The Ghost of Magical Realism in World Literature

Stefano Cagnato
Advisor: Dr. Hoyt Long
Preceptor: Tristan Schweiger

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF ARTS

MASTER OF ARTS PROGRAM IN THE HUMANITIES
UNIVERSITY OF CHICAGO
MAY 20, 2016

Table of Contents

The Ghost of Magical Realism in World Literature **3**
Introduction *4*
Outlining the Mode *5*
Measuring the Mode *12*
Quantifying Distance *16*
Topic Modeling *19*
Principal Component Analysis *23*
Lo real maravilloso *26*
Possibilities for Further Study *30*

Bibliography **33**

Appendix **37**
A. Details on Corpora *37*
B. Topic Modeling Results *38*
C. Top Words *42*

The Ghost of Magical Realism in World Literature

Introduction

Defining magical realism has proved difficult for scholars[1]. Still, most scholars can agree on the reproducibility of magical realism—that is, that is the mode's ability to be replicated and recognized by writers and readers around the world. Mariano Siskind believes this reproducibility demonstrates either a shared sense of the supernatural or the shared conditions of colonialism. Scholars such as Wendy Faris and Kim Sasser have determined magical realist texts share reproducible stylistic qualities such as the disturbance of space and time. Detecting magical realism is perhaps more difficult than detecting other more established modes or genres because the mode's stylistic qualities are not necessarily tied to quantifiable lexical parameters. And still, the popularity and widespread appeal of the mode induce critics to wonder whether there are ghostlike qualities to magical realism—qualities that can be felt and experienced but cannot be articulated.

In this vein, I propose to test the hypothesis that magical realism has inherent quantifiable characteristics unseen to the naked eye. To test this hypothesis, I conduct a combination of unsupervised computational methods such as topic modeling and text mining, along with a selective close reading. Unsupervised methods are better suited to find undiscovered characteristics. Using a distant reading approach, this study found themes of death and violence, along with a transition into urbanized settings, as additional potential markers of magical realism. As such, this study aims to understand magical realism as a mode of resistance across cultures and historical backgrounds.

Rather than offer an irrefutable solution, it is my aim to give credibility to computational methods within the magical realism archive. Text mining and topic modeling have been applied to several genres and modes of literature. However, this project marks the first time digital humanities

---

[1] Magical realism has been a source of debate from the moment Frank Roh first applied the term (*Magischer Realismus*) to art criticism in 1925. The debate, from a global-historical perspective, stems from the misrecognition of magical realism as part of other genres such as fantasy and *lo real maravilloso*. Another issue arises from the Latin American Boom's publishing houses knowingly miscategorizing texts as "magical realism" because it would increase sales. For a detailed history of the mode's transition from Roh through García Márquez onto contemporary world literature, see Reeds and Siskind. For a detailed study on the Latin American Boom and its economic enterprise, see Cohn and Rama.

studies have been applied specifically to magical realism. My intervention will lay a scholarly foundation for the analysis of magical realist texts on a larger level and allow the emergence of a community dedicated to prove that the empirical form of magical realism is the same across regional and global versions, and across historical and cultural contexts.

Outlining the Mode

The driving force behind this study is to find an appropriately empirical formulation of magical realism in order to facilitate a reading of multiple magical realist novels. Though most scholars prefer close readings that privilege historical and cultural context, a list of characteristics of the mode that can be replicated is more useful for a study that involves magical realist novels from different cultural backgrounds. However, due to recent scholarship on the mode, identifying this list can be difficult. Wendy Faris and Kim Sasser have produced viable attempts at such a list of characteristics, though not without repercussions. Still, deciding on an all-encompassing list of characteristics for the mode is imperative to ground the computational methods into a critical and theoretical framework.

The confusion surrounding magical realism stems from the claim by Latin American authors and scholars that the mode was autochthonous to the continent. Magical realism was not a way to describe the new writing of Latin America but rather a way to separate Latin America culturally from Europe and the United States. The claim was an attempt toward a cultural emancipation in order to demonstrate that Latin America was culturally relevant and independent. In this search for legitimacy, however, Latin American critics overemphasized the continent's claim over magical realism. In 1955, Ángel Flores ended an article on magical realism with the following phrase: "Latin America now possesses an authentic expression, one that is uniquely civilized, exciting, and, let us hope, perennial."[2] Amaryll Chanady believes this line of thinking to be a conscious "territorializing"

---

[2] Flores, 192.

of magical realism. This claim rejects "the hierarchical dichotomy between civilization and barbarism...by demonstrating the acceptability of Latin American literature in its present state of evolution within the universal canon."[3] Though this claim still has some supporters, the international appeal of such a reproducible mode creates an understanding of magical realism as unconquerable, a realm where flags cannot be planted.

> Is magical realism a universal aesthetic that unveils the supernatural core of the real everywhere thanks to its universal antipositivistic (transhistorical, and often ahistorical) appeal? Or is it an aesthetic that belongs organically to marginal cultures marked by traumatic collective experiences of oppression (colonial or otherwise)?[4]

Siskind invites a reading where the mode belongs to cultures oppressed by colonialism rather than to a specific geographical location. On the other hand, Carpentier's ontological claim that the reality of Latin America is already magical[5], along with the reproducibility of the mode, harkens back to the popular notion of the supernatural and its universal appeal. In 1995, Lois Parkinson Zamora and Wendy Faris describe magical realism as an "international commodity"[6] in their landmark collection of essays on the mode, thus ending the "Latin American monopoly"[7] of such an eristic term.

Another important aspect of current scholarship on magical realism is that it is understood as a mode rather than genre. Considering magical realism as a genre is a miscategorization because magical realism exists under multiple conventions such as postcolonial and postmodernist literature. Magical realism is not exclusively tied to any of these conventions. Furthermore, placing magical realism within these conventions creates more confusion due to the many understandings of

---

[3] Chanady, 128.

[4] Siskind, 62-63.

[5] Carpentier argues that the magical element of magical realism was not epistemological but rather ontological. I elaborate on this point in a later section on *lo real maravilloso*. For more information on Carpentier's claim, see Reeds 182.

[6] Zamora and Faris, 1.

[7] Zamora and Faris, 2.

postcolonial and postmodern literature. As such, in contemporary scholarship, magical realism is considered as a literary mode. The identification of magical realism as a mode occurred after it was theorized that magical realism was not exclusively tied to Latin America. In the past twenty years, magical realism has made a definite transition into other cultural and historical settings. If magical realism is referred to as a genre, it is only to draw a distinction between Latin American magical realism (which is sometimes considered its own literary genre) and other more contemporary forms of the mode. By the time of Zamora and Faris' publication, magical realism had already been linked with postcolonialism, along with postmodernism and counterrealism. Although scholarship has worked to create a more global taxonomy of magical realism, scholars still cannot come to a consensus on how to define the mode.

Understanding magical realism "relies upon the prior understanding of what is meant by 'magic' and what is meant by 'realism'."[8] Zamora and Faris argue that realism often "functions ideologically and hegemonically"[9] and promotes an objective representation of reality. However, in magical realism, realism is more open to interpretation and is thus "not centralizing but eccentric."[10] Bowers discusses two types of literary realism: one that stems from Aristotelian mimesis and the expression of fictionalized real events; and one where realism is possible not because it describes real events but rather because it is made up of elements that are familiar in our understanding of language. "Magical realism therefore relies upon realism but only so that it can stretch what is acceptable as real to its limits."[11] The realism expressed in magical realism is made up of familiar elements while simultaneously challenging what those elements can express.

---

[8] Bowers, 20.

[9] Zamora and Faris, 3.

[10] Zamora and Faris, 3

[11] Bowers, 22.

The element of magic is more nuanced and subtle, though we can begin with the premise that magic is any supernatural phenomenon that occurs in a text. Bowers claims magic in magical realism should not overextend into the realist aspects. Allegories or fables are often not considered magical realism because they privilege the magic in the narrative over the real. Hunter S. Thompson's hallucinatory tales remind the reader that the fantastical elements in his prose are a figment of a drug-induced imagination. Likewise, Alice's realization that her journey to Wonderland was all a dream gives privilege to the real. This idea comes from Ato Quayson's "principle of equivalence," which states that though magical realism shares elements with other modes and genres, it "generates a scrupulous equivalence between the two domains."[12] In other words, giving privilege to either the realist or the magical aspects draw us farther away from the liminal nature of magical realism.

Once each term is defined separately, the question of what exactly magical realism is remains a divisive query. Stylistically, magical realism is often characterized by a socially realistic setting, fantastical elements, authorial reticence, political and social critique, metafiction, hybridity, and awareness of mystery. Christopher Warnes aims to define the mode from a close reading perspective, asking questions specific to the texts: "what is the source of the supernatural in this text? What kind of dialogue does it enter into with the text's more realistic elements? What literary, cultural, and historical factors lead to the appearance of this narrative strategy in this particular work of literature?"[13] However, when considering magical realism's *reproducible* qualities, the most important hermeneutic aid is Irlemar Chiampi's notion that magical realism colloquializes the supernatural while concurrently exalting the mundane. As Zamora and Faris state, "magic is no longer quixotic madness, but normative and normalizing."[14] In magical realist texts, the magical elements are not uncomplicated but they are commonplace.

---

[12] Quayson, 728.

[13] Warnes, 153.

[14] Zamora and Faris, 3.

Still, these stylistic characteristics often appear in dissimilar quantities and in myriad ways. The varying literary characteristics that are present discourage a specific set of guidelines and, as such, scholars are reluctant to apply a global definition. As of now, there seem to be two different ways in which magical realism is defined: anthropological and formalist. The anthropological approach analyzes the historical context surrounding the novel at hand. These accounts often produce a definition of magical realism that is deeply rooted in cultural context and aims to describe the culture that produces the mode rather than the mode itself. While anthropological approaches are valuable for historical analyses, they often fall short when producing a strict list of characteristics that belong to the mode. On the other hand, the formalist approach is concerned with the linguistic elements of a text and how these formal linguistic choices constitute a distinct mode. This approach allows us to focus not so much on the cultural details of magical realist works, but rather on the reproducible qualities of the mode.

Wendy Faris provides a reproducible list of markers concerning magical realism that allows for an empirical analysis of the mode. Using her framework encourages a discussion of the mode that is not confined to the genre or context of the novel.

> First, the text contains an "irreducible element" of magic; second, the descriptions in magical realism detail a strong presence of the phenomenal world; third, the reader may experience some unsettling doubts in the effort to reconcile two contradictory understandings of events; fourth, the narrative merges different realms; and, finally, magical realism disturbs received ideas about time, space, and identity.[15]

Warnes considers only the first two as essential.[16] Nevertheless, Faris' reproducible list of characteristics is ideal for a computational analysis of the shared features throughout the many types of global expressions of the mode. Faris' "clear and unambiguous definition"[17] provides a secular

---

[15] Faris, 7.

[16] Warnes, 5.

[17] Warnes, 5.

set of characteristics that allow us to begin working with the mode. The aim is not to strip texts of their inherent cultural qualities; rather, the aim is to analyze how magical realism operates as a technique in fiction *independent* of cultural context. Surely there must be a compromise—a claim that doesn't acknowledge the cultural context of a novel cannot then posit a thorough critical approach.[18]

I insist on treating magical realism independent of cultural context because it allows for us to focus solely on the reproducibility of magical realism. This reproducibility has led to the mode becoming "the literary language of the postcolonial world."[19] Mariano Siskind asks what it takes for a genre (or mode) to become a global form of literary expression. Siskind argues that, "as evidenced by its ubiquity," magical realism "is the most established and stable world literary genre."[20] Clearly, he posits, the mode must have reproducible qualities, a set of parameters that can be applied in different cultural contexts. Furthermore, the mode should be inherently "suited to exploring—and transgressing—boundaries."[21] This assessment of magical realism is a testament to its reproducibility. Not only are the formal aspects of magical realism easy to replicate; the cultural conditions under which magical realism flourishes are also already reproduced throughout the world.

