

Lexical Richness as a Lens: Exploring the Influence of Social Class on Authors' Writings*

Qi Er (Emma) Teng Wentao Sun Yang Cheng

March 12, 2024

In this study, we explore the influence of social class on lexical richness in the works of authors after the Great Depression, utilizing a comprehensive analysis of lexical data from literary works and social context of authors. Our findings reveal upper-class authors exhibiting a higher Corrected Type-Token Ratio which indicates greater lexical richness, and reflecting their social milieu. This study employs a thorough lexical analysis, aiming to elucidate the complex interplay between class background and language use in literature. This research contributes to our understanding of how social class background shapes language use in literature, offering insights into the broader cultural forces that drive linguistic creativity.

Table of contents

1	Introduction	2
2	Estimand	3
3	Data	3
4	TODO: explain	3
4.1	Source and Methodology	3
5	TODO: explain	3
5.1	Variables	3
6	TODO: explain For better visualization	5
6.1	Measurements	5

*Code and data are available at: <https://github.com/dwz92/Analyzing-the-impact-of-social-class-on-authors-writings>.

6.2	Model set-up	6
7	TODO: formula	6
7.1	Model Justification	6
8	TODO: expectation and why	6
9	Results	6
10	Expalin	6
10.1	Hugh Garners CTTR and Length	6
10.2	David Herbert Lawrence CTTR and Length	6
10.3	Nancy Mitford CTTR and Length	6
10.4	Powell Anthony CTTR and Length	6
11	Discussion	7
11.1	Findings	7
11.2	Economic Impact Insights	7
11.3	Societal and Technological Influences	7
11.4	Weaknesses and Future Research Directions	7
	Appendix	8
A	Data Manipulation and Cleaning	8
B	TODO: explainn	8
	References	9

1 Introduction

The Great Depression was a major economic event of the 20th century that had a profound effect on the global economic situation and class divisions. The question of whether the writing styles of different authors are affected by their class of origin and the general economic climate deserves an in-depth discussion. The purpose of this paper is to gain a deeper understanding of how people lived and thought after the Great Depression by examining whether upper-class writers had a richer writing vocabulary to utilize in their essays than middle-class writers.

Utilizing a combination of quantitative data analysis and authorship studies, we investigated the vocabulary richness of texts used by upper-class and middle-class authors after the Great Depression. Our findings suggest that works produced by upper-class authors during the period 1930-1990 will have a richer vocabulary compared to middle-class writers. This phenomenon

suggests that different classes of writers have different writing styles, and that economic downturns and international instability have had an impact on writers' writing styles. It also had a profound effect on the writers' personal lives, including education, life and thought. These influences, brought about by the broader context of the times, affect not only the choice of vocabulary used in the essay, but also the expression of the central idea of the whole essay, which in turn affects the minds of many civilian readers.

The paper is structured to facilitate a comprehensive understanding of the study and its implications. Following Section 1, Section 3 presents the data, detailing the data sources, analytical techniques, and the rationale behind the chosen methods. Section 9 discusses the results, elaborating on the observed trends and patterns in the authors writings. Section 11 provides an in-depth discussion of these findings, exploring potential factors influencing these trends, drawing connections to broader socio-economic issues, and providing suggestions for future research in this area.

2 Estimand

In this study, our main object of estimation is the causal effect of different classes on the vocabulary use of writers' writing, i.e., changes that can be attributed to the writers' own class of origin, divorced from the economic factors of their times. This involves analyzing the richness of the writing vocabulary of authors from both the upper and middle classes, in order to understand how differences in education, life, and personal perceptions brought about by different class origins affect writers' ideas and writing styles.

3 Data

4 TODO: explain

4.1 Source and Methodology

5 TODO: explain

5.1 Variables

#TODO: explain raw text, sample raw text

Table 1: First Ten Row of Waste No tear by Hugh Garners

Text

“TO LOOK AT ME, you’d never know that I’d ever been anything but a skid row bum. I haven’t been one long, as years are reckoned down here, and I won’t be one much longer. As soon as I get out of this hospital and get a drink or two into me, I’m going to blow the top off this town. Yeah, I know. I’ve been saying the same thing for months, but this time I mean it.”

“ ”

“In our city the skid row is located ten blocks from Waltham Avenue. I was born on Waltham Avenue thirty-seven years ago, and it took me the last twenty to get to skid row. Twenty years is a third of a lifetime, and my twenty should have been the formative and pleasurable ones of my life. Instead, they were something else again, something I am going to try to put down here. It may help me, if I lay out my life in black type on white paper. It may give me the guts to do the things I’ve got to do.”

