

A New Literary Analysis of *The Great Gatsby*
by Natural Language Processing

A Thesis
Submitted to
the Department of English
Doshisha University

In Partial Fulfillment
of the Requirements for the Degree
of
Bachelor of Arts

Yuhao Dai
(No. 1102170010)
(Adviser: Professor Noriko Ishizuka)

December 2020

CONTENTS

Abstract	i
1. Introduction.....	1
2. NLP Techniques.....	2
3. Representative Meanings of Green.....	4
4. American Dream.....	8
5. Inequality of American Dream.....	10
6. Conclusion.....	11
Notes.....	13
Works Cited.....	14

Abstract

A New Literary Analysis of *The Great Gatsby* by Natural Language Processing

Applications of computer science are changing the ways people obtain information and the ways of communication. It also greatly improves productivity on multiple fronts. Natural Language Processing (NLP) application in literature, however, is still limited. This thesis introduces the methods that encode words into numbers for machines to process, and furthermore to use the pre-trained two-layered neural network, Word2Vec, to explore one of the greatest novels in American literature, *The Great Gatsby*, to examine the representative meanings of the color green and the connections between the protagonist Jay Gatsby and the American Dream.

The first chapter introduces the previous papers on both NLP and literary studies that helps to analyze the text in this thesis. NLP papers helps in developing the Python program, which would be trained solely on the text of *The Great Gatsby* and generates the semantically closest words of a given word. The results would be the bedrock of this thesis's literary analysis about both green light and American Dream. Previous studies helped in gaining general interpretations of *The Great Gatsby*, and to cross compare the differences between these interpretations with the results from the machine learning model, therefore, to show how differently a machine learning model would see the text. This chapter also explains two different techniques to encode text into numbers, and a pre-trained two-layered neural network called Word2Vec.

The second chapter introduces some results of the machine learning model I trained based on the text of *The Great Gatsby*. These results help me to understand the text better and give me a new perspective on the text. Then, this

chapter further explores the most relevant outputs of the word green, which are “light,” “shore,” “unknown,” and “orgastic”. And all these words are then put into the machine learning model, showing the results to be “dream” and “orgastic,” which lead to the connection of color green and American Dream.

The third chapter introduces representative meanings of the color green, using the results from the machine learning model, which includes “light,” “shore,” “unknown,” and “orgastic”. Each keyword is carefully analyzed, leading the conclusion that green is the concept of hope, and Gatsby’s “extraordinary gift for hope” embodies the idea of American Dream itself, which leads to the next chapter.

In American Dream chapter, this thesis introduces the original idea of American Dream, how its ideology embodies equality while in reality hides a sense of inequality. Out of this hidden attribute which Gatsby fails to understand, he fails. In order to better understand the reasons for his failure, this thesis compares him with another similar character, Dexter Green, from Fitzgerald’s short story “Winter Dreams.”

The last chapter explains the inequality hidden inside the ideology of American dream, and the message Fitzgerald conveys through *The Great Gatsby*: The American Dream is something one can thrive for, but keep in mind that even after everything achieved, one can never be truly accepted by the elite group.

(487 words)

A New Literary Analysis of *The Great Gatsby*

By Natural Language Processing

Introduction

The advancement in computing power led us into a new world, where cars can drive themselves, packages are delivered by drones and artificial intelligence can help diagnosis cancers. Recent developments in natural language processing are also changing the ways people think about the capacity of computers. For instance, a none profit organization called Open AI recently developed a language processing deep learning model called GPT-3, which contains 175 billion machine learning parameters. The model is able to produce short texts and small applications which would enhance productivity profoundly. However, its application on literary textual analysis is still limited. This thesis introduces the machine learning methods and shows results from analyzing the text of *The Great Gatsby*, and how a perspective by machine can differ from the ones by human readers.

For the lack of similar papers that deal with both NLP and literature, I divided the field of research into two categories: NLP and Literary Textual Analysis.

NLP

For NLP, I read Google and Open AI's papers (Devlin et al.), to confirm such application's authority, and papers relating to semantic relation extraction (Sergeev) to see the current application. To free myself from training a brand-new neural network, which could be both expensive and time consuming, I used a highly recognized pre-trained 2-layered neural network called Word2Vec and read both papers by its creator, Thomas Mikolov's team (Mikolov et al.). All the codes used to present the results are written by myself.