It is difficult to ignore the relationship between the conditions of colonial societies and the places where magical realism flourishes. Siskind would argue the transformation of magical realism from localized genre to global literature mode can be condensed to an argument about shared colonialism and the conditions which it includes:

> a world in which religion and superstition dominated people's lives; also a world in which there was a powerful and complicated history of colonialism; also a world in which there were colossal differences between the very poor and the very rich, and not much in between; also a world bedeviled by dictators

---

[18] Faris has been faulted for taking too much of a distant approach. Scholars who disagree with her argument claim that magical realism is closely tied to cultural and historical context.

[19] Babha, 7.

[20] Siskind, 59.

[21] Zamora and Faris, 5.

and corruption.[22]

Magical realism works well to discuss issues of oppression at a more universal level. The magical elements allow authors to conceal allegorical messages critiquing socioeconomic or political elements of culture. Warnes convincingly asserts that magical realism also originates from violence and cruelty.[23] The postcolonial understanding of magical realism—drawn from the conditions in which the mode often arises—has changed the magical realist scholarly landscape for the better. This connection proves that magical realism is not endemic but rather worldly, global. Postcolonialism provides a seemingly tight-knit definition for a cultural perspective on magical realism.

However, categorizing magical realism within postcolonial theory can be problematic. Kim Sasser warns us not to fall under the spell of postcolonial binaries of the rational colonizers and the supernatural colonized.[24] Using a flexible set of parameters such as Faris' five-part definition "one can remove magical realism from underneath any one particular overlay, such as the postcolonial and postmodernist politics of resistance…while recognizing the significant role these usages have played in the mode's history and development."[25] Thus, not only can magical realism be treated without cultural context—it can also be considered without any pre-existing attachment to other literary movements, genres, or agendas. (To be clear, the point here is not to strip cultural context from magical realist novels; rather, this study seeks to achieve some distance from existing forms of reading while acknowledging that magical realism exists within these forms.) But, Sasser asks, "does it detract from magical realism's historically postcolonial usages to argue…that the mode is capable

---

[22] From "Inverted Realism", a talk given by Salman Rushdie at a PEN Tribute to Gabriel García Márquez.

[23] Warnes, 151.

[24] Sasser, 23.

[25] Sasser, 24-25.

of exceeding a prescriptive type of interpretation?"[26] This is an important consideration; we must acknowledge that quantifying literature comes at a price.

Conversely, computational methods and close readings, along with the theoretical framework that has been laid out, give this study a better grasp of the mode's possibilities without depreciating its cultural functional space. Faris' empirical definition of magical realism works not to exclude cultural and historical context but rather to treat magical realism as a global mode of expression. Likewise, using Sasser's theoretical framework, this study uses Faris' approach to acknowledge the forms of reading, such as postcolonialism, under which magical realism flourishes.

Measuring the Mode

Now that we have delimited the term, let's turn to the question of whether or not magical realism is quantifiable. Since magical realism's reassessment from genre into mode, close readings have been more successful in drawing out the specific concerns of the mode within each novel relative to the concerns of the novel as a whole. Close readings allow for discovering connections between the novel and the cultural context surrounding the novel. Distant reading, an analysis of literary texts *en masse*, provides a removed approach. My concern is not with how the novel or magical realism function within a novel, but rather how magical realism functions in its totality. If scholars insist solely on close-reading or thinking of the mode as belonging to a specific time and place, the academic community will miss out on seeing magical realism as a mode of resistance across multiple cultures.

A text mining approach to formalism can illuminate the search for a potential way to catalogue findings at an empirical level. Text mining is a computational method used to derive quantifiable data from a text. Examples of this data range from the number of times a word is used in a novel to what types of words occur in close proximity. To capture such a large scope and make

---

[26] Sasser, 40.

it understandable to humans, distant reading through text analysis is not only viable but necessary. While a larger scale project could use supervised techniques, I am intentionally limiting my study to unsupervised techniques. Unsupervised techniques are often used in exploratory projects—they find patterns without any human intervention. This framework will allow future scholars to expand this untapped field, especially if they were to use supervised techniques. Furthermore, unsupervised techniques allow for the discovery of potentially overlooked characteristics or patterns, those ghostlike characteristics this study is interested in.

At this time, most stylistic analyses have focused on genres rather than modes. Jockers argues that genres might be distinguished through the analysis of specific features such as pronouns, articles, and other words with little lexical meaning more commonly referred to as function words. Other attempts have been made to differentiate genres from a group of texts, such as Moretti's genre analysis of Gothic literature and Michael Witmore and Jonathan Hope's genre categorization of Shakespeare plays[27]. Witmore was able to detect whether a Shakespearean play was a comedy or a tragedy with a high percentage of accuracy. It is clear from these past experiments that genre analysis works, but one must wonder how much of the distinction is due to the conventions of genre discovery and how much is "attributable to other factors, such as author, time period, or author gender."[28] Due to these past studies, I treat magical realism here as a genre for the purposes of computational findings. Treating magical realism as a mode in computation analyses would conflict with the unsupervised approach this study seeks.

> If genres, like authors, appear to have distinguishable linguistic expressions that can be detected at the level of common function words, then a way was needed for disambiguating the "author" effects from those of genre and for determining whether such effects were diluting or otherwise obscuring the genre signals we observed.[29]

---

[27] See Hope and Witmore.

[28] Jockers, 68.

[29] Jockers, 70.

The "if" is important here—scholars are not entirely sure when these methods are expressing new information and when they are simply fitting neatly within a biased understanding of the texts under analysis. This problem exists every time an artistic structure endures throughout time and is reproduced in different contexts. But close reading is only one answer to this obstacle. Distant reading through text analysis and machine learning techniques is a way to analyze literary modes through fundamental configurations. These fundamental configurations exist at the lexical level. Distant reading "generates a theoretical space where apprehensions of texts as constitutive of larger structures or as radically singular can coexist, illuminating world literature as both structure and divergence, diffusion and variance."[30] Taking a distant approach to world literature by using relative word frequencies and topic modeling might illuminate the reproducibility of magical realism and global modes and genres at large.

For any genre/mode study, choosing the right texts is imperative. The corpus of this study focuses on works explicitly related to magical realism, whether through the author's statements or through published scholarly work. There are two corpora involved in this study: a Latin American corpus and an Anglophone corpus. To avoid generational bias, the Latin American corpus contains novels from 1949 to 1987, a period that encompasses the pre-*Boom* and post-*Boom* eras. This variety allows calculations about magical realism to encompass the mode from its embryonic stages into its later stages in Latin American literature heralded by *Cien años de soledad*. Likewise, all the works in the Anglophone corpus have been discussed in relation to magical realism. To ensure a similar generational unspecificity, the Anglophone corpus includes works from 1981 to 2002, a period that begins a year before Isabel Allende published *La casa de los espíritus*—that is, a period of time in which *Cien años de soledad* had become an inspiring work for other novelists with similar literary agendas. The Anglophone corpus then reaches the late 90s and early 2000s, a period where

---

[30] Long and So, 3.

multicultural and multilingual works begin to fill the fray of world literature. This variety provides the study with possibilities in finding distinctions between magical realist novels. In other words, if similarities are found within these novels (which are quite different in terms of setting, narrative style, and even length) a defining marker of the similarity could be the shared mode of magical realism.

Though there are shortcomings in using a corpus of only eighteen novels, the prestige and relevance of the novels in the academic community speak to their importance in this study. In distant reading analyses, the idea is to use computational power to upscale literary investigation. If a model is created based on a few texts, the viability of the model is tested on a larger corpus. With a small corpus, it is more difficult to catalog the efficiency of machine learning algorithms. Usually, a model that only works on eighteen novels is not considered replicable. However, the canonical nature of these texts leads me to infer that any conclusion drawn about these texts will certainly have consequences for the mode at large. In other words, a model that defines this group of texts will have repercussions beyond the scope of the corpus.

Another difficulty arises when comparing corpora in different languages. In these unsupervised methods, the algorithm is looking for relationships between what it considers to be words. These algorithms do not know *eyes* and *ojos* have the same meaning in different languages. This is for a good reason—though machine-learning algorithms have progressed in the past few years, they still cannot account for nuances in languages and translation. Therefore, the methods used in this study can only be carried out with a corpus in one language. Thus, the texts had to be separated into two corpora. Though this limitation complicates the method, I can nonetheless experiment with texts in translation in order to compare the significance of the distance measurements.

Pursuing an unsupervised approach creates limitations for this study. Despite these limitations, the findings suggest potential paths magical realist scholarship can move toward in order

to account for all the global forms of the mode. This study is not concerned with how magical realism works to critique or comment on a single cultural background. Rather, the focus of this study lies on using distant reading and computational methods to account for the mode's replications throughout the literary world. A study that includes both close and distant readings provides a better opportunity to understand magical realism as a global mode of resistance.

Quantifying Distance

As a means to discover potentially obscured patterns within magical realist literature, a distant and unsupervised approach seemed vital and pertinent. Unsupervised exploration of magical realism can lead to discoveries of new markers that may belong to an empirical model of the mode. Three methods stand out as standards in unsupervised text mining: relative word frequency, topic modeling, and principal component analysis (PCA). When analyzing relative word frequency, the novels tend to overlap, which means they are similar at a lexical level. Following this, topic modeling demonstrates the novels share topics related to family, body, violence, and time. Lastly, PCA provides a different way of looking at how the novels are related to topics. In this study, I used text mining in order to test the hypothesis of a global magical realism discoverable through quantifiable means. The methods underlined above all point to the novels in each corpus being related at a basic level via word frequency and the existence of similar topics. These similarities are not necessarily tied solely to stylistic elements of magical realism, but the similarities between these canonical novels provides a proof of concept for potential quantifiable parameters of magical realism.

A common way to find similarities and differences between novels is through the use of the shared vocabulary between the novels, otherwise known as relative word frequency. Using this approach, a scholar can use a list of words that occurs in every novel in the corpus and then find the differences in use between those novels. For example, the word "time" will occur in different frequencies in each novel. If "death" appears 100 times in Novel A and 6 times in Novel B, and a

Novel C with the word "death" 8 times is introduced, it is clear Novel C is more lexically related to Novel B. To take advantage of the computational method and to improve the power of this analysis, the list of words can be expanded to encompass a much larger part of the shared vocabulary. For this study, I am only looking at the most popular thousand shared words in my corpus. The algorithm analyzes the relative word frequencies and creates a distance matrix that calculates the distance or similarity between each text.

In preparation for this analysis, a few measures were taken in order to produce a more cohesive result. First, the texts were split into smaller sections called chunks, each 1000 words in length. This allows for me to compare shorter texts with longer texts without any discrepancy. Second, in order to focus on content words and possible themes that arise from the analysis, I removed character names and function words, such as articles and pronouns. Finally, I focus here on *relative* word frequencies. In other words, I am using the word frequency within each chunk, not the frequency within the entire novel. Analyzing relative word frequency will allow us to find similarities between novels at the lexical level.