“ ”

“When I was a kid Waltham Avenue had not yet become a slum street. It was situated in a poor working-class district, but its tenants were decent people who believed in work, thrift, and of some day moving to one of the new residential districts that were springing up in the suburbs of our city. Many of them moved away over the years when I was a boy, but my family were left behind. It might have been my old man’s drinking that stopped the Matterson family from getting ahead, or, as he claimed, it might have been bad luck. Whatever it was, it kept us chained down to our little rented house on Waltham Avenue. We were still there when Waltham became an industrial street, with the soap factory and the planing mill and the fish wholesaler’s where there had formerly been rows of little houses like ours.”

“ ”

“In the summer the smells from the factories used to fill the streets, and Nicky Manulesco used to stop his bicycle in front of our steps and sniff the air. “Smell that, Tom?” he would ask. “That goddamn soap company is cleaning its vats again. Between the smell of the soap and the smell of the fish from down the street, this street stinks like bath night in a girls’ school.”

“ ”

“When I was fourteen I left school and got a job riding a bicycle for the telegraph company. It was good to get away from school, and the money I made helped us at home. By this time my old man was a confirmed lush, and he had stopped working with the first day of the Depression. From then on the Matterson family lived on relief hand-outs and from the money I made delivering telegraphs.”

“ ”

Text

“My mother might have been young once, but she was old from the first time I remember her. When I think of her now, I see a little beaten-up old woman working over the coal range in the kitchen or scrubbing the tattered linoleum in the front hall. She’d had a hard life, and had dragged up four kids, of which I was the eldest. Besides my sister Elsie, who was two years younger than me, there were two other boys, Eric and Paul. My mother died in 1930, and my two young brothers were taken away by the Children’s Aid. Elsie went to live with an aunt in another part of town, and my old man disappeared. I learned years later that he had died in a Midwest jail. I moved in with a friend of mine, Alec Ranning.”

Table 1, created with **kableExtra** (Zhu 2021), showcases the first ten rows of #TODO: Explain raw data

6 TODO: explain For better visualization

Table 2: First Ten Rows of Authors’ Corrected Type Token Ratio

author	title	corrected type token ratio
David Herbert Lawrence	england_my_england	21.4
David Herbert Lawrence	kangaroo	40.3
David Herbert Lawrence	lady_chatterleys_lover	35.1
David Herbert Lawrence	odour_of_chrysanthemums	20.5
David Herbert Lawrence	sons_and_lovers	35.2
David Herbert Lawrence	the_plumed_serpent	37.0
David Herbert Lawrence	the_rainbow	35.1
David Herbert Lawrence	the_virgin_and_the_gipsy	28.3
Hugh Garners	waste_no_tears	21.9
Hugh Garners	the_conversion_of_willie_heaps	15.0

Table 2 # TODO: Explain

6.1 Measurements

#TODO: explainnnnn

#Model # TODO: Explain neg binom model

6.2 Model set-up

7 TODO: formula

7.1 Model Justification

8 TODO: expectation and why

9 Results

10 Expalin

Section [9](#)

10.1 Hugh Garners CTTR and Length

10.2 David Herbert Lawrence CTTR and Length

10.3 Nancy Mitford CTTR and Length

10.4 Powell Anthony CTTR and Length

11 Discussion

The findings of this study offer a layered perspective on the decline in birth rates among young people in the United States, particularly in the context of the Great Recession. Our analysis provides an understanding of the multifaceted influences on demographic trends.

11.1 Findings

11.2 Economic Impact Insights

11.3 Societal and Technological Influences

11.4 Weaknesses and Future Research Directions

One limitation of our study is the potential influence of unobserved variables that may affect the lexical richness in authors' works, such as individual educational background or specific life experiences, which were not fully captured in our analysis. Moreover, while we used an improved TTR formula, its inherent sensitivity to text length variability and the possibility of other underlying linguistic factors were not entirely addressed. These could include stylistic choices or genre-specific language use that our study does not explore in depth. Future research should aim to unravel these complex interactions further, perhaps by integrating more granular data on authors' personal histories or by employing a wider range of linguistic analysis tools. Additionally, exploring the application of advanced computational methods, such as natural language processing algorithms, could offer more detailed insights into how social class impacts lexical choice and style in literature. Understanding these nuanced relationships is vital for comprehensively interpreting the intersection of social factors and literary expression.

Appendix

A Data Manipulation and Cleaning

B TODO: explainn

References

Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*.
<http://haozhu233.github.io/kableExtra/>, <https://github.com/haozhu233/kableExtra>.