Literary Textual Analysis

For literature analysis, I read papers about color symbolism in *The Great Gatsby*, which analyses representative meanings of the color green like life, youth, growth, hope, safety, spring and happiness (Yamanaka 220). Another paper indicates the color green, especially the green light, is considered to represent the color of banknotes in America, and furthermore, the American Dream, which means American ideal of freedom and equal opportunity for success, is just the money, which is interchangeable to almost all physical materials, like food and luxurious lifestyles (Miyamoto 166).

With a regard to these opinions about color green and American Dream, let's turn our focus to the results from the machine learning model. But first, we need to see the techniques used to encode word into numbers.

NLP Techniques

To use machine learning, the text must be tokenized before feeding into a 2-layered of pre-trained neural network model called Word2Vec invented at Google by Tomas Mikolov. This neural network model transforms words into vectors of numbers, which made it possible for computer to train and understand the meaning of words.

Tokenization Techniques

Generally, two algorithms are used in tokenization: Bag of Words and Skip-Gram. To understand tokenization, consider the following example : “To be or not to be.” First method is called Bag of Words and is the foundation for literary textual analysis in the past. This method would require a larger corpus, the main operation is to put all the words inside a bag, shake them up (put all the words back into its original forms, such as getting rid of tense and spelling errors), and then pour the words out. The outcome has lost its original in-word connections, so all we can do is count how many times each word has appeared, and to denote each and every one of them with their count.

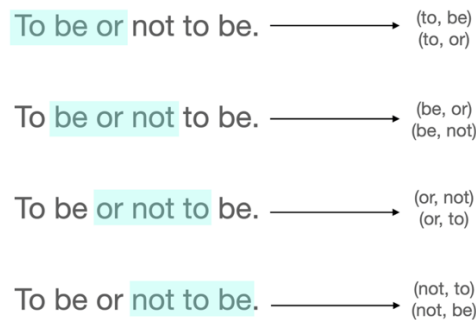
“to” : 2

“be” : 2

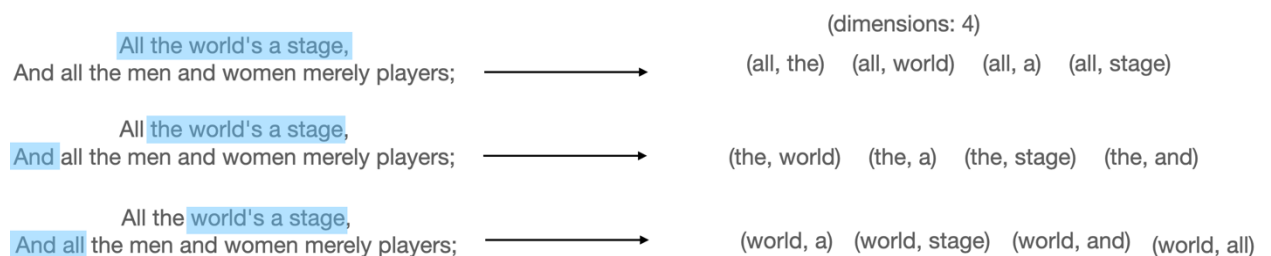
“or” : 1

“not” : 1

However, as in literature analysis, word order is important to consider. Therefore, another method of tokenization that takes word order into account called Skip-Gram is preferred, which would tokenize the quote as shown:



This way of tokenization offers both words' occurrences and their interrelations. Here, one parameter works as a key, the "dimension," which is just the number of adjacent words to consider when tokenizing a word; here in this example, the dimension is 2, since every word is tokenized into two tuples of words. In order to understand this concept, let's look at another, a longer example from *As You Like It*:



In this instance the dimension parameter is tuned to 4, meaning that for each word, 4 words of its adjacent are encoded in numbers. This gives each word more connections to other words and therefore each word has a better understanding of the context it is in. In practice, I encoded each word with 1000 dimensions, meaning for each word; there are 1000 words next to it considered as adjacent words and therefore closer in the vector space. The cosine similarity is, essentially, the scaled distance

between words in vector space. This distance represents how close a connection is, both semantically and contextually, between words, and therefore is helpful in our case of literary analysis.

My Usage of the Model

I trained the model based only on the text of *The Great Gatsby*. The model would encode each word in 1000 dimensions. In order to extract the most similar words to each word, the model calculates the distance between that word and every other word in the text in the 1000-dimension vector space, using cosine-similarity, whose availability and credibility are confirmed in the Mikolov et al. paper. After 1000 iterations, I tested some keywords for its accuracy.