In Figure 1, the Anglophone corpus clusters toward the center, implying a basic similarity at the lexical level. The first pattern revealed itself when plotting the distance matrix of relative word frequencies belonging to the contemporary magical realist corpus. Figure 1 demonstrates how contemporary magical realist works are relatively close, and thus similar. Though the works don't fully overlap, they do converge toward the center, which means the novels share a substantial

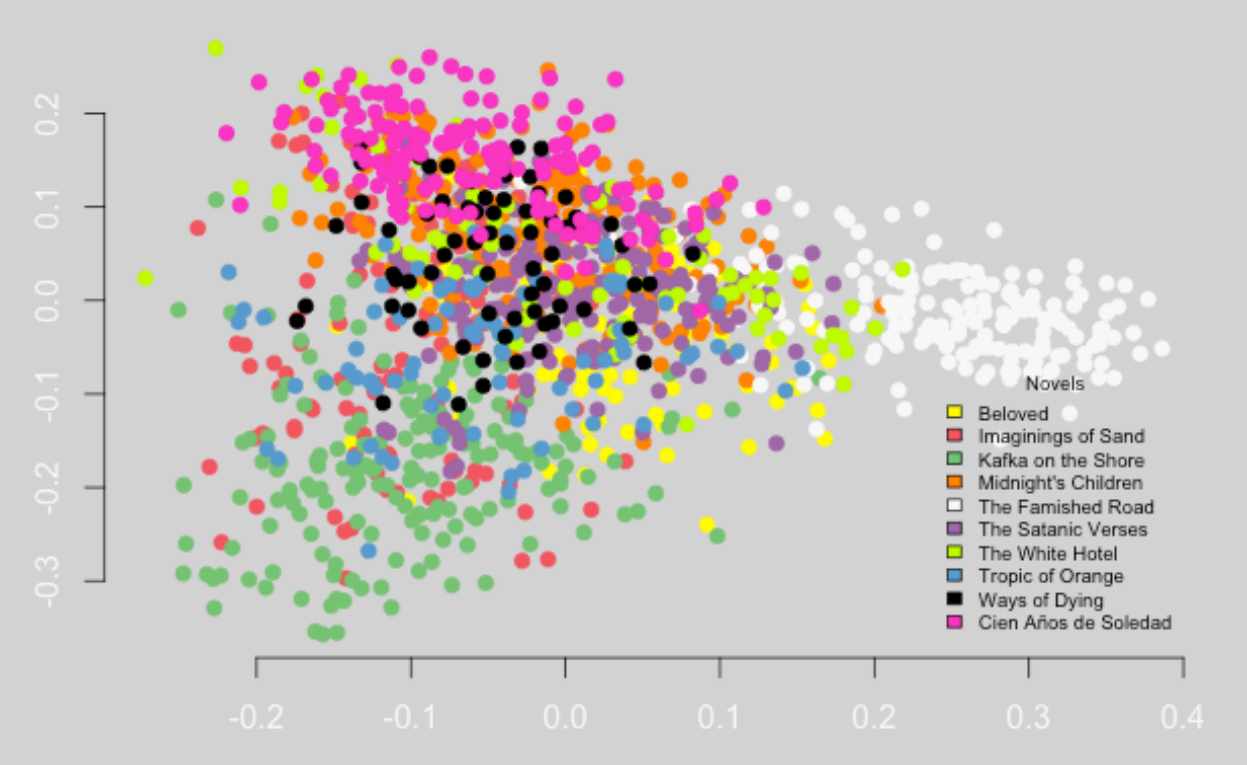


Figure 1: Anglophone corpus distance matrix including *Cien años de soledad*

amount of lexical characteristics. Curiously, when adding *Cien años de soledad* (translated into English) to the distance matrix, the novel seems to fit right in with the rest of the texts. This visualization demonstrates the possibility for the relativity of magical realism. In other words, although the corpus is small, Figure 1 implies magical realism via relative word frequencies has not changed as much as scholars might think it has. Now, we must be careful with this assertion. We must remember we are only looking at similarities through the use of a shared vocabulary.

The Latin American corpus behaves similarly. In the visualization that includes the Latin American magical realist corpus (Figure 2), the works tend to overlap. The works seem to cluster around other works produced within the same generation. As with the Anglophone corpus, a translation was introduced in order to compare where these distances would fall if I was able to plot both languages together. The translation of *The Satanic Verses* falls very much in line with the rest of the Latin American magical realist corpus, indicating the texts have a potentially significant similarity.

But the questions linger: what do these findings tell us about magical realism? What happens when non-magical realist novels are compared using these methods? A control set could give a study such as this one a more valid claim. The goal is to decidedly prove that the similarity displayed in the visualizations is specifically related to magical realist stylistic qualities. Still, text mining is immensely useful—it demonstrates that, in lexical terms, these novels are inherently related. The question remains: how are the texts related? To find the answer this study invokes other methods such as topic modeling and Principal Component Analysis.

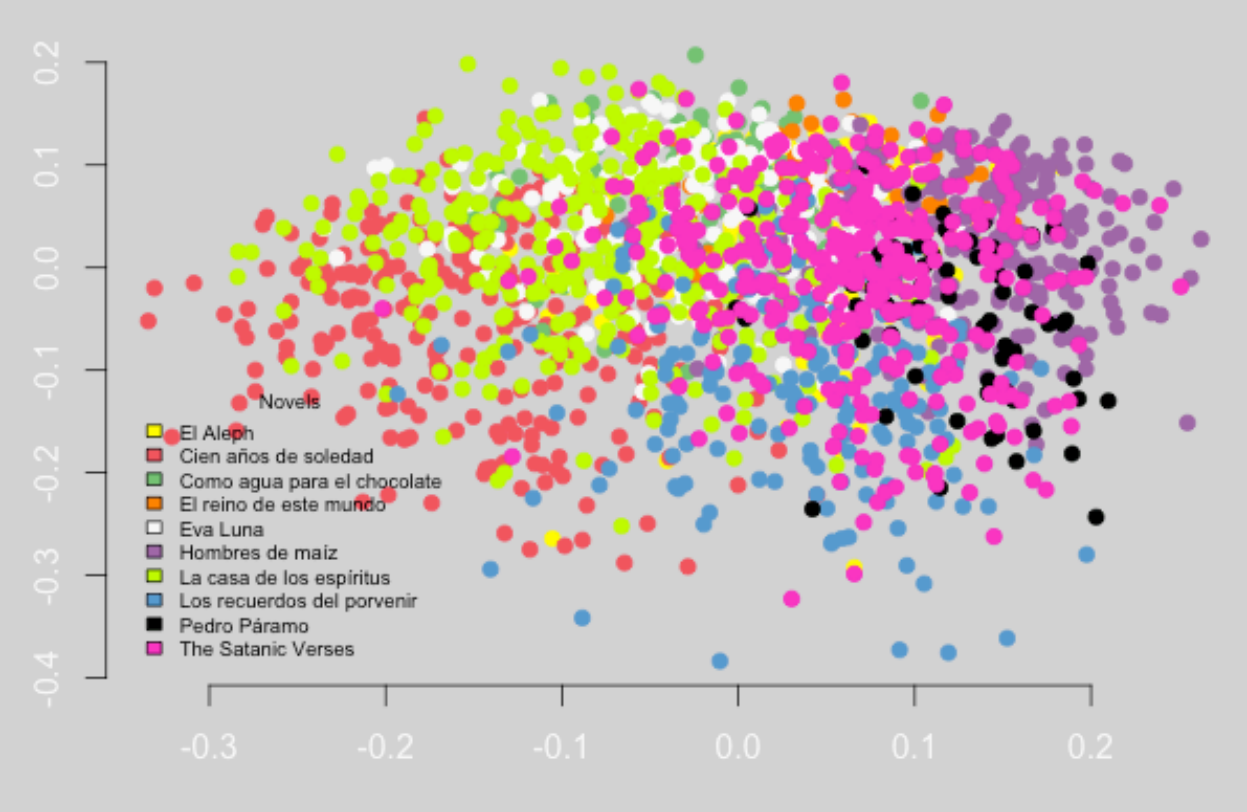


Figure 2: Latin American corpus distance matrix including *The Satanic Verses*

Topic Modeling

Another unsupervised approach to text mining is topic modeling. Here, a computer algorithm finds "topics", or lists of words that frequently co-occur. The algorithm uses statistical probability to discern whether words belong to a topic. These topics are not necessarily semantically significant—it is up to the scholar to decide how to categorize these collections of words. The

topics can then be mapped back to the original text to see where certain topics occur in a high frequency. The amount of topics the algorithm looks for is one of the major choices in topic modeling, along with the amount of iterations the algorithm goes through. For this study, I opted to use fifteen topics. Using less than fifteen topics only provided one major topic that occurred throughout, while using more than fifteen didn't produce any new results. Fifteen topics leave room for topics to emerge freely without limiting the algorithm's capacity. In addition, the algorithm went through one thousand iterations of the same texts to find frequently co-occurring words.

I specifically used topic modeling as a tool in order to provide more contextual evidence for the relativity of magical realism. Since topic modeling provides lists of words, one can then use these in order to delve into close readings or to make arguments about the shared vocabularies of the corpora. Just as with the previous method, I have provided visualizations to enhance the understanding of how these topics illustrate the relationship between the texts (the words belonging to each topic are in the appendix). The topics found by the algorithm demonstrate that these texts are related lexically in ghostlike ways that are obfuscated to human readers. Each corpus contains two topics that occur throughout the novels. Furthermore, the presence of death-related themes provides a new set of potential characteristics for the development of an empirical form of magical realism.

When running the chunked novels through the algorithm at 1000 iterations, topic modeling consistently provided two topics for each corpus that occurred in all the works. These topics were not necessarily the most common in every novel. For example, it is clear from Figure 3 that the most common topic in *The Famished Road* is Topic 2. This topic does not occur at a noticeable rate in any of the other novels. Most topics are related to a single novel due to the limited corpus size. However, Topics 6 and 7 both occur throughout the corpus. Topic 6 and 7 have words that belong to certain themes represented in the corpus. At first glance, I can denote Topic 7 the "body" topic. Words such as *eyes*, *head*, *face*, *hand*, *hair*, *mouth*, and *body* give us the sense that this topic stems from

physical descriptions of characters. However, there are some groups of words that do not fit under the "body" category. Words such as *door*, *room*, *window*, *place*, and *bed* imply a descriptive scene of a house or bedroom. Even more interesting: *Blood*, *dark*, and *dead* seem to suggest a somber scene, though the word *blood* is also related to family and lineage. Topic 6 is more difficult to decipher, though it is clear that the concepts of family and time are highly present. There's a long list of words relating to family and its members: *mother*, *woman*, *father*, *child*, *family*, and *wife*. As for temporality, words such as *time*, *year*, *old*, *day*, *end*, *remember*, and *ago* are all ways to describe temporal aspects.

