From Homer to Goethe and Beyond to AI

Jakob Margolis

From Homer to Goethe and Beyond to AI

A New Comparative Approach to the Study of Poetic Language

Wissenschaftlicher Verlag Berlin
wvb

Bibliografische Informationen der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.ddb.de abrufbar

Bilder auf dem Umschlag:

Homer Caetani, Louvre Museum ARK ID 010277994, Fotografie von Marie-Lan Nguyen. Wikimedia Commons

Bust of Virgil by Tito Angelini, 1930, Vergiliano park (Naples), Flickr Parco della Grotta di Posillipo. Wikimedia Commons

Porträt des Dante Alighieri, idealisierende Darstellung von Sandro Botticelli, um 1495, Wikimedia Commons

Der Porträt-Stich Shakespeares von Martin Droeshout auf dem Titel der „First Folio“ (1623) Shakespeares Unterschrift, Wikimedia Commons

Karl Joseph Stieler, Johann Wolfgang von Goethe at age 79, Bayerische Staatsgemäldesammlungen, Wikimedia Commons

Orest Adamowitsch Kiprenski, Porträt des Alexander Pushkin, 1827, Tretjakow-Galerie, Wikimedia Commons

The author's photo on the back cover of the book is the work of Viktor Starshenko

ISBN 978-3-96138-447-1

© 2025 Wissenschaftlicher Verlag Berlin
Olaf Gaudig & Peter Veit GbR
www.wvberlin.de / www.wvberlin.com
Alle Rechte vorbehalten.
Dieses Werk ist urheberrechtlich geschützt.
Jede Verwertung, auch einzelner Teile, ist ohne Zustimmung des Verlages unzulässig. Dies gilt insbesondere für fotomechanische Vervielfältigung sowie Übernahme und Verarbeitung in EDV-Systemen.

Druck und Bindung: SDL – Digitaler Buchdruck, Berlin
Printed in Germany
€ 46,00

This innovative volume provides a comparative and diachronic analysis of the main works of outstanding European poets of different eras, nationalities, cultures, and religions. By tracking the use of the most common nouns (key concepts) within and between the works of Homer, Virgil, Dante, Shakespeare, Pushkin, and Goethe (in their original languages), the author reveals previously unknown facts and hidden patterns in their language.

The book takes as its starting point the axiom of French poet Charles Baudelaire: "To understand the soul of a poet, or at least his main concern, let us search in his works to see which word or words are most often represented there."

By tracing the evolution of key concepts through their frequency of use in the texts of outstanding European poets over time, the reader gains access to areas of poetic consciousness that were previously inaccessible.

This type of study has only become possible in the digital era of the 21st century thanks to the advent of new computer technologies and linguistically marked corpora of poets' texts. This book can serve both as a textbook and a long-term reference work for students and scholars specializing in computational and corpus linguistics, cross-linguistic comparative literature, lexicology and semantics, poetics, and cultural studies.

In addition, the book examines and compares poems written by various artificial intelligence programs: ChatGPT, Gemini, Claude, and DeepSeek.

'National literature is not worth much now. We are now entering the epoch of world literature, and everyone must now help to hasten the appearance of this epoch.'

Goethe, J. W., Gespräche. Mit Johann Peter Eckermann, 31. January 1827

"It is possible to see European literature as a whole only if one acquires the right of citizenship in ***all its eras from Homer to Goethe****."*

Curtius E. R. Europäische Literatur und lateinisches Mittelalter. Bern: Francke, 1948

"To understand the soul of a poet, or at least his main concern, let us search in his works to see which word or words are most often represented there."

Charles Baudelaire, Collected Works, vol. 3, Paris, 1966.

"There is as much truth in science as there is mathematics."

Immanuel Kant "Metaphysische Anfangsgründe der Naturwissenschaft", 1786.

"If what distinguishes us from the rest of the animal kingdom is speech, then poetry is the highest form of speech — a kind of genetic difference from animals. Moreover, poetry is a colossal accelerator of consciousness, both for the writer and the reader. ***You discover connections and dependencies you didn't even suspect existed — data embedded in language and speech. Poetry is a unique tool of cognition****."*

Joseph Brodsky, Interview with Moscow News, No. 50, June 23–30, 1995.

"People would get rid of half their troubles if we could agree on the meaning of words."