For the word “dream,” the model returns “fail,” “grasp,” and “intriguing” as its most similar words, and for the word “Gatsby,” the model successfully returns that the most similar words are “Gatz,” “empty,” and “shore”. The word “empty” shows up only 4 times in the entire novel, and only once is connected to Gatsby, the book describes him as “man with a hard empty face — the pioneer debauchee....” (Fitzgerald 100), and the word “shore” appears where he meets Dan Cody for the first time and changes his name to Jay Gatsby. With the given evidences for credibility, a further exploration on the word “green” was processed.

Representative Meanings of Green

As mentioned in the “Introduction” section, the color green is previously considered to symbolize “youth,” “vitality,” and “money”. At the same time, the machine learning model tells us the most similar words of “green” are “light,” “shore,” and “unknown”. Also, the word “orgastic” shows up with a similarity of 93% as well. The word “light” is the green light that shines at the end of Daisy’s dock, whose representative meanings are explored in the following section. The word “shore” is explained in the previous section. The word “unknown” shows up twice in the book;

one is at the end of the book, where Gatsby is shot and Nick sits on his blue lawn and says: “as I sat there brooding on the old, *unknown* world...” (Italics mine; 180). The connection between the words “green” and “unknown” is explored further in the “Orgastic Future” section.

Light

In the past of *The Great Gatsby*'s literary analysis, the green light is considered as a dream to repeat the past in the future, as well as a dream to pursue infinite possibilities in the future (Yamanaka 220), and this representation of the green light is successful because “it embodies the profound naiveté of Gatsby's sense of the future, while simultaneously suggesting the historicity of his hope” (Bewley 224).

Despite previously interpretation that green light symbolizes the past, to Gatsby, the green light does not relate to the past, but rather to hope itself. When he first met Daisy in her house, he was a “penniless young man without a past” (149); all he had was the hope for Daisy. The sentence “He took what he could get, ravenously and unscrupulously — eventually he took Daisy one still October night” (149) suggests Gatsby's ownership of Daisy's virginity. Gatsby, a penniless young man, “deliberately [gave] Daisy a sense of security” (149), that he is “from much the same strata as [Daisy]” (149), and he would be able to support Daisy's luxurious lifestyle. Although, since he is not satisfied when he holds Daisy in his colossal mansion, the green light does not simply mean money or Gatsby's past. He wants more, like when he holds Daisy and tells her about the green light, he says: “If it wasn't for the mist we could see ... a green light that burns all night” (92), and becomes “absorbed in what he just said” (92). At this moment, despite the fact that Gatsby is holding the girl he supposedly loves in his mansion, just like what he used to dream about, he is lost in his own imagination. The girl, the mansion, the beautiful shirts do not matter anymore, since at this moment, Gatsby has achieved these objects he once dreamt of, and now is the time to find a new object. This is the

“extraordinary gift for hope” (2) Nick talks about. Green light’s representative meaning is neither past nor money, but rather the concept of hope itself. The problem is that he would always want more. The object of light is a great analogy here, since light and hope are both something that one can *have* in a point of one’s life, but not *preserved*. In order to have them, one must seek for them eternally.

Shore

To Gatsby, shores have some significances. And from different stages of his life, meanings of shores change depending on the stage of his dream.

Firstly, it’s the shore of Lake Superior, the place where James Gatz works “as a clam-digger and a salmon-fisher” (98) and drifts with nothing to do, until he sees Dan Cody’s boat. On the day that “witnessed the beginning of his career” (98), he wears a green jersey. At the moment he borrows a rowboat and pulls out Dan’s boat, he is already Jay Gatsby. This shore represents both the past Gatsby wants to forget and the turning point of his life. This shore is the beginning of his dream, the American Dream where the opportunity of success is provided regardless of the state of wealth, ethnicity and gender.

Then, after Gatsby achieves worldly success through illegally selling alcohols, the shore is the Long Island shore, where he built a colossal mansion with pool and lawn. Gatsby would stand on his blue lawn in a cold winter night to see and try to grasp the green light that shines at the end of Daisy’s dock, which is at the other side of the shore. Here, Gatsby’s shore represents his ambition and “all the beauty and glamour in the world” (100), and the fact that through hard work --- weather it’s legal or not ---, one can turn his life around; this shore in his mansion’s shadow is the destination of American Dream, where one can have a better life if they work hard enough.

Lastly, a shore is a part of the land that Dutch sailors first see, which leads us other keywords: unknown and orgasmic future.

Unknown, and Orgastic Future

In the text of *The Great Gatsby*, the word “unknown” shows up twice: one at the beginning and the other at the end. Let’s focus on the second one, after Gatsby’s funeral, and Nick decided to leave New York and the “careless people” (179) behind, Nick “[sits] [at Gatsby’s blue lawn] brooding on the old, unknown world...” (180). The “unknown world” here refers to the piece of Long Island Nick sees, and it represents New York, and furthermore the new continent. The “unknown world” is the world that once “flowered for the once for Dutch sailor’ eyes” (180), and to explore it is “the last and greatest of all human dreams” (180); it embodies all human dreams. The dream to explore a new world contains same meanings with Gatsby’s dreams, which is to become anyone that is not previously determined by birth or occupation. Dutch sailors share their dreams with Gatsby, for in the new continent they can be whoever they want to be.

Gatsby does not know that he would turn out to be what he is at the end, but he hopes for the best and prepares himself for it, as we can see in “a book [Gatsby] had when he was a boy” (173); he strives to be a better man. Gatsby does so because, despite that he could never be sure of what he can expect in the future, he “was bound to get ahead” (173), because the unknown world is filled with infinite hope.

The end of *The Great Gatsby* reads: “the green light, the orgastic future that year by year recedes before us” (180). Here, the orgastic future is not just for Gatsby, as the narrator Nick pictures the time when the Dutch sailors first appeared on the new continent. When Europeans believed in the possibilities of the greatest geographical discovery, this orgastic future is the “last and greatest of all human dreams” (180). Dutch’s sailors not only literally cleared way for Gatsby’s house by cutting down trees, but also laid the foundation of where New York now lays, and later, the United States of America. To see the green shore is “for the last time in history with something commensurate to his capacity for wonder” (180). The green shore of the new continent is the promise of hopes and opportunities.

To sum up, green is associated with the words “light,” “shore,” “unknown” and “orgastic”. And all of these words symbolize one concept: hope.

American Dream

USA is special in a sense, because unlike other countries, it was consciously built by the founding fathers. The *Declaration of Independence*, and later, the *Constitution of the United States*, express the promises towards its citizens, that the new country built by Americans will be a place where “all men are created equal.” This is a highly ambitious phrase, a high expectation of the new country considering in present day people are still protesting on behalf of racially motivated violence towards African Americans using the slogan “Black Lives Matter.” At a time when women are not considered a member of society, and African Americans are traded as slave against their will, these ideals of equality show the vision and high expectations of the founding fathers. Their ideals and dreams of what USA could be in the future inspired the American Dream.

The American Dream means equal opportunity regardless of one’s birth and financial circumstances. It images a society with upward mobility through hard work, taking risks and self-sacrifice. As mentioned in the “Representative Meanings of Green” section, Gatsby does not try to achieve American Dream himself, but rather to embody the concept of American Dream. Using the machine learning model to extract the most related words with “dream,” as mentioned in the “NLP Techniques” section, returns “fail” and “grasp.”

Considering these two words together, we have an image of Gatsby, trying to grasp the dream of his and fails. He fails not out of incompetence, but because of the illusions he has over his dreams. In order to see the mistakes he makes more clearly, let’s compare him with Dexter Green from “Winter Dreams,” a similar character by Fitzgerald.

Firstly, Gatsby invents himself through illusions. During the course of chasing American Dream, instead of accepting the fact of his birth and making the most out of what he was born as, he invents himself with this persona that holds all the attributes he craves. Instead of admitting that he is just a poor boy from mid-west whose parents are farmers, he despises women that gather around him, lies about himself to everyone with every chance he has. He lies about his name, his birth, his family and his education. Dexter Green, on the other hand, is honest towards people; he never hides his true identity, but rather try to be successful through hard work (of course there's a sense of insecurity Gatsby has with his true identity while Dexter doesn't).

Secondly, Gatsby takes shortcuts when possible. When looking up the most similar word to "dream" in the machine learning model, it shows "fail" with 96% similarity. In *The Great Gatsby*, the author does not explicitly say how Gatsby has got his money, except for the insinuation. The whole book mentions the word "career" exactly twice: when Gatsby reveals his real name: "... at the specific moment that witnessed the beginning of his career" (98), and when Gatsby stops hosting parties on Saturday nights for Daisy: "... as obscurely as it had begun, his career as Trimalchio⁽¹⁾ was over" (113). By comparing Gatsby to Trimalchio, the author cryptically tells us that Gatsby has earned his money through devious means (selling liquor in an underground business). On the other hand, Dexter Green earns his money through entrepreneurship, by taking risks and investing in a new business. And because of his success in an early stage of his life, he is able to confirm with Judy that "[he is] probably making more money than any man [his] age in the Northwest" (5).

Lastly, Gatsby misunderstands his dreams. He sees people with nice things as his dreams. When he sees Dan Cody, he sees TUOLOMEE, the yacht that represents "all the beauty and glamour in the world" (100). Just as when he meets Daisy for the first time, he is astounded since he "[has] never been in such a beautiful house before"

(148). He doesn't love Daisy, but rather the idea of her: "the first *nice* girl he [has] ever known" (Italic mine; 148), and to him, her value is determined by how nice a house she lives in and the fact that "many men had already loved [her]" (149) and the fact that Daisy has multiple suitors "increased her value in his eyes" (149). Gatsby constantly tries to impress people by showing his achievements from the past, like constantly carrying the decoration of Montenegro and a souvenir of Oxford days. This is because of his insecurity that someday, others might find out who he really is. While in the case for Dexter, he doesn't need to bring anything on him to show his identity and solid standing, when he works as a caddy, people know him as "the best caddy in the club" (Fitzgerald, 1); when he buys in the laundry partnership, people talk about him as "*there's* a boy" (2). Everyone already knows Dexter as the best, therefore he doesn't need some things to impress other people, and he could deal with Judy with confidence.

In shorter terms, Dexter Green achieves the American dream through hard work, taking risks and self-devotion, while Gatsby misunderstands the people with nice things for people with value; when he is trying to achieve his goals, he takes shortcuts whenever he could; and he has to invent his own persona in order to participate with others. In the end, Gatsby never believes in American dream, but rather the idea of dream itself; it keeps him moving, so he could continue on "[Gatsby's] Father's business ... a vast, vulgar and meretricious beauty," (98) and therefore he fails at last.

Inequality of American Dream

As discussed in the previous section, American Dream is supposedly about equality, but there is a hidden sense of inequality in it. A Chinese writer Lu Xun once said that one person is born with something that another person pursues for life. This is exactly the case with the inequality in American Dream. This inequality is shown in *The Great Gatsby* as the conflicts between "Old Money" and "Newly Rich," as well

as deeply in the narrator Nick's mind, where "a sense of the fundamental decencies is parceled out unequally at birth" (2), meaning that money is not the only thing that people are born into, some people are just more *decent* than others, and the amount of money they possess cannot change the strata difference they were born into. Gatsby thinks that the real difference between himself and Tom Buchanan is money, but it's actually the social class they were born into. When Gatsby is trying to defense himself from accusations that Tom made, Daisy "[is] drawing further and further into herself" (134). The only reason Daisy agrees to Gatsby was that she thinks "[Gatsby] was a person from much the same stratum as herself" (149), therefore as soon as Daisy finds out about the truths about Gatsby's business, she decides to leave him. People are *born* differently, and no efforts, not something as grand as what Gatsby did, can overcome the gaps between different stratum at birth. In *The Great Gatsby*, the ones who survive at the end of book all belong to "Old Money," including Nick, Tom, Daisy and Jordan; the ones who try to make a living through dedication and hard work are dead: Gatsby, George and Myrtle. This is because of the message the book is trying to convey: that American Dream is something one can try for, but keep in mind that even after everything achieved, one can never be truly accepted by the elite group. This is the true sense of inequality: brought upon at birth and irrevocable.

Conclusion

The advancement in computing power enables this thesis's research through natural language processing, which includes both transforming texts into numbers using a technique called tokenization, and through such technique, a pre-trained neural network called Word2Vec is utilized to fully demonstrate the new perspective carried by computer. After the program is successfully compiled, further exploration on *The Great Gatsby* is conducted. Most similar words to the keywords that greatly affect the book's theme such as the color green and dream are further explored by

machine learning model. For green, it returns “light,” “shore,” “unknown,” and “orgastic,” and for dream, it returns “fail” and “grasp”.

Then, this thesis continues exploring the most similar words of green returned by the machine learning model. For light, while previous researches argue the green light to be the symbol of Gatsby’s past, machine learning model says it’s the concept of hope itself. Shore represents different stages of dreams Gatsby beholds; at first the persuasion of success, and then the taste of success itself; and lastly the unknown and orgastic future that encompasses unlimited possibilities and hopes. The unknown continent refers the colonization of America. A brand-new continent without pre-determined social classes, where everyone has equal opportunities and unlimited hopes. This then leads to the foundation of USA and American dream.

Continuing discussing the American Dream, this thesis compares Gatsby with Dexter Green from “Winter Dreams,” to show how they are similar and different, and how Gatsby’s misunderstanding in American Dream and his own dreams eventually lead to his failure.

Lastly, this thesis deals with the inequality of American dream, the machine learning model returns “certain” and “unalienable” when processes “equal” within the text of *Declaration of Independence*; the so-called equality here ignores a sense of hidden inequality. The last chapter of this thesis deals with the inequality caused by differences at birth, and the true message behind the entire book of *The Great Gatsby*.

As shown in this thesis, the application of NLP on literary textual analysis is still limited, and while the difficulty of conducting such research is greatly resolved due to the advancement of computing power even in personal computers, certain area of knowledge like programming and neural network are still required. But as technology keeps on advancing, more and newer perspective that does not carry bias can be expected in the future of literature studies.

(3868 words)

Notes

1. Trimalchio is a character in Petronius's *Satyricon*, who earns his wealth through ways most people would find distasteful.

Works Cited

- Bewley, Marius. "Scott Fitzgerald's Criticism of America", *The Sewanee Review*, vol.62, no.2, 1954, pp.223-246. "<https://www.jstor.org/stable/27538346?seq=1>", Accessed 16 October, 2020.
- Devlin, Jacob et al. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding", 24 May, 2019. "<https://arxiv.org/pdf/1810.04805.pdf>". Accessed 20 June 2020.
- Fitzgerald, F. Scott. *The Great Gatsby*, Scribner's, 2018.
- . "Winter Dreams." 1998, *the Board of Trustees of the University of South Carolina*, <https://public.wsu.edu/~campbelld/engl494/winterdreams.pdf>. Accessed 17th December 2020.
- Iaccarino, Anthony. "The Founding Fathers and Slavery", July 28, 2016. Encyclopædia Britannica. "<https://www.britannica.com/topic/The-Founding-Fathers-and-Slavery-1269536>", Accessed 16 October 2020.
- French, David. "America Is Being Pulled Apart. Here's How We Can Start to Heal Our Nation", September 10, 2020. Time. "<https://time.com/5887428/american-political-division>", Accessed 20 November 2020.
- Hedge, Chris et al. *Days of Destruction, Days of Revolt*, Nation Books, 2012.
- Mikolov, Tomas et al. "Efficient Estimation of Word Representations in Vector Space", 2013. "<https://arxiv.org/abs/1301.3781>", Accessed 20 August 2020.
- Mikolov, Tomas et al. "Distributed Representations of Words and Phrases and their Compositionality", 2013. "<https://arxiv.org/abs/1310.4546>", Accessed 20 April 2020.
- Miyamoto, Aya. "Midori no Tomori to Ogon no Kagayaki – *The Great Gatsby* ni Okeru Kahei to Amerika no Yume." (The Green Light and the Golden Glitter : Money and the American Dream in *The Great Gatsby*) *Pacific and American Studies*, vol.9, 2009, pp.160-173.
- Sergeev, Aleksandr. "An Application of Semantic Relation Extraction Models", -, "https://portfolios.cs.earlham.edu/wp-content/uploads/2020/05/Project_Paper_Final.pdf", Accessed 27 July 2020.

Yamanaka, Yuko. “_The Great Gatsby_ ni Okeru Sikisai no Imaji” (The Image of Colors in *The Great Gatsby*) *Studies in the Humanities and Sciences*, vol.49, no.2, 2009, pp.213-225.

Zhang, Yin et al. “Understanding Bag-of Words Model: A Statistical Framework,”
“<https://cs.nju.edu.cn/zhoush/zhoush.files/publication/ijmlc10.pdf>”, Accessed 20 April 2020.