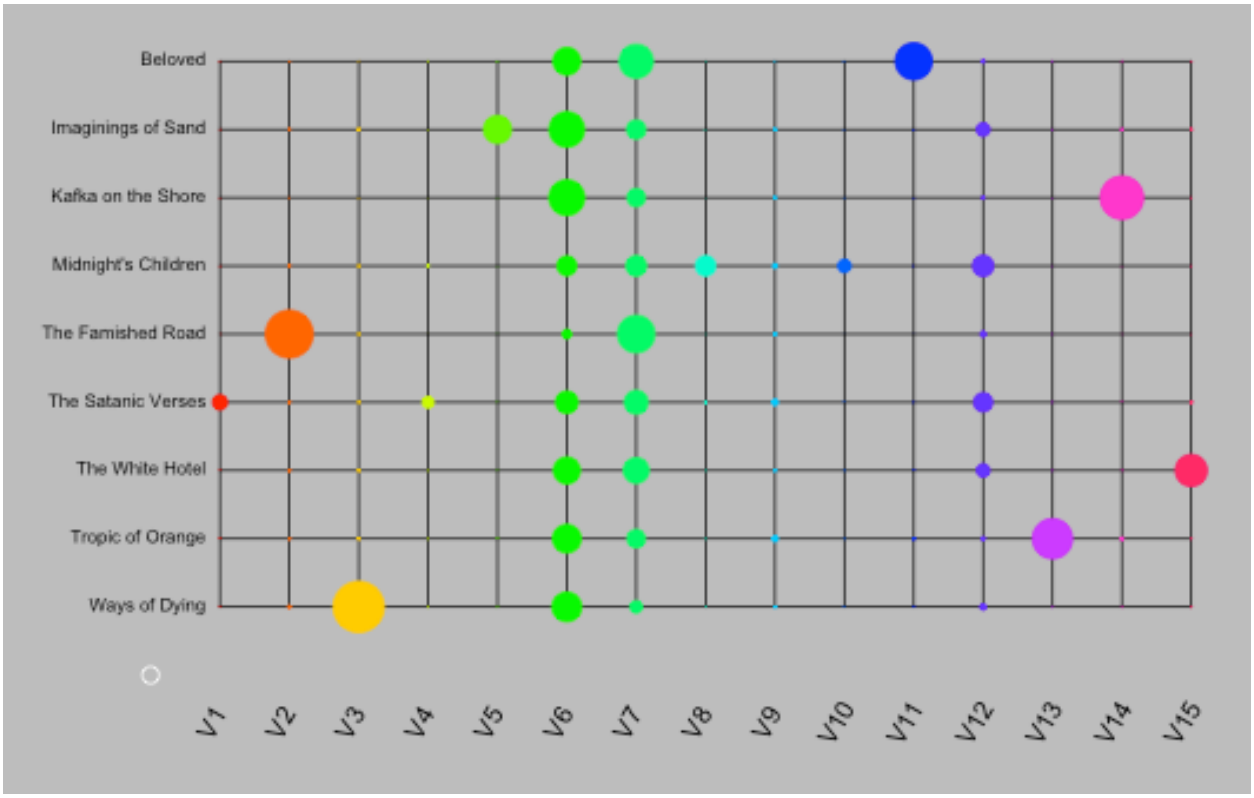


Figure 3: Topic modeling of Anglophone corpus. Topics 6 and 7 occur in all nine novels.

The Latin American corpus topic modeling behaved in a similar fashion. Two topics emerged that occurred throughout the corpus: Topics 3 and 10. Topic 10 contains very similar body parts to the topic in the Anglophone corpus: *ojos*, *cabeza, mano*, *boca, pie*. I am inclined to acknowledge this similarity as a token for fiction but not necessarily magical realism. Similarly, Topic 3 lists some

words related to family: *madre*, *familia, hijo, hija, marido, hermano*. Since magical realist novels intend to set up a realistic social environment, descriptions of character's family members is crucial. At the same time, these descriptions are common conventions in many types of fiction. More nuanced is the appearance of words related to temporal aspects in both topics. *Años*, *noche*, *día*, *tiempo*, *tarde, hora, meses,* and *pronto* are all related to the list of temporal words highlighted by the topics in the Anglophone corpus. In novels, discussing the passing of time is imperative to allow the reader to follow the action and get a sense of the pace of the unfolding events. In magical realism, this is even more important.

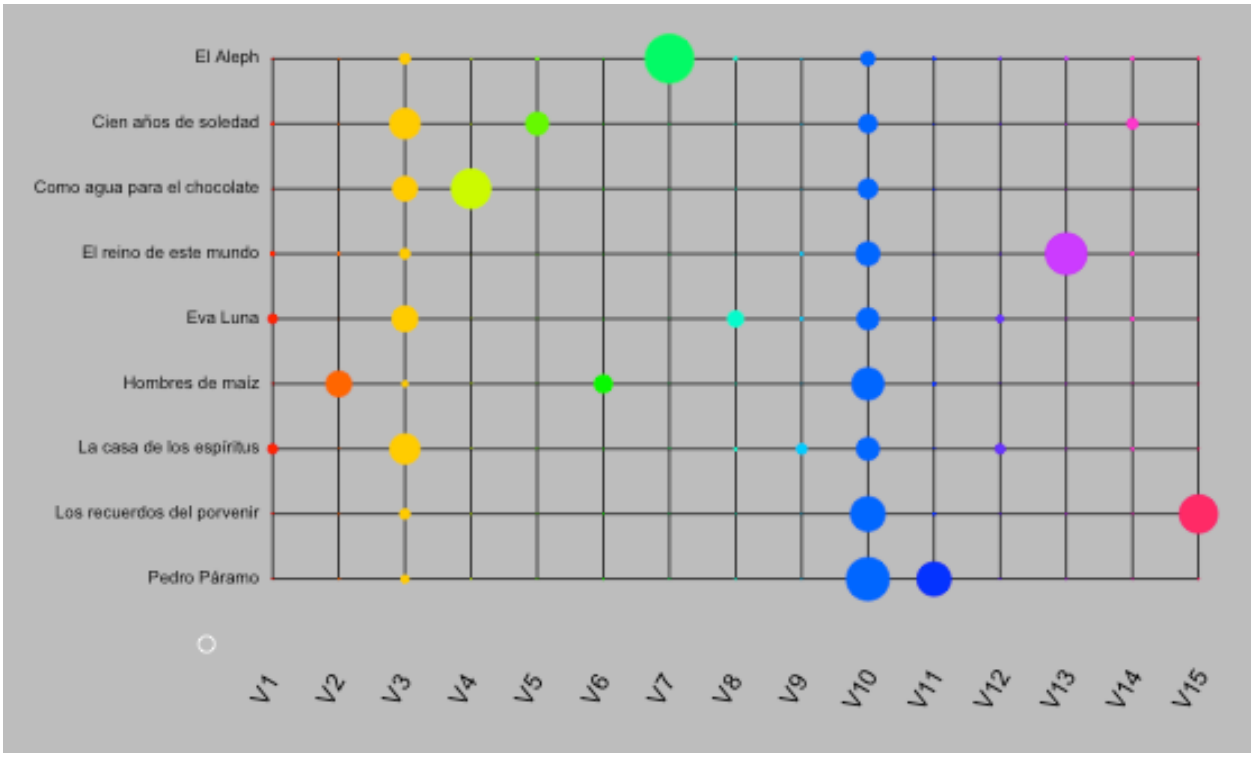


Figure 4: Topic modeling of Latin American corpus. Topics 3 and 10 occur in all nine novels.

Death and temporality are common themes in the most prevalent topics of both corpora. In topics 6 and 7 of the Anglophone corpus, words related to death and temporality occur in abundance. In topics 3 and 10 in the Latin American corpus, words like *muerto*, *sangre*, and *muerte* and the vocabulary related to temporal aspects are similar to those in the Anglophone corpus. Not only this—the top word for the Latin American corpus is *guerra*, signaling toward the violence expressed

in these novels (Appendix C). Temporality, referring back to Faris' definition, is one of the concepts which magical realism works to disturb. Death, however, is not mentioned in Faris' definition, but its inclusion as a definitive marker of magical realism is not surprising. Language of death is closely tied to Warnes' stipulation of magical realism as an exploration or commentary on violence. Due to the prevalence of death-related vocabulary, the findings suggest death and violence are as important as temporality, the body, and family to magical realist fiction. Topic modeling, then, becomes a useful tool in discerning the theme of death/violence as an undiscovered empirical and functional category within magical realism. Topic modeling confirms temporality as a marker of magical realism and introduces death and violence as potential ghostlike characteristics of the mode.

Principal Component Analysis

Another type of textual analysis mixes the distance matrix and the topic modeling approach. Principal Component Analysis (PCA) serves to demonstrate how often a topic occurs in each novel. Each PCA visualization shows the novels (in colored circles) and topics (represented by white crosshairs). The proximity of a novel to a topic in the visualization is directly related to the topic's prevalence in that novel. PCA is the ideal method to converge an analysis of relative word frequencies and topic modeling because it shows both results in one visualization.

In Figure 5, the novels are all clustered near the right side of the plot. *Imaginings of Sand, Ways of Dying, Kafka on the Shore, and Tropic of Orange* have clustered neatly near Topic 5, 3, 13, and 14.

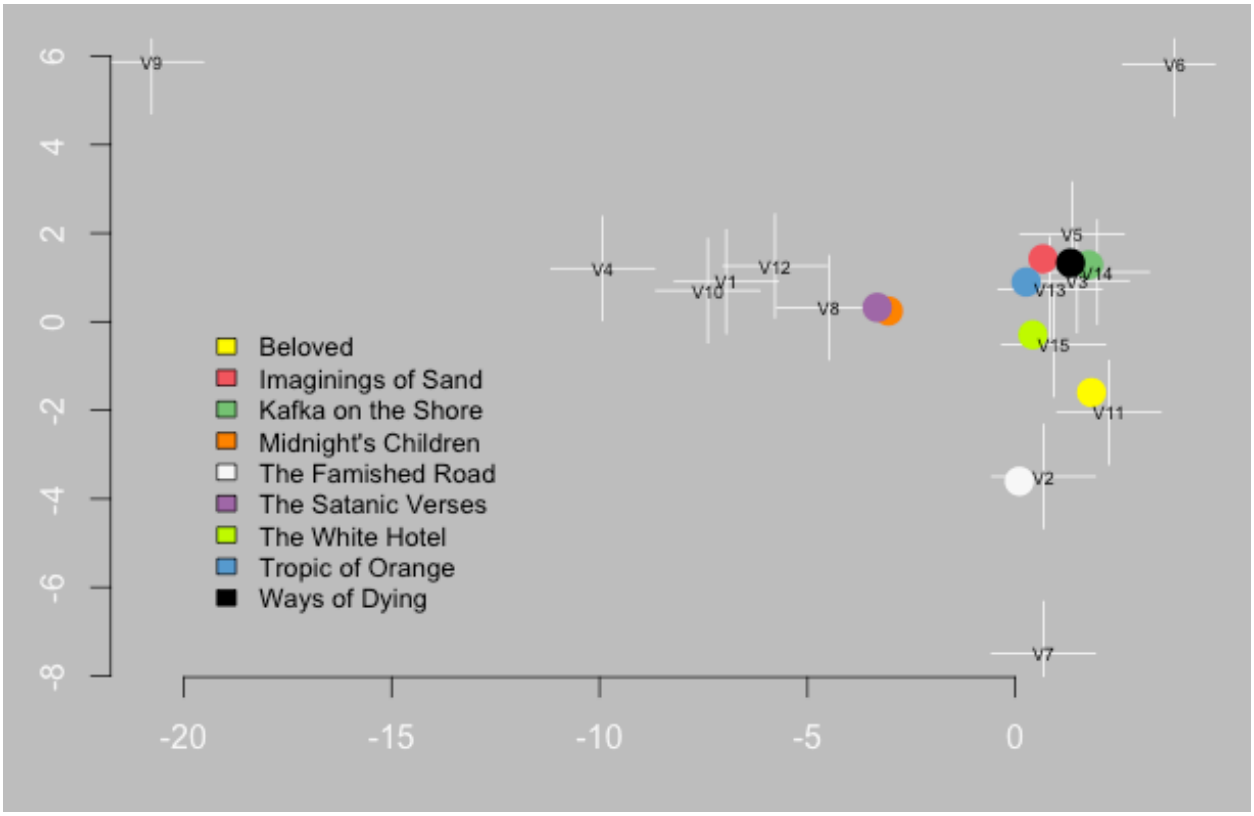


Figure 5: Anglophone corpus PCA visualization

Likewise, Both Salman Rushdie novels have clustered toward the center of the plot. Due to the small corpus size, the topic modeling algorithm found topics that occur only in one novel. Since each novel is eleven percent of the corpus and the algorithm is finding fifteen topics, it would be difficult for this not to occur. The PCA visualization makes it clear: Topic 2 corresponds to *The Famished Road*, Topic 11 to *Beloved*, and so on. While Topic 6 and 7 occur in every text, they exist in the peripheries of the corpus in the PCA visualization. This is interesting—I hypothesized that these topics, central to these novels, would have been more central to the visualizations, with the novels and other topics existing around these topics. Instead, Topic 6 and 7 surround the corpus. However, it is important to consider once again that magical realism is not a genre but a mode. Magical realism doesn't occur throughout a novel but in select places. Topic 9, which also seems to occur in every novel at a very small rate, demonstrates itself as an urban topic with words such as *car*, *city*, *crowd*,

street, *shop*, and *police*. Topic 9 exists far out toward the top left corner of the plot, creating somewhat of a triangle with Topic 6 and 7. This triangle creates a new hypothesis: along with the themes presented in Topic 6 and 7, the urban themes of Topic 9 could be another potential marker of magical realism, especially considering the mode's transition into urban settings. The distance from the novels to Topic 9 does not automatically support this hypothesis, but the corpus only contains one decidedly urban magical realist novel: *Tropic of Orange*. If the corpus were larger, perhaps urban topics would be much more of a cornerstone to the PCA visualization.

Figure 6 works a bit differently than Figure 5. Here, the novels are more widely spread out. Similar to Figure 5 and due to our small corpus size, there is an abundance of topics that only occur in one novel. When taking a look at the topics that occur in every novel (Topic 3 and 10), they are

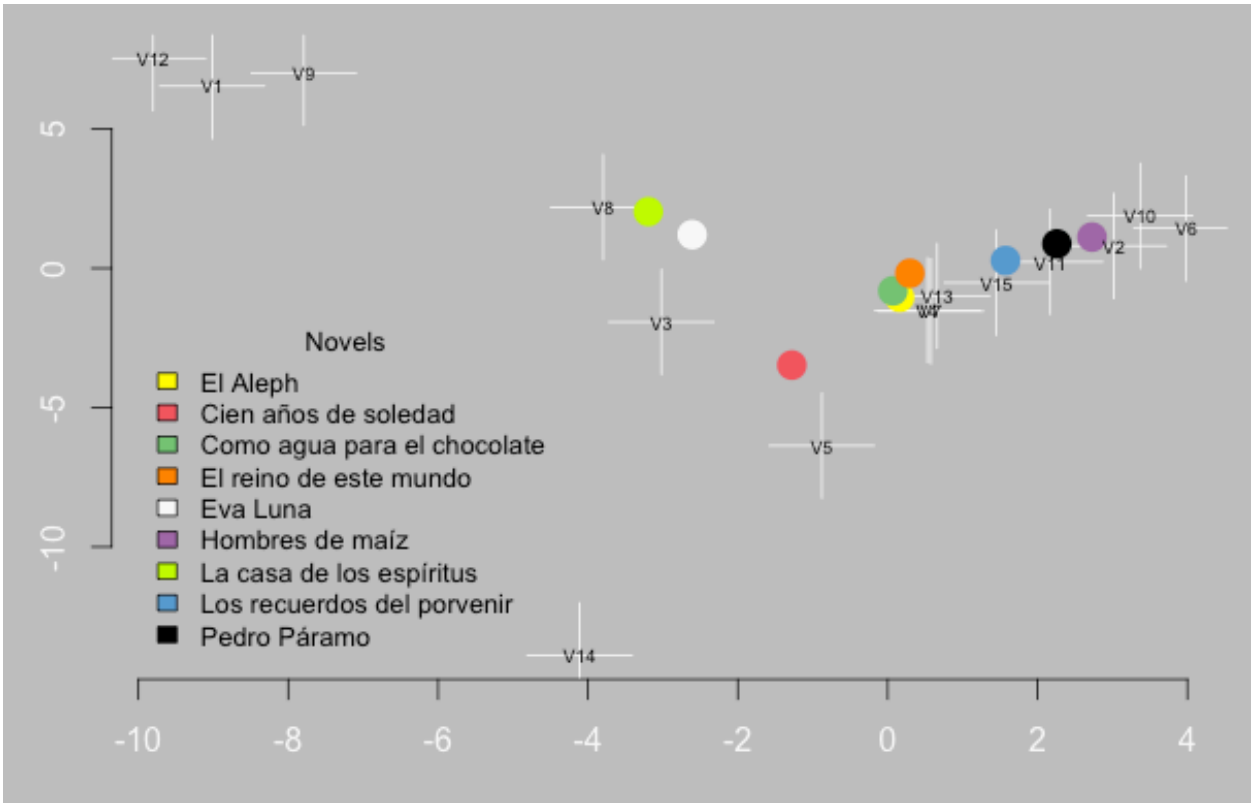


Figure 6: Latin American corpus PCA visualization

not as far from the corpus as they were in the Anglophone corpus. Topic 3 lies in the center of the plot, while Topic 10 hides toward the right end, close to *Hombres de maíz*. Interestingly, there are four

topics that are very far from all the novels; Topic 14 lies by itself toward the bottom, while Topic 12, 1, and 9 cluster together toward the top left corner.

Proximity, however, is not enough for us to draw conclusions about the mode, especially with such a small corpus. An important next step is to return to the text to find if the algorithm truly found significant similarities. The difficulty comes when interpreting what the algorithm considers a similarity. When returning to the text and comparing the chunks that fall close to the center of the cluster in Figure 1, there were noticeable similarities—however, these similarities are not necessarily related to magical realism. The chunks were mostly moments of description, and while some of the events described were indeed magical or supernatural, others were not. This is an interesting moment in our study—though the algorithm is not necessarily finding similarities between moments of supernatural acts, it is finding similarities through other lexical and content-based elements which have gone unnoticed without the use of computational tools. Though PCA results are not exceptionally encouraging, they point to a potential triangulation of topics: temporality and death, body and family, and a transition into an urban setting—a new feature in magical realist scholarship.

Lo real maravilloso

To complement this distant reading, a closer look at two of the texts would help contextualize that findings and simultaneously locate the magical realist patterns under a postcolonial framework. In this case, an analysis of *The Famished Road* seems pertinent due to the separation of the novel from the rest in Figure 1. Analyzing *The Famished Road* in closer detail will provide an explanation for the novel's separation from the rest of the corpus in Figure 1. An analysis of the novel demonstrates that categorizing this novel under magical realism can be problematic. In fact, close readings by Warnes and Sasser, along with text mining, demonstrate that *The Famished Road* and *El reino de este mundo* are lexically and conceptually more similar to each other, which can be

explained by the novels' relationship to *lo real maravilloso*, a branch of magical realism. This assertion is confirmed in part by computational tools.

In Figure 1, *The Famished Road*—published in 1991 by Nigerian writer Ben Okri—stands out as an outlier in the visualizations and text analysis. The novel, represented by white dots, stands farther out to the right than the rest of the corpus. This separation is interesting because distance implies difference. So, how is the novel different from the rest of the corpus? At first, it might seem as though this distance is created through authorial style. Magical realism and *lo real maravilloso* both offer a simple formula to augment realism, though magical realism is more concerned with an experience of the amalgamation of the magical and the real while *lo real maravilloso* claims that the reality (of Latin America) is, as confirmed by the narrative, magical. Situating the novel within "existing narrative traditions" is "one of the most pressing concerns"[31] when discussing *The Famished Road*, and magical realism is an effort in this vein. However, Okri himself denies the novel's relationship to magical realism, claiming "the problem of realism is that it does not catch the full richness of reality"[32] and that there is no "absolute reality"[33]. As Warnes points out, there are different ways of reading and placing *The Famished Road* in context. In some interpretations, the novel fits within the parameters of West African fiction and "represents the advent of a postcolonial postmodernism in African writing."[34] In others, Okri's narrative uses a localized set of mythologies to establish global strategies that defy realism, i.e. magical realism. These readings coexist within the concerns of world literature, commenting on both local and global cultures.

The Famished Road is more in tune with *lo real maravilloso*, or marvelous realism. In Sasser's reading of *The Famished Road* she makes a strong case for the relationship between Okri and Alejo

---

[31] Warnes, 124.

[32] Kapur.

[33] Sethi.

[34] Warnes, 124.

Carpentier, the creator of marvelous realism or *lo real maravilloso*. *Lo real maravilloso* and its relationship to magical realism is a contested one—many argue the difference between the two is an ontological/phenomenological one. While magical realism is more concerned with the experiencing of "reality", Carpentier argued "it was natural to Latin America's history, geography, people, and politics that unlikely combinations of events occurred producing marvelous results."[35] The ending of *El reino de este mundo* encourages an acceptance of the supernatural possibilities of Mackandal's execution/escape. Carpentier juxtaposes the rationality of the French soldiers with the supernatural existence of the slaves to demonstrate the marvelous reality of Cuba, and thus, of Latin America. In *The Famished Road*, Okri creates a setting where both humans and otherworldly spirit children inhabit. Okri encourages the reader to understand the reality of his culture through the understanding of the spirit realm. By doing so, Okri joins a legacy of writers concerned with ontological truths about spiritual phenomena.

In comparing the cultural perspective of both *The Famished Road* and *El reino de este mundo*, a shared use of *lo real maravilloso* becomes apparent. Though Carpentier vehemently claims *lo real maravilloso* can only pertain to a Latin American reality and Okri simultaneously denies any relationship with any type of supernatural realism, the similarities between their manifestos are difficult to ignore. Haiti, Carpentier's muse, can be compared to West African traditions, especially considering the presence of the supernatural and voodoo. The concept of "animism" is present in both cultures, and the desire to approach the supernatural as a collective "real" is a concern of both Okri and Carpentier.

Without question, *The Famished Road* enacts separate codes of magic and realism throughout its narrative. The question then becomes, to what extent do these codes intermingle, and how can a scholar consider this intermingling in the context of magical realism. Considering the cultural and historical conditions present within Okri's childhood and narrative, one can definitely understand

---

[35] Reeds, 182.

how this novel might fit within the parameters of magical realism. Okri writes from a Nigerian cultural perspective, which fits neatly into a later form of postindependence and postcolonial and African fiction. One can surmise from the narrative that Okri is concerned with what constitutes the "true Nigerian nation".[36] This postrealist approach to ardent postindependence nationalism examines the conditions of newly built nations, a similar approach to that of canonical Latin American magical realist authors such as Carpentier.

Through a characterization of African traditions in Latin America, Carpentier acts as a beacon for proto-global West African literature. He provides a framework under which the West African tradition exists globally. This connection is not only a racial one, but rather a cultural one as well since African slaves kept traditions as they were transported from Africa to the Caribbean. The relationship between these frameworks is important—it allows us to think of *The Famished Road* as a novel that diverges slightly from the more traditional and phenomenological approaches to magical realism. A potential representation of this displacement can be demonstrated through quantitative methods. A distance matrix that includes *The Famished Road* and *El reino de este mundo* (Figure 7) demonstrates a similarity in terms of relative word frequency. *El reino de este mundo* is much shorter in length than *The Famished Road*—the former is 180 pages and the latter is 512—which explains the disparity between the number of points in the visualization. Still, the measures described earlier should prevent length from being a variable in the analysis. For the purposes of analyzing novels of such different lengths, each chunk consisted of 500 words.

When returning to the texts, the findings were not substantial, though it's not surprising that text mining has difficulty finding patterns when only using two novels. The relationship between the chunks that appeared close to each other in the visualization was not immediately apparent. The

---

[36] Sasser, 73.

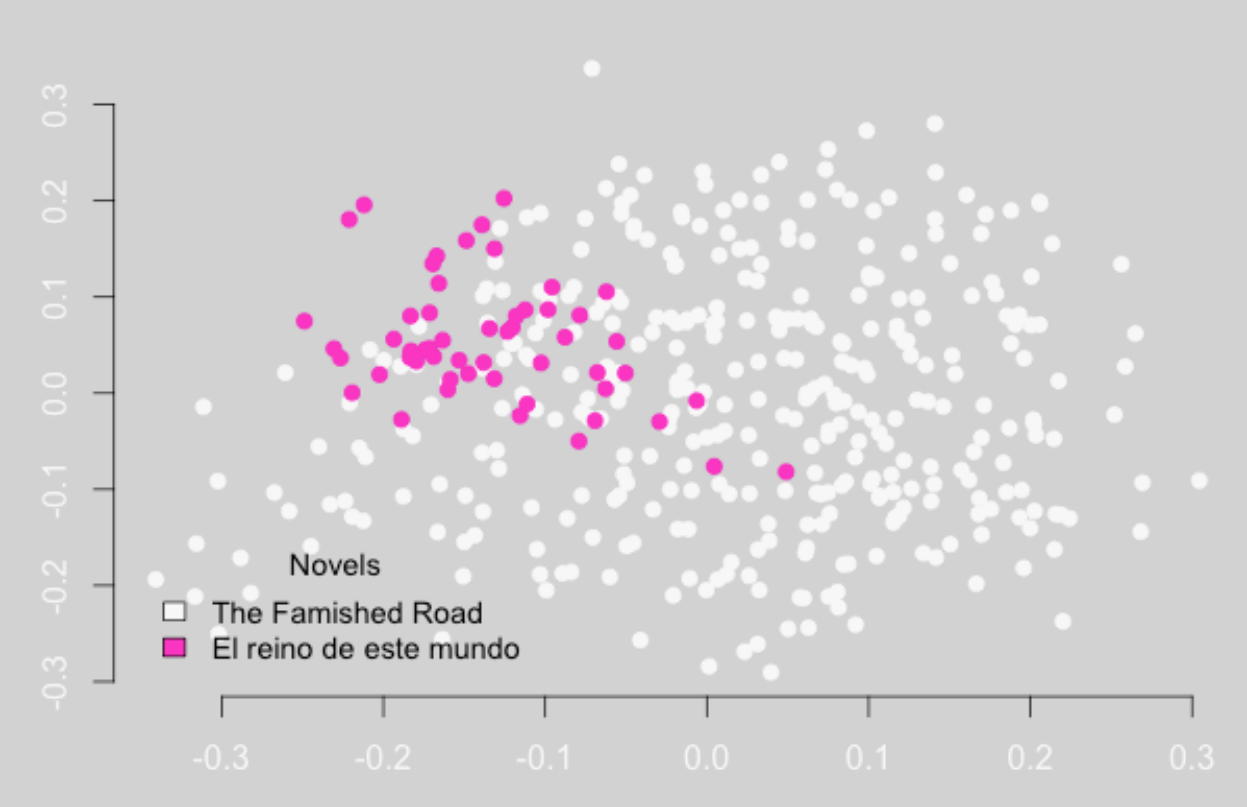


Figure 7: Visualization of a distance matrix between *The Famished Road* and *El reino de este mundo*

short length of each chunk also prevents the chunks as a whole from containing a cohesive segment of writing, which means the lexical variety will be much smaller and therefore more difficult to generate a coherent analysis. The similarities presented in the visualization are indications of deeper concepts, but would require further investigation to find potential correlations with a larger corpus. Along with the analysis of magical realist corpora, this study points toward potential pathways a scholar can follow in order to discover differences and similarities between magical realism and *lo real maravilloso* at the lexical and quantifiable level.

Possibilities for Further Study

Further investigation should be performed using the information obtained from this study. Primarily, the use of machine-learning strategies and the creation of a magical realist statistical model would provide a more robust analysis of the mode. With the present information, I can make

inferences about the mode based on lexical variety. For example, the importance of death and violence as a theme and a surging interest in the urban as a vehicle for magical realist fiction denote important characteristics discovered through unsupervised computational methods. However, I cannot make inferences about the stylistic characteristics of the mode, which are truly at the core of magical realism. A statistical model would consist of choosing magical realist passages—already the study becomes much more specific—and training an algorithm to decide whether or not a piece of literature is magical realist. This type of experiment would give us a much clearer idea as to how magical realism operates at the computational level. Given the findings of this study, creating a model that features death and violence as a primary target for understanding magical realism would be incredibly beneficial. Furthermore, an exploration of a potential shift in the traditional magical realist setting toward an urban setting would be pertinent to capture differences in the conditions under which magical realism flourishes. These analyses would no longer be unsupervised because there is a clear set of lexical features the algorithm would be in search of.

Another noteworthy research question that arose from this study is the possibility of bias after the publication of *Cien años de soledad* in 1967. In the period before the Latin American *Boom*, magical realism was *very* undefined, still in its nascent stages. Since the publication of *Cien años de soledad*, magical realism has become synonymous with García Márquez's novel. Since then, scholars' idea of what a magical realist novel is might be tainted by their experience of having read *Cien años de soledad*. Perhaps these novels are not canonical because they represent the truest form of magical realism, but rather their place in canonical literature rests on their similarity to García Márquez's magnum opus. The data supports this hypothesis—Figures 1 and 2 demonstrate that *Cien años de soledad* clusters closer to text by Salman Rushdie and Isabel Allende, both authors who have mentioned García Márquez as a strong influence. This issue can be studied through translations as well. The translation of *Kingdom of This World* used here in comparison to *The Famished Road* was published in 1989, so the similarity between these (and the other novels that were compared in this

study) could be due to translational and authorial bias. I firmly believe this concern can be addressed through the use of computational tools as well, specifically an in-depth lexical analysis of texts and the translations of those texts pre- and post-García Márquez.

Still, by using only unsupervised methods, I have managed to demonstrate the possibilities for a continuous, globalized magical realism. Analyzing magical realism separately from cultural context allows for the possibility to recognize certain characteristics that exist throughout the magical realist corpus, from Latin America to the rest of the world. Since this is such a large undertaking, using text analysis and computational methods should be considered an appropriate tool for the discovery of patterns in magical realism. These patterns are paramount not only for our understanding of magical realism, but, due of the mode's power of resistance, these patterns can also offer a way to understand how authors criticize and comment on the conditions under which they live, locally and globally. And here I invoke the Old Norse ancestor of haunt: *heimta,* to lead home. The ghost of magical realism haunts world literature not through fear or intimidation, but through a comforting, lingering presence. For world literature, magical realism is a home.

Bibliography

Allison, Sarah et al. "Quantitative Formalism: An Experiment." *n+1* 13 (2012): n. pag. Print.

Argamon, Shlomo et al. "Gender, Race, and Nationality in Black Drama, 1950-2006: Mining Differences in Language Use in Authors and Their Characters." *Digital Humanities Quarterly* 3.2 (2009): n. pag. Print.

Bowers, Maggie Ann. *Magic(al) Realism*. London ; New York: Routledge, 2004. Print. The New Critical Idiom.

Chanady, Amaryll. "The Territorialization of the Imaginary in Latin America: Self-Affirmation and Resistance to Metropolitan Paradigms." *Magical Realism: Theory, History, Community*. Ed. Lois Parkinson Zamora and Wendy B. Faris. Duke UP, 1995. 125–144. Web.

Cohn, Deborah. *The Latin American Literary Boom and U.S. Nationalism during the Cold War*. Nashville: Vanderbilt University Press, 2012. Print.

Faris, Wendy B. *Ordinary Enchantments: Magical Realism and the Remystification of Narrative*. 1st ed. Nashville: Vanderbilt University Press, 2004. Print.

Feinerer, Ingo, Kurt Hornik, and David Meyer. "Text Mining Infrastructure in R." *Journal of Statistical Software* 25.5 (2008): 1–54. Print.

Flores, Angel. "Magical Realism in Spanish American Fiction." *Hispania* 38.2 (1955): 187. *CrossRef*. Web.

Foster, John Burt, Jr. "Magical Realism, Compensatory Vision, and Felt History: Classical Realism Transformed in The White Hotel." *Magical Realism: Theory, History, Community*. Ed. Lois Parkinson Zamora and Wendy B. Faris. Duke UP, 1995. 267–283. Web.

Grzęda, Paulina. "Magical Realism: A Narrative of Celebration or Disillusionment? South African Literature in the Transition Period." *ariel: A Review of International English Literature* 44.1 (2013): 153–183. *CrossRef*. Web.

Hope, J., and M. Witmore. "Shakespeare by the Numbers: On the Linguistic Texture of the Late Plays." *Early Modern Tragicomedy*. Ed. Subha Mukherji and Raphael Lyne. N.p., 2007. 133–153. Web.

Hun, Kim Sang, and Kornelije Kvas. "Magical Realism in the Comparative Context." *Philological Studies/Filološki Studii/Filoloske Pripombe/Filologicheskie Zametki/Filoloske Studije* 1 (2011): 179–186. Print.

Jockers, Matthew L. *Macroanalysis: Digital Methods and Literary History*. U of Illinois P, 2013. Web. Topics in the Digital Humanities~~Topics in the Digital Humanities.

---. *Text Analysis with R for Students of Literature*. Springer, 2014. Web.

Kapur, Vikram. "'Reality Is Not in the Realism.'" *The Hindu* 28 Jan. 2012. Web.

Long, Hoyt, and Richard Jean So. "Literary Pattern Recognition: Modernism between Close Reading and Machine Learning." *Critical Inquiry* 42.Winter (2016): 1–33. Print.

Merivale, Patricia. "Saleem Fathered by Oskar: Intertextual Strategies in Midnight's Children and The Tin Drum." *Reading Rushdie: Perspectives on the Fiction of Salman Rushdie*. Ed. M. D. Fletcher. Rodopi, 1995. 83–96. Web. Cross/Cultures: Readings in the Post/Colonial Literatures in English~~CrossC~~16.

Moretti, Franco. *Distant Reading*. London ; New York: Verso, 2013. Print.

---. *Graphs, Maps, Trees: Abstract Models for Literary History*. Paperback ed. London: Verso, 2007. Print.

Napier, Susan J. "The Magic of Identity: Magic Realism in Modern Japanese Fiction." *Magical Realism: Theory, History, Community*. Ed. Lois Parkinson Zamora and Wendy B. Faris. Duke UP, 1995. 451–475. Web.

Quayson, Ato. "Fecundities of the Unexpected: Magical Realism, Narrative, and History." *The Novel: Volume 1: History, Geography, and Culture*. Ed. Franco Moretti. Princeton, NJ: N.p., 2006. 726–758. Print.

Ramsay, Stephen. "In Praise of Pattern." *TEXT Technology* 14.2 (2005): 177–190. Print.

R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2015. Web.

Reeds, Kenneth. "Magical Realism: A Problem of Definition." *Neophilologus* (2006): n. pag. Web. Literature Online.

Sasser, Kim. *Magical Realism and Cosmopolitanism: Strategized Belonging*. New York: Palgrave Macmillan, 2014. Print.

Sculley, D., and Bradley M. Pasanek. "Meaning and Mining: The Impact of Implicit Assumptions in Data Mining for the Humanities." *Literary & Linguistic Computing* 23.4 (2008): 409–424. Print.

Sebastiani, Fabrizio. "Machine Learning in Automated Text Categorization." *ACM Computing Surveys* 34.1 (2002): 1. Print.

Sethi, Anita. "Ben Okri: Novelist as Dream Weaver." *The National* 1 Sept. 2011. Web.

Siskind, Mariano. "The Global Life of Genres and the Material Travels of Magical Realism." *Cosmopolitan Desires: Global Modernity and World Literature in Latin America*. Evanston, Illinois: Northwestern University Press, 2014. 59–100. Print.

Strecher, Matthew C. "Magical Realism and the Search for Identity in the Fiction of Murakami Haruki." *Journal of Japanese Studies* 25.2 (1999): 263. *CrossRef*. Web.

Tekdemir, Hande. "Magical Realism in the Peripheries of the Metropolis: A Comparative Approach to Tropic of Orange and Berji Kristin: Tales from the Garbage Hills." *The Comparatist* 35.1 (2011): 40–54. *CrossRef*. Web.

Upstone, Sara. "Magical Realism and Postcolonial Studies: Twenty-First Century Perspectives." *Journal of Commonwealth and Postcolonial Studies* 17.1 (2011): 153–163. Print.

Walker, Steven F. "Magical Archetypes: Midlife Miracles in The Satanic Verses." *Magical Realism: Theory, History, Community*. Ed. Lois Parkinson Zamora and Wendy B. Faris. Duke UP, 1995. 347–369. Web.

Warnes, Christopher. *Magical Realism and the Postcolonial Novel: Between Faith and Irreverence*. Basingstoke ; New York: Palgrave Macmillan, 2009. Print.

Wickham, Hadley. *Stringr: Simple, Consistent Wrappers for Common String Operations*. N.p., 2015. Web.

Yu, Bei. "An Evaluation of Text Classification Methods for Literary Study." *Literary & Linguistic Computing* 23.3 (2008): 327–343. Print.

Zamora, Lois Parkinson, and Wendy B. Faris, eds. *Magical Realism: Theory, History, Community*. Durham, N.C: Duke University Press, 1995. Print.

Appendix

A. Details on Corpora

Spanish language corpus:

El Aleph y cuentos (1949) by Jorge Luis Borges
Cien años de soledad (1967) by Gabriel García Márquez
Como agua para chocolate (1989) by Laura Esquivel
El reino de este mundo (1949) by Alejo Carpentier
Eva Luna (1987) by Isabel Allende
La casa de los espíritus (1982) by Isabel Allende
Hombres de maíz (1949) by Miguel Ángel Asturias
Los recuerdos del porvenir (1963) by Elena Garro
Pedro Páramo (1955) by Juan Rulfo.

Anglophone corpus:

Beloved (1987) by Toni Morrison
Imaginings of Sand (1996) by André Brink
Kafka on the Shore (2002) by Haruki Murakami
Midnight's Children (1981) by Salman Rushdie
The Satanic Verses (1988) by Salman Rushdie
The Famished Road (1991) by Ben Okri
The White Hotel (1981) by D. M. Thomas
Tropic of Orange (1997) by Karen Tei Yamashita
Ways of Dying (1995) by Zakes Mda.

B. Topic Modeling Results

Anglophone Corpus Topics:

Topic 1	london, english, sir, beneath, movie, fucking, bloody, film, hell, body, human, indian, everest, point, really, devil, damn, fell, telephone, baba, right, evil, self, actor, class, television, bastard, hit, british, movies, spite, soul, creature, fellow, beings, sort, person, shouted, okay, show, flight, pretty, type, mountain, century, high, horns, goddamn, wearing, plane
Topic 2	people, came, bar, spirit, saw, men, looked, eyes, room, women, road, round, began, money, sat, face, wind, door, dad's, darkness, street, heard, stayed, wine, compound, world, air, food, voices, spirits, fell, children, strange, gave, thugs, silent, stared, party, table, kept, chair, drank, palm, suddenly, beggars, stopped, going, candle, music, blind
Topic 3	people, children, village, city, woman, funeral, shack, going, school, mountain, women, mother, father, death, men, child, settlement, day, food, work, funerals, nurse, town, money, want, taxi, figurines, brother, police, came, church, bus, professional, shop, says, beautiful, walk, friend, clothes, family, community, freedom, tribal, sit, wife, sing, government, chief, business, pictures
Topic 4	city, god, wife, pilgrims, butterflies, verses, men, archangel, water, prophet, sea, village, stone, villagers, grandee, pilgrimage, desert, feet, curtain, messenger, lat, idea, tent, sees, poet, offer, sand, fellow, adversary, allah, boom, persian, kind, vision, mecca, damn, falls, bullock, closed, says, feels, asks, poetry, goddess, whores, business, heart, devil, weak, human
Topic 5	says, family, farm, small, men, years, birds, course, asks, mother, coffin, woman, quite, town, girl, cape, nurse, house, brought, kitchen, sandile, stories, outside, dust, bird, cattle, women, sheep, turns, came, really, happened, trek, god, shadow, deep, suppose, police, realise, river, difficult, round, child, mirror, mere, takes, palace, jethro, africa, sound
Topic 6	time, life, mother, year, house, old, woman, day, came, father, child, people, knew, death, take, family, little, right, took, mind, end, men, away, really, fact, girl, story, wife, brought, find, matter, believe, course, business, talk, work, look, quite, going, dead, things, lost, possible, call, remember, living, ago
Topic 7	eyes, head, away, face, time, old, night, place, room, hand, black, look, white, day, light, water, hair, door, little, felt, going, saw, bed, open, came, dark, voice, blood, small, hear, turned, mouth, window, want, hard, sleep, took, side, body, seen, front, boy, started, past, arms, dead
Topic 8	mother, father, nose, monkey, doctor, grandfather, reverend, green, days, beneath, little, estate, old, whatsitsname, inside, finger, mother's, uncle, midnight, brass, lips, washing, children, family, dr, despite, snake, big, film, baba, burns, skin, birth, servants, god, sheet, commander, saffron, daughter, hand, business, dust, filled, pickle, mumtaz, conference, villa, road, year, grandfather's

Topic 9	car, city, crowd, street, men, driver, road, police, drive, public, high, cars, ride, party, windows, yelled, glass, shop, scene, speed, political, direction, number, arrived, type, hot, announced, mass, vehicle, green, newspapers, chief, drove, famous, streets, island, bags, myth, view, west, names, traffic, driving, orders, local, pointed, bridge, plastic, national, shit
Topic 10	buddha, picture, war, midnight, general, uncle, magicians, army, sister, witch, indian, captain, city, ghetto, th, singer, nose, mosque, sir, knees, aunt, soldiers, despite, jungle, spittoon, air, basket, ho, prime, ears, optimism, family, south, smell, past, streets, fellow, tiger, mere, minister, military, allah, slum, silver, hut, field, pakistani, snake, president, hero
Topic 11	knew, house, right, sweet, looked, took, mother, children, girl, men, little, shoes, women, looking, place, yard, want, feet, work, milk, years, saw, woman, boys, dress, water, dead, hands, hear, wanted, suggs', hurt, em, believed, door, face, needed, porch, holding, walked, bit, heart, beat, somebody, stove, daughter, road, sat, stay, things
Topic 12	life, began, world, years, end, death, child, god, old, women, dream, children, father, wife, born, history, remained, returned, husband, power, days, secret, true, second, possible, given, state, poor, fact, known, return, begun, refused, certain, mother, voice, learned, country, present, longer, dreams, continued, human, presence, name, simply, birth, appeared, unable, parents
Topic 13	orange, freeway, looked, homeless, people, big, brother, work, oranges, north, mexico, phone, car, palm, kid, hey, trees, south, border, yeah, cars, tv, stuff, bus, entire, traffic, gabe, gran, body, side, gets, kids, gabriel's, talk, damn, pulled, mexican, overpass, crazy, plan, check, knew, hour, corn, today, net, chinese, newsnow, live, probably
Topic 14	right, says, stone, head, library, lot, inside, cat, kind, cats, time, world, pretty, going, things, take, really, nod, okay, open, looks, feel, desk, want, room, yeah, entrance, music, sky, sit, read, eyes, need, coffee, crow, tokyo, sure, able, front, different, anybody, anymore, shirt, reading, bag, hard, idea, somebody, tall, asleep
Topic 15	mother, hotel, aunt, woman, lake, felt, train, quite, breast, husband, white, dress, mother's, happy, uncle, patient, father, letter, able, couple, dear, case, music, friends, course, opera, mountains, lovers, maid, jewish, friend, kind, stars, guests, holiday, priest, heard, marriage, short, hardly, mountain, kiev, probably, singing, reading, pains, journal, vogel, feeling, warm

Latin American Corpus Topics:

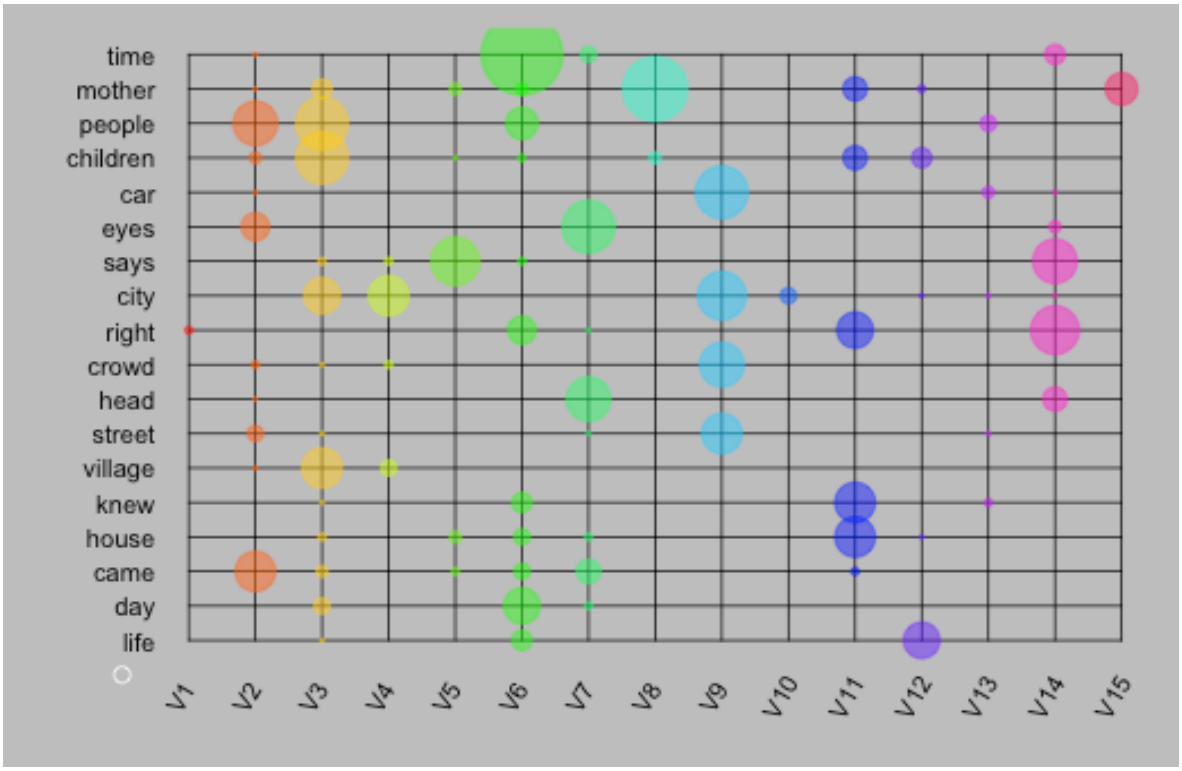
Topic 1	habitación, tío, largo, profesor, jardín, patas, joven, comenzaba, patrón, expresión, seres, periódico, cuentos, pelo, techo, negocio, inglés, uñas, ataúd, desierto, cortinas, muebles, usaba, mantenía, mansión, habitaciones, nieta, mente, esquina, estilo, mesas, colocó, simpatía, escalera, vivido, barriga, menudo, capacidad, hospital, extranjero, salón, invitados, parque, salones, especialmente, empleados, sirvió, hogar, delantal, columnas

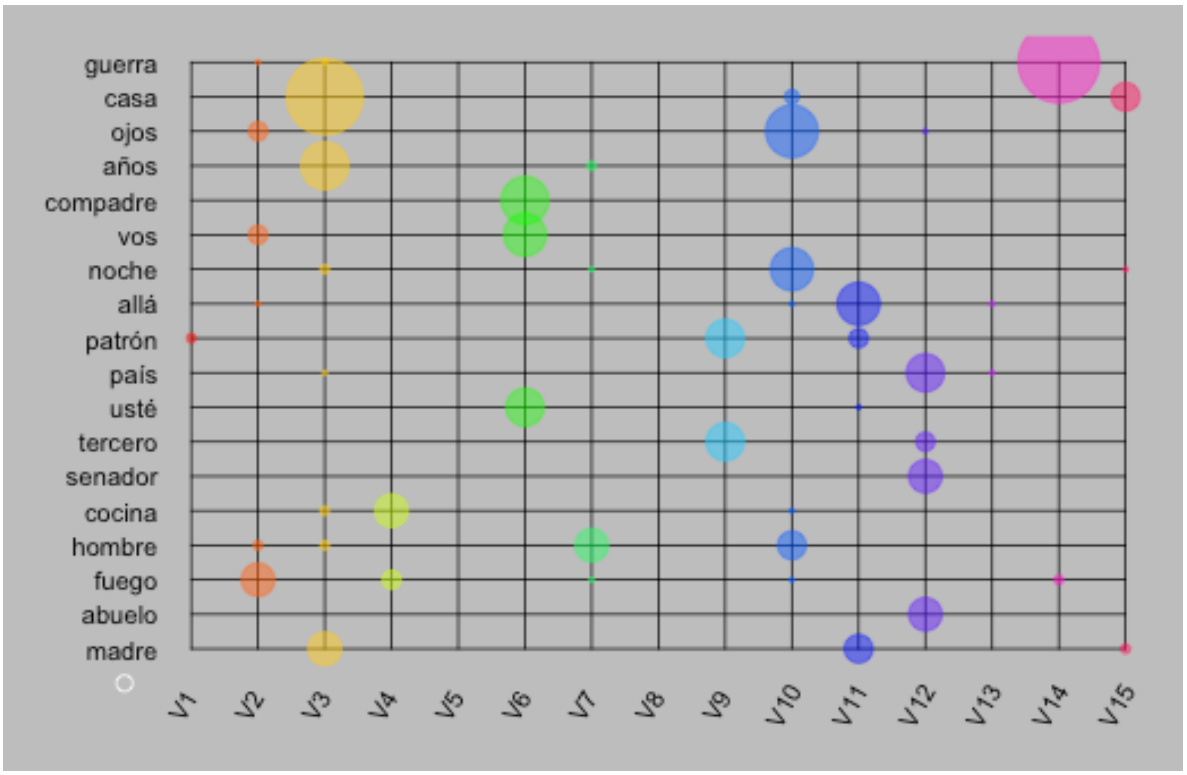
Topic 2	fuego, agua, venado, curandero, vos, hijos, ojos, ilóm, señor, maíz, macho, dientes, orejas, muerto, pelo, cajón, sombrero, color, luciérnagas, camino, gu, brujos, animal, ciego, piedra, tembladero, hombre, bestia, gaudencio, musús, rancho, vaca, rozas, iba, oro, nubes, carne, árboles, conejos, cabeza, veneno, ojo, subteniente, herbolario, sombra, mano, montada, bestias, ruin, río
Topic 3	casa, años, madre, tiempo, día, podía, días, familia, amor, hacía, único, cama, hija, tarde, mundo, parecía, niños, mesa, decía, hijos, vio, ropa, horas, idea, meses, hombre, niña, única, tiempos, punto, forma, hijo, trabajo, cuarto, modo, marido, pronto, empezó, piel, cambio, llevó, cocina, realidad, puso, hermano, nombre, tantos, hora, pasaba, semana
Topic 4	cocina, agua, rancho, fuego, habla, hermana, doctor, boda, patio, sentía, palabras, día, chiles, mesa, masa, preparar, lágrimas, quería, cuerpo, mirada, frijoles, poder, pone, azúcar, ingredientes, tenia, pudiera, punto, cebolla, sabía, niño, podía, comida, carne, tomó, almíbar, nuevamente, rápidamente, ponen, amor, caldo, manos, mente, escuchó, invitados, claro, leche, huevos, huevo, tía
Topic 5	volvió, dormitorio, cuarto, aquel, aquella, pueblo, encontró, taller, oro, meses, modo, corazón, mundo, lluvia, compañía, vio, llevaba, gitanos, tren, tiempo, bananera, tarde, corredor, regreso, realidad, ciénaga, pergaminos, calle, ansiedad, tratando, aldea, cabo, hora, llevaron, sala, cierta, siquiera, triste, conciencia, cartas, hielo, tienda, patio, voluntad, oyó, buscando, antigua, contestó, seguía, ocasión
Topic 6	compadre, vos, usté, correo, señor, trago, revolorio, coyote, tatacuatzín, garrafón, cumbre, pesos, muía, aquino, aguardiente, acatan, indio, mingo, miguelita, tecuna, mal, ái, casualidón, arrieros, diablo, arriero, sacayón, olegario, amor, malo, cuevas, puro, ramos, presos, correspondencia, negocio, gallo, onde, sos, guaro, peor, chai, pelo, deféric, contestó, decía, cartas, sombrero, español, palabra
Topic 7	hombre, hombres, dios, historia, palabras, laberinto, quizá, ciudad, nombre, muerte, otálora, mundo, acto, universo, sueño, acaso, zahir, bandeira, infinito, muerto, destino, años, libro, siglo, sé, poema, obra, arena, temor, pensó, argentino, averroes, recordó, hechos, aquel, murió, principio, espejo, espinosa, noches, cuyo, símbolo, judas, abenjacán, vano, palabra, día, pensé, moneda, tesoro
Topic 8	señora, madrina, pude, calle, comandante, sentí, allí, kamal, aravena, patrona, agua, melecio, maestra, oriente, selva, resultaba, guerrilleros, puedo, comenzó, conmigo, pajarito, guardias, traté, supe, almacén, cuentos, colonia, jefe, penal, seguro, radio, dije, encontré, lograba, aldea, historias, rodillas, modos, fábrica, di, quise, turco, horas, pensé, ocurre, verme, oficio, llegué, difícil, quedé
Topic 9	patrón, tercero, campo, valle, campesinos, comenzó, hermana, cuevas, trucha, inquilinos, doctor, capital, propiedad, barrabás, tren, gallinas, sirvientes, coche, rabia, conde, estación, viejo, té, colegio, cuadernos, hermanos, espíritus, administrador, cuero, terremoto, cuñada, perro, fuerte, esquina, corredores, fundo, fiebre, sistema, patrones, excepto, podía, vacas, ojo, bestias, río, hermanas, respondió, fierro, verano, equipaje

Topic 10	ojos, noche, mujer, hombre, manos, hombres, cabeza, puerta, mano, día, tiempo, cuerpo, agua, voz, iba, gente, allí, aire, pueblo, mujeres, boca, dios, frente, muerte, mañana, medio, casa, sangre, pasar, ir, aquel, volvió, paso, silencio, suelo, salió, camino, pie, saber, llevaba, mal, salir, negro, lejos, vio, pies, muerto, aquella, sueño, cielo
Topic 11	allá, dije, sé, madre, patrón, conmigo, cielo, cosa, sabe, murió, lluvia, oí, dónde, aquello, modo, señor, viene, media, comala, quizá, venido, cama, voy, di, quieres, vine, digo, dices, pensé, verdad, supe, pregunté, arriba, alguien, oyendo, puedo, creí, decía, vive, iré, sabes, sentí, oía, quiera, gusta, ves, cuánto, dile, oye, falta
Topic 12	país, senador, abuelo, ciudad, policía, tercero, nieta, edificio, presidente, gobierno, armas, tío, abuela, militares, políticos, ideas, estudiantes, comenzó, poder, respondió, podían, carabineros, pareció, política, montaña, militar, dictadura, prisioneros, revolución, compañero, guardia, universidad, teléfono, radio, replicó, puertas, calles, derecha, compañeros, media, acontecimientos, siguientes, automóvil, palacio, libertad, congreso, democracia, violencia, sitio, celda
Topic 13	negros, cabo, esclavos, llanura, negro, hacienda, amo, montaña, ogún, ciudad, palacio, tambores, mar, cuyo, ciertas, demasiado, isla, cuyas, manco, aquel, aquella, verde, allá, solía, gobernador, poderes, ciertos, pronto, parís, negras, francia, comenzó, blancos, souci, lluvias, orden, alto, veneno, fortaleza, mandinga, sans, ciudadela, norte, arriba, aquellos, anciano, caminos, tantas, honor, causa
Topic 14	guerra, soldados, gobierno, calle, militar, oficial, orden, armas, capitán, autoridad, pelotón, oficiales, muertos, liberales, muerte, ejército, hombres, fusilamiento, previsto, madrugada, trabajadores, jochen, noticia, creyó, hamaca, cuartel, presidente, moneada, amanecer, jefe, término, fuego, régimen, medianoche, mando, veinte, fuerzas, estación, liberal, civil, militares, muchedumbre, noticias, armados, república, público, escuela, niño, conservadores, cuyos
Topic 15	palabras, señora, joven, miró, casa, cuarto, jardín, hotel, soldados, cama, calle, luchi, señor, contestó, volvió, justo, militares, memoria, voz, habitación, calles, jóvenes, preguntó, fiesta, silencio, orden, días, militar, plaza, mujeres, mañana, capitán, miraron, respondió, paso, quería, salió, flores, presidente, acercó, criados, frente, corredor, repitió, pepe, amigos, miedo, miraba, extranjero, reír

C. Top Words

Here are two visualizations that demonstrate the prevalence of each top word in each topic.




guerra
casa
ojos
años
compadre
vos
noche
allá
patrón
país
usté
tercero
senador
cocina
hombre
fuego
abuelo
madre
V1
V2
V3
V4
V5
V6
V7
V8
V9
V10
V11
V12
V13
V14
V15