René Descartes, *Discourse on the Method for Rightly Directing One's Reason and Searching for Truth in the Sciences*, Leiden, 1637.

*"****There are no facts, only interpretations****." („Es gibt keine Tatsachen, nur Interpretationen.")*

Friedrich Nietzsche, 'Notes' 1888, S. 458.

"All knowledge is gained by comparison."

A catch Phrase

"I may be wrong and you may be right, and by an effort, we may get nearer to the truth."

Karl Popper, „The myth of the framework", 1994.

I dedicate the book to my children, who speak many European languages.

Author

TABLE OF CONTENTS

Preface by the author 15

Acknowledgments 19

INTRODUCTION. WHY AND FOR WHOM WAS THE BOOK WRITTEN? WHAT'S NEW IN THE BOOK? 23

CHAPTER 1. METHODOLOGY FOR RESEARCHING THE FREQUENCY OF KEY CONCEPTS AND LANGUAGE EVOLUTION IN THE DIGITAL AGE. THE METHOD OF KEY CONCEPTS. 25

1.1 Object of study and boundaries of the language of outstanding European poets: Homer, Virgil, Dante, Shakespeare, Goethe, Pushkin, and their works 26

1.2. Definitions of poetry, key concepts, frequency lists 28

1.3. Setting the aims and objectives of the study, the problems encountered in the case 30

1.4. Postulates of the study. Correspondence with Chomsky and Melchuk 32

1.5 New opportunities for studying the language of poets in the digital era 43

1.6. Comparison of key concepts in different languages, eras, cultures, and religions 44

1.7 Evolution of the first and second kinds of key concepts in the languages of the poets 47

1.8. Zipf's Law, the well-known notion of hapax and the new notion of lemmax 48

1.9. Cores and categories of key concepts in the languages of the poets, two types of lemmatizations: 1) by word; 2) by the meaning of the word in the context 49

1.10. Poetic corpora of texts, frequency, and rank of lemmas 51

1.11. Empirical poetics: facts first, then interpretations 52

1.12. Stages of the Methodology for Researching the Frequency of Key Language Concepts 53

CHAPTER 2. HOMER: COMPILATION AND COMPARISON OF FREQUENT LISTS OF NOUNS, KEY CONCEPTS OF THE ILIAD AND THE ODYSSEY 59

2.1. Sources of ancient Greek texts, and the corpora of the Iliad and Odyssey 60

2.2. Frequency lists of lemmas of nouns in the Iliad and Odyssey..... 61

2.3. Analysing changes in key concepts from the Iliad to the Odyssey 64

2.4. Mandelstam's Hypothesis, or Is Homer's Language Driven by Love? 65

2.5. Poetic letter from Odysseus to his son Telemachus from the European Football Championship 2004 67

CHAPTER 3. VIRGIL: COMPARATIVE ANALYSIS – KEY CONCEPTS IN POETIC LANGUAGE 69

3.1 Sources of Latin texts and corpus of Virgil's works..... 70

3.1.1 Source Latin texts and corpus of LASLA Opera Latina 70

3.1.3 Open Latin Perseus corpus, Digital Library Scaife Viewer (SV) .. 71

3.2. Frequency lists of key concepts in Virgil's poems 71

3.3. Comparative Analysis of Key Concepts in the *Bucolics*, *Georgics*, and *Aeneid*: The Evolution of Virgil's Poetic Language and Themes Over Time 73

3.4. Evolution of the lemma "*life*" (vita) and the lemma *"death*" (letum) 74

3.5. The category 'nature' in Virgil's poetic language 76

3.6. The category 'metaphysics': 'fate' and 'luck' in Virgil's language 77

3.7. Male proper names in the language of Virgil..... 80

3.8. Comparison of Key Concepts Homer's Odyssey vs. Virgil's Aeneid 83

CHAPTER 4. DANTE'S *DIVINE COMEDY*: HELL, PURGATORY, PARADISE — COMPILATION AND COMPARISON OF FREQUENCY LISTS OF KEY CONCEPTS 85

4.1 Sources of the Italian texts of La Divina Commedia..... 86

4.2. Frequency lists of noun usage in Dante's Divine Comedy (DC): 87

4.3. Dante's Sieve, or separating the 'grains from the chaff' 87

4.4. Increase of positive and decrease of negative lemmas from Hell to Paradise 89

4.5. Classification of Lemmas in The Divine Comedy by Frequency Change Dante's Paradox: The Decreasing Frequency of the Lemma "Soul" from Hell to Paradise 92

