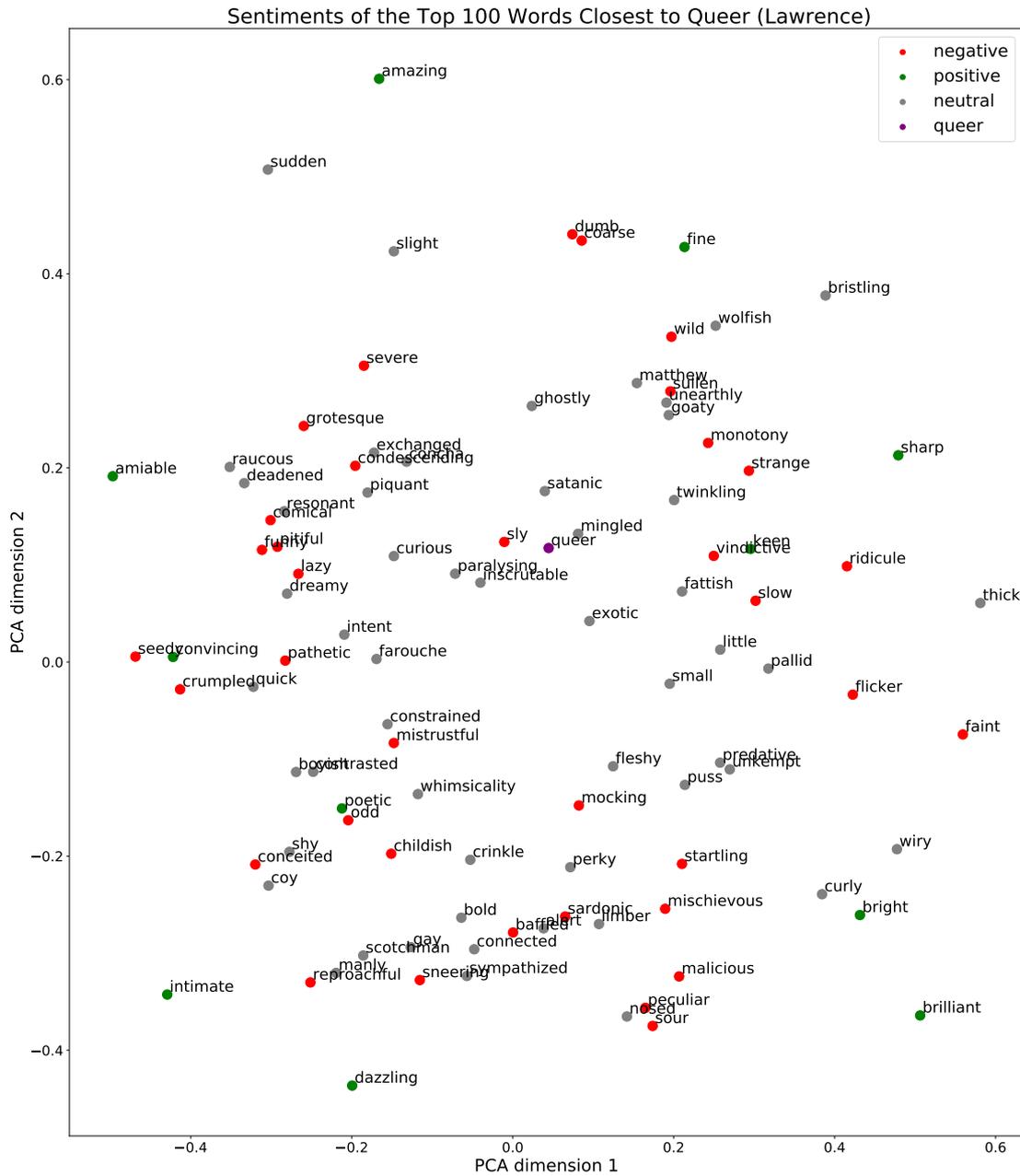
Figure 7: Plot of sentiments of the top 100 words closest to queer in Lawrence's text.

"irrational," "something to be feared," and "the idealized noble savage" (Torgovnick 1991, 159). The second is a masculine version: the primitive as "regeneration" (Torgovnick 1991, 159). The emergence of "savage" and "barbaric" as words closest to "queer" in Lawrence, along with Lawrence's negative sentiment towards "queer," demonstrates that "queer," for Lawrence, is associated more with the negative version of the primitive – the feminine –, which is shaped by his frustration with disappearing Western values – the conventional idea of masculinity and femininity where the former is associated with regeneration, the latter, reproduction– with the arrival of the modern (Torgovnick 1991, 153).

The plots for Stein (Figure 8) and Mansfield (Figure 9) are also provided. How and why each corpus exhibits this pattern, other than what I mentioned earlier about Stein's tendency to align class and nationality with "queer," requires a separate investigation. Nonetheless, the outcome that the net positivities of these women authors' corpus are the lowest suggest that female authors do not necessarily have a positive sentiment towards what is considered "out of the ordinary," that the author's gender does not necessarily correlate with their sentiment towards "queer."

5. What "Queer" Represents in Woolf

Here, I take the approach of a literary critic, to validate my outcome with close reading, to argue that Woolf, as a renowned feminist writer and queer author, had a keen sense of how the norm manifests itself as various forms of power to oppress those who do not conform to it. Unlike the male authors who were spoiled for choice, Woolf grappled with the absence of a strong female tradition and keenly sensed herself in conflict with the masculinist, heteronormative climate of the British Empire and as permanently in exile. A series of medical treatments she had received due to her recurrent mental and physical illness, albeit a disaster in her personal life, offered her a powerful tool to interrogate the tyranny of the norm as a form of social repression (Lee 1997, 186).

As a form of resistance, Woolf deploys "queer" to create desires, personalities, and relationships – bodily, aesthetic, and epiphanic – that exist outside of the paradigmatic markers dictated by normativity. In her diary entry on December 21, 1925, Woolf employs "queer" to mean both bodily consummation and esthetic fulfillment after spending her first night with Vita Sackville-West at Long Barn:

> There is her maturity & full breastedness… there is some voluptuousness about her. But then she……so lavishes on me the maternal protection which, for some reason, is what I have always most wished from everyone…. I shall be hung about with trailing clouds of glory from Long Barn wh. always disorientates me & makes me more than usually nervous: Then I am—altogether so queer in some ways. One emotion succeeds another (Woolf 2018, 11654).

In *Mrs. Dalloway*, "queer" is employed in a sympathetic and lovable note to describe the truth behind her characters who are viewed as failures by the social norm. Earlier, we saw "Maisie" plotted as one of the top 100 words closest to "queer" in Woolf's corpus, along with "invalids." Maisie is a low-class woman from Edinburgh, who appears very
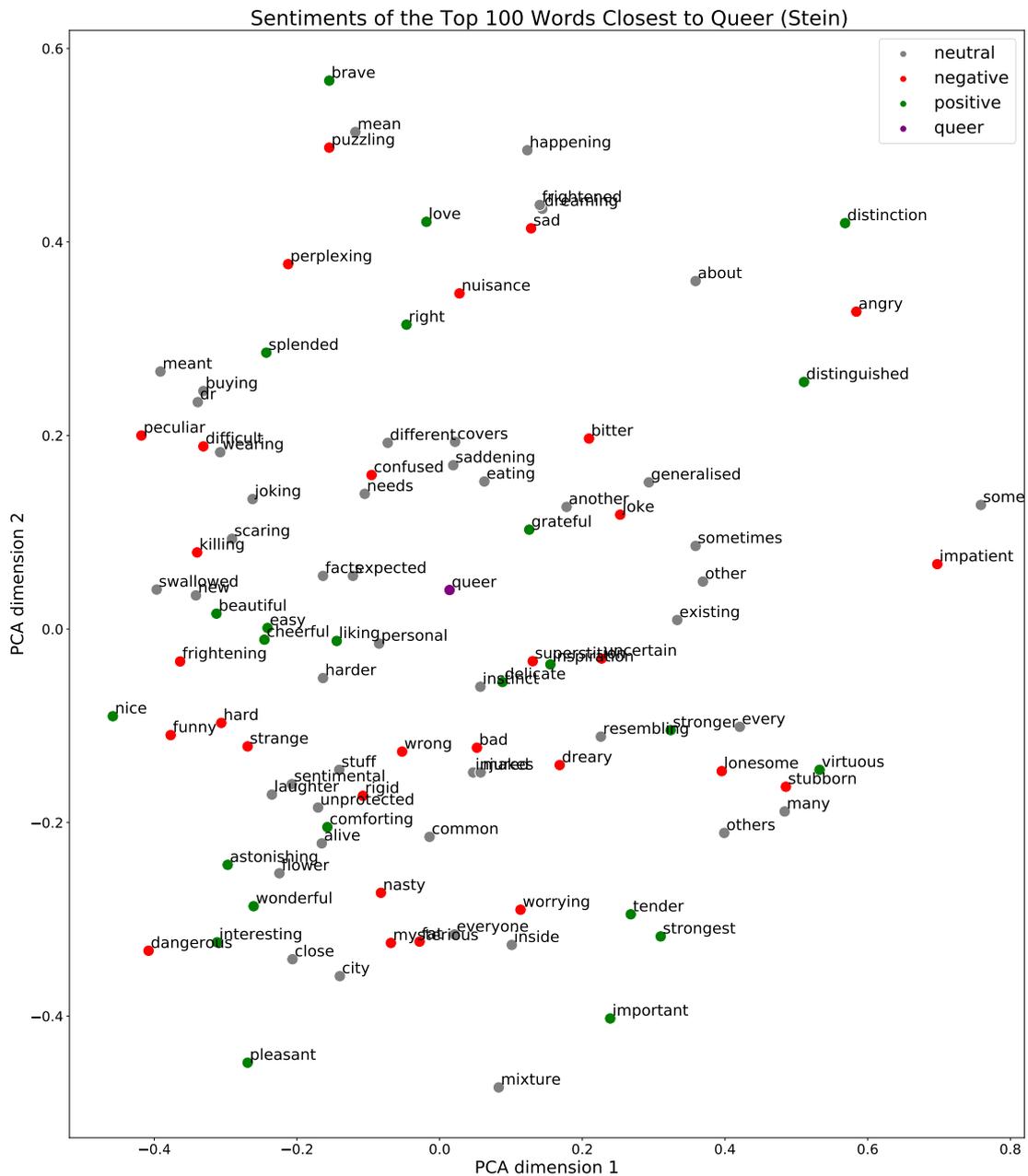
Figure 8: Plot of sentiments of the top 100 words closest to queer in Stein's text.

Sentiments of the Top 100 Words Closest to Queer (Mansfield)

Figure 9: Plot of sentiments of the top 100 words closest to queer in Mansfield's text.

briefly at the beginning of *Mrs. Dalloway*. What is specifically remarkable is the tangible link that Woolf establishes between the term "queer'' and those who were parceled into the category of "queer'' in the oppressive British interwar regimes, through Maisie Johnson's stream of consciousness in her first encounter with London.

> They seemed queer, Maisie Johnson thought. Everything seemed very queer. In London for the first time, come to take up a post at her uncle's in Leadenhall Street, and now walking through Regent's Park in the morning, this couple on the chairs gave her quite a turn; the young woman seeming foreign, the man looking queer....... For she was only nineteen and had got her way at last, to come to London; and now how queer it was, this couple she had asked the way of, and the girl started and jerked her hand, and the man—he seemed awfully odd; quarrelling, perhaps; parting forever, perhaps; something was up, she knew; and now all these people (for she returned to the Broad Walk), the stone basins, the prim flowers, the old men and women, invalids most of them in Bath chairs—all seemed, after Edinburgh, so queer (Woolf 1981, 26).

Remarkably, in this short passage, "queer'' is employed five times in total. Here, Maisie Johnson calls Septimus Warren Smith, a veteran of World War I and his Italian wife Rezia each queer and then all the people she comes across in Regent's Park: "The old men and women, invalids most of them in Bath chairs.'' For Maisie Johnson, "queer'' is a term that binds all these people who appear out of time and out of place – invalids sitting in Bath chairs, the foreign (Rezia), and the awfully odd and mad (Septimus), who suffers from shellshock. Remarkably, as the story unfolds, readers also notice that a link between "queer'' and a same-sex desire is tellingly made in Septimus when his close relationship with his wartime officer Evans is repeatedly highlighted. Ultimately, Septimus commits suicide in defiance of Dr. Holmes and Sir William Bradshaw's desire to "straighten'' his "shell shock,'' his madness. Here, in his triumphant choice of death over treatment, we see "being queer'' is also equated with a willing choice and a vehicle for resistance.

Maisie, Septimus, and Rezia are not the only characters associated with "queer'' on a sympathetic note. In numerous instances throughout *Mrs. Dalloway*, the characters' impregnable queerness – Clarissa's bisexuality, Richard's anxiety over his masculinity, the adventurous queer child within Peter Walsh and Elizabeth, and Miss Kilman's misandry and obsession with food – is directly described as "queer'' or finds its way out as spatial metaphors in its askew relation and stubborn resistance to normativity. Earlier, we saw "Richard,'' a politician who is also Clarissa's husband, and Peter "Walsh,'' Clarissa's friend, identified as terms close to "queer'' in Woolf's corpus. Strikingly, in Woolf's manuscript of *Mrs. Dalloway*, Richard emerges as a queer trope out of place: "Richard had all the marks of that queer breed'' (Woolf and Wussow 1996, 75). Indeed, it is repeatedly implied throughout the novel that politics does not suit Richard's simple character and love for nature. Throughout the novel, Richard's nostalgia for Norfork's sky and movements of grass and breeze is constantly placed in opposition to his awkwardness in London. When Richard unreluctantly visits a jewelry shop with Hugh Whitebread on Conduit street on their way back from Lady Bruton's luncheon in Mayfair, for instance, he feels old and "torpid,'' unable to "think or move.''

With Peter, Woolf goes further. Like Maisie Johnson, Peter Walsh sees through other people's queerness. Elizabeth's bisexuality is remarkably hinted at by Peter's observation: "She's a queer-looking girl, [Peter] thought, suddenly remembering Elizabeth as she came into the room and stood by her mother" (Woolf 1981, 56). Woolf also constructs Peter as a rebellious queer child who takes pleasure in cruising through the city and refuses to conform to normative developmental stages, by stubbornly holding onto his "youth." Notably, as we already saw, the term "young" is identified as one of the closest terms to "queer" in Woolf's model.

> & Peter Walsh, thought Peter, I haven't felt so young for years, thought Peter; & yet he was no child could have had yet it was not youth, young, this feeling of irresponsible adventure; rather it was (not) a child's feeling: but a man's; & it & not a normal man's but.... a queer man's.... who after being wound himself about with ties & responsib[ilities] duties, burdens, & privileges, suddenly perceives their vanity (&) his freedom, as a child.... but only for a moment. <second> (Woolf and Wussow 1996, 15).

Another instance where the queer child in Peter is tellingly evoked is when Clarissa meets Peter after 30 years, she thinks "Exactly the same....; the same queer look; the same check suit; a little out of the straight his face is, a little thinner, dryer, perhaps, but he looks awfully well, and just the same" (Woolf 1981, 40).

We can locate another important theme that runs through Woolf's works when looking closely at a subset of words that models on Woolf each identify as a word close to "queer": "painter," "dressmaker," "craftsman," and "archaeologist." Indeed, in *A Room of One's Own*, *To the Lighthouse*, *Three Guineas*, *The Years*, and "Craftsmanship," Woolf deploys "queer" to imagine women author tropes – novelist, painter, archeologist, and dressmaker – who work to uncover the truth beyond the established "archives and repositories of knowledge" by reading between the lines of "patriarchal discourse" (Kaufman 2018, 333).

Elsewhere, the term "queer" is evoked to represent a beautiful harmony made out of incompatible things in life: "The voices of birds and the sound of wheels chime and chatter in a queer harmony, grow louder and louder and the sleeper feels himself drawing to the shores of life" (Woolf 1981, 69). In *Orlando*, "queer" is used to imply the spontaneous, private, and fictional side of all sort of things with respect to their factual, public, normative sides:

> Nature, who has played so many queer tricks upon us, making us so un-equally of clay and diamonds, of rainbow and granite, and stuffed them into a case, often of the most incongruous, for the poet has a butcher's face and the butcher a poet's; nature, who delights in muddle and mystery, so that even now (the first of November 1927) we know not why we go upstairs (Woolf 1928, 58).

It is notable to note that "diamonds" and "rainbow" are words identified as close to "queer" in certain model iterations on Woolf. "Diamonds" is also a recurring trope in *To the Lighthouse*, which signifies security and privacy out of sync with publicity. So is "rainbow" in *Orlando* and "New Biography," which is directly placed in sharp

opposition to cold facts, public school, and diplomacy. For Woolf, "queer'' is almost always placed in fierce confrontation with normativity.

6. Conclusion

As the above analyses and discussion demonstrate, I was able to statistically prove my thesis that "queer'' is more positive than negative for Woolf, and that Woolf's idea of "queerness'' was progressive, a thesis that would otherwise rely solely on interpretation. The top 100 words closest to "queer'' that the model on Woolf extracts turned out to be also extremely useful, when used to aid close reading of the author's works. I hope my paper helps identify a space where data science and the Humanities can be brought together to enrich Digital Humanities.

7. Data Availability

Data can be found here: `https://github.com/heejoungs/woolf_queer`.

8. Software Availability

Software can be found here: `https://github.com/heejoungs/woolf_queer`.

9. Acknowledgements

I greatly appreciate Ted Underwood, Professor of Information Science and English at the University of Illinois at Urbana-Champaign, for providing technical advice on many aspects of this research. I am very thankful for his insight and generosity.

10. Author Contributions

Heejoung Shin: Conceptualization, Writing – original draft

References

Arindam, Paul (2019). "Do I Have to Remove Stop Words in in Order to Train Word Vectors Using Word Embedding?" In: *Quora*. `https://www.quora.com/Do-I-have-to-remove-stop-words-in-order-to-train-word-vectors-using-word-embedding` (visited on 11/23/2022).

Barker, Meg-John and Julia Scheele (2016). *Queer: A Graphic History*. Icon.

Boag, Peter (2021). "Gay and Lesbian Rights Movement". In: *Oregon Encyclopedia*. `https://www.oregonencyclopedia.org/articles/gay_lesbian_rights_movement/` (visited on 11/23/2022).

Fitzgerald, F. Scott (1989). *The Short Stories of F. Scott Fitzgerald*. Scribner.

Hamilton, William L., Jure Leskovec, and Dan Jurafsky (2016a). "Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change". In: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*. Vol. 1: Long Papers, 1489–1501. `10.18653/v1/P16-1141`.

— (2016b). "HistWords: Word Embeddings for Historical Text". In: *The Stanford NLP Group*. `https://nlp.stanford.edu/projects/histwords/` (visited on 11/23/2022).

Houlbrook, Matt (2005). *Queer London: Perils and Pleasures in the Sexual Metropolis, 1918-1957*. University of Chicago Press.

Kaufman, Mark David (2018). "True Lies: Virginia Woolf, Espionage, and Feminist Agency". In: *Twentieth-Century Literature* 3 (64), 317–346. `10.1215/0041462X-7142072`.

Lamos, Colleen (1998). *Deviant Modernism: Sexual and Textural Errancy in T. S. Eliot, James Joyce, and Marcel Proust*. Cambridge University Press.

Lee, Hermione (1997). *Virginia Woolf*. Alfred A. Knopf.

Liu, Bing, Minqing Hu, and Junsheng Cheng (2005). "Opinion Observer: Analyzing and Comparing Opinions on the Web". In: *Proceedings of the 14th International World Wide Web conference* (*WWW-2005*). `https://www.cs.uic.edu/~liub/publications/www05-p536.pdf` (visited on 11/23/2022).

Norris, David (1994). "The "Unhappy Mania" and Mr. Bloom's Cigar: Homosexuality in the Works of James Joyce". In: *James Joyce Quarterly* 3 (31), 357–373. `https://www.jstor.org/stable/25473572` (visited on 11/23/2022).

Queer (2022). In: *HathiTrust+BookWorm*. HathiTrust Research Center. `https://bookworm.htrc.illinois.edu/develop/#?%7B%22search_limits%22:%5B%7B%22word%22:%5B%22queer%22%5D,%22date_year%22:%7B%22$gte%22:1760,%22$lte%22:2010%7D%7D%5D%7D` (visited on 11/23/2022).

Queer (2022). In: *Google Books Ngram Viewer*. Google. `https://books.google.com/ngrams/graph?content=queer&year%5C%5Fstart=1800&year%5C%5Fend=2019&corpus=27&smoothing=3&direct%5C%5Furl=t1%3B%2Cqueer%3B%2Cc0#t1%3B%2Cqueer%3B%2Cc0` (visited on 11/23/2022).

Řehůřek, Radim (2022). *Gensim: Topic Modeling for Humans*. `https://radimrehurek.com/gensim/` (visited on 11/23/2022).

Rubin, Gayle (1975). "The Traffic in Women: Notes Toward a Political Economy of Sex". In: *Toward an Anthropology of Women*. Ed. by Rayna R. Reiter. Monthly Review Press, 157–210.

Schmidt, Ben (2015). "Vector Space Models for the Digital Humanities". In: *Ben's Bookworm Blog*. `http://bookworm.benschmidt.org/posts/2015-10-25-Word-Embeddings.html` (visited on 11/23/2022).

— (2017). "Word2Vec Workshop". In: *benschmidt.org*. `http://benschmidt.org/slides/Word2Vec%5C%5FWorkshop.html` (visited on 11/23/2022).

Torgovnick, Marianna (1991). *Gone Primitive: Savage Intellects, Modern Lives*. University of Chicago Press.

Valente, Joseph (2004). *Joyce and Sexuality*. The Cambridge Companion to James Joyce. Cambridge University Press.

Wendlandt (Burdick), Laura, Jonathan K. Kummerfeld, and Rada Mihalcea (2018). "Factors Influencing the Surprising Instability of Word Embeddings". In: *Proceedings*

of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (*NAACL-HLT*). Vol. 1: Long Papers. Association for Computational Linguistics, 2092–2102. 10.18653/v1/N18-1190.

Woolf, Virginia (1928). *Orlando: A Biography*. Harcourt.

— (1981). *Mrs. Dalloway*. Harcourt.

— (2018). *Virginia Woolf: The Complete Works*. MyBooks Classics.

Woolf, Virginia and Helen Wussow (1996). *Virginia Woolf "The Hours": the British Museum Manuscript of Mrs. Dalloway*. Pace University Press.

A. Appendix: Complete List of Literary Texts Used in this Study from Project Gutenberg Australia

Fitzgerald, F. Scott. "The Adjuster." 1926.

———. "The Complete Pat Hobby Stories." 1940-41.

———. *Collected Stories*.

———. *The Great Gatsby*. 1944.

———. "The Guest in Room Nineteen." 1937.

———. "Hot and Cold Blood." 1926.

———. "Presumption." 1926.

———. "The Pusher-in-the-Face." 1925.

———. "Shaggy's Morning." 1935.

———. "A Snobbish Story." 1930.

———. "Strange Sanctuary." 1939.

———. *Tender is the Night*. 1933.

———. "Three Acts of Music." 1936.

———. "Too Cute for Words." 1936.

Joyce, James. *Dubliners*. 1914.

———. *A Portrait of the Artist as a Young Man*. 1916.

———. *Ulysses*. 1922.

Lawrence. D. H. *Aaron's Rod*. 1922.

———. *Amores: Poems*. 1916

———. *Birds, Beasts and Flowers*. 1923.

———. *Bay: A Book of Poems*. 1919.

———. *The Captain's Doll*. 1923.

———. *Collected Short Stories*.

———. *A Collier's Friday Night*. 1934.

———. *The Daughter-in-law*. 1912.

———. *David*. 1926.

———. *England My England*. 1922.

———. *Etruscan Places*. 1932

———. *Fantasia of the Unconscious*. 1922

———. *The Fight for Barbara*. 1912.

———. *The Fox*. 1923.

———. *Kangaroo*. 1923.

———. *Lady Chatterley's Lover*. 1928.

———. *Look! We Have Come Through!* 1917.

———. *The Lost Girl*. 1920.

———. *The Ladybird*. 1923.

———. *The Man Who Died*. 1929.

———. *The Married Man*. 1926.

———. *The Merry-go-round*. 1912.

———. *Mornings in Mexico*. 1927.

———. *New Poems*. 1918.

———. *The Plumed Serpent*. 1926.

———. *The Prussian Officer and Other Stories*. 1914.

——. *The Rainbow*. 1926.

——. *St Mawr*. 1925.

——. *Sea and Sardinia*. 1921.

——. *Sons and Lovers*. 1913.

——. *Tortoises*. 1921.

——. *Touch and Go*. 1920.

——. *The Trespasser*. 1912.

——. *Twilight in Italy*. 1916

——. *The Virgin and the Gypsy*. 1930.

——. *The White Peacock*. 1911.

——. *The Widowing of Mrs. Holroyd*. 1914.

——. *The Woman Who Rode Away and Other Stories*. 1928.

——. *Women in Love*. 1920.

Mansfield, Katherine. *Bliss and Other Stories*. 1920.

——. *The Doves' Nest, and Other Stories*. 1923.

——. *The Garden Party and Other Stories*. 1922.

——. *In a German Pension*. 1911

——. *Something Childish and Other Stories*. 1924.

Stein, Gertrude. *The Autobiography of Alice B. Toklas*. 1933.

——. *The Making of Americans*. 1925.

——. *Geography and Plays*. 1922.

——. *Three Lives*. 1909.

Woolf, Virginia. *Between the Acts*. 1941.

——. *Collected Essays*.

——. *Collected Short Stories*.

——. *The Common Reader*. 1925.

——. *The Common Reader Second Series*. 1935.

——. *The Death of the Moth and Other Essays*.

——. *Flush: A Biography*. 1933.

——. *The Haunted House and Other Short Stories*.

——. *Jacob's Room*. 1922.

——. *The Moment and Other Essays*. 1947.

——. *Monday or Tuesday*. 1921.

——. *Mrs. Dalloway*. 1925.

——. *Night and Day*. 1919.

——. *Mrs. Dalloway*. 1925.

——. *Orlando: A Biography*. 1928.

——. *A Room of One's Own*. 1929.

——. *To the Lighthouse*. 1927.

——. *Three Guineas*. 1938.

——. *The Voyage Out*. 1915.

——. *Walter Sickert: A Conversation*. 1934.

——. *The Waves*. 1931.

——. *The Years*. 1937.