4.6. "Dante's Scissors": evolution of the lemmas 'love' and 'death' 96

CHAPTER 5. SHAKESPEARE: COMPILATION AND COMPARISON OF KEY CONCEPT FREQUENCY LISTS IN THE SONNETS 99

5.1. Folger Shakespeare, Open-Source Shakespeare, and Sketch Engine corpora 99

5.2. Frequency lists of key concepts in Shakespeare's sonnets 100

5.3. The 'love' lemma and the separation of the 'grain from the chaff', the noun from the verb 101

5.4. An Archaeology of Feeling in Shakespeare's Sonnets: Which Emotion Most Frequently Motivates the Poet's Use of the Key Concept 'Love'? .. 102

5.5. The words *kiss* and *lips* in the sonnets addressed to the "friend" and to the "dark lady" 104

5.6. A phenomenon: the 'disappearance of time' in Shakespeare's sonnets, or 'The time is out of joint' 105

5.7. Analogies between tragedies-comedies in Shakespeare's plays and Hell-Paradise in Dante's Divine Comedy 107

CHAPTER 6. GOETHE: COMPILING AND COMPARING THE FREQUENCY LISTS OF 'URFAUST', 'FAUST, PART ONE', AND 'FAUST, PART TWO' 109

6.1. The Goethe Dictionary, Faust source texts, and the Sketch Engine corpus 109

6.2. Frequency lists of noun lemmas in Goethe's Faust 113

6.3. Changes in key concepts in the language of young, mature, and elderly Goethe 114

6.4. Multiple meaning lemmas in Goethe's Faust 117

6.5. Analogy of the changes in the lemmas ‘life’ and ‘death’ in Goethe's Faust and Dante's Divine Comedy 118

6.6. The frequency of proper names in Goethe's Faust 121

6.7. The lemmas ‘spirit’ and ‘soul’ in Goethe's Faust 121

6.8. Results and perspectives of the study of Goethe's Faust 124

CHAPTER 7. PUSHKIN: COMPILATION AND COMPARISON OF THE FREQUENCY LIST OF KEY CONCEPTS OF THE POET'S ENTIRE POETIC CORPUS 129

7.1. The poetic corpus of the Russian language and its possibilities 130

7.2. Frequency lists of key concepts of Pushkin and Russian poets of other epochs 130

7.3. The change of key concepts of Russian poets of the 18th – 21st centuries 131

7.4. The core of key concepts in the language of Russian poets 133

CHAPTER 8. COMPARATIVE ANALYSES AND CHANGES IN KEY CONCEPTS OF THE EUROPEAN POETS BY TIME AND FREQUENCY FROM HOMER TO GOETHE AND BEYOND 137

8.1 Frequency lists of European poets from Homer to Goethe and Pushkin .. 137

8.2. Changes in the frequency of key concepts in the language of European poets 140

8.3. The paradox of the lemma ‘soul’ in Dante, Shakespeare, and Pushkin 141

8.4. The cores of key concepts in the language of European poets of different nations, epochs, religions, and cultures 142

8.5. Comparison of semantic categories of key concepts in the language of European poets 146

8.6. Frequency of mention of the lemma ‘God’ in the languages of European poets 148

8.7. Lemmaxes in the works of European poets 151

8.8. Poets' frequency lists and a new argument to the ’Homeric question’ 152

8.9. References to women's names in the works of European poets 156

8.10. Comparison of key concepts in European and Chinese poetry 159

CHAPTER 9. POETIC LANGUAGE AND ARTIFICIAL INTELLIGENCE:
A COMPARISON OF THE RESPONSES OF ARTIFICIAL INTELLIGENCE
WITH THE RESULTS OF THE RESEARCH IN THE BOOK 165

9.1. Questions to ChatGPT about this book and answers from ChatGPT 165

9.2. Questions to Gemini about this book and answers from Gemini 168

9.3. A comparison of the key concepts of the poems written by AI 170

9.4. AI Claude and its poems 176

9.5. DeepSeek – Chinese AI model..... 179

9.6. New possibilities for researching poetic language with the help of AI include the Synthesis of poetic styles from different languages, historical periods, mentalities, and cultures. 181

CHAPTER 10. FOUR EXAMPLES OF THE STUDY OF POETIC LANGUAGE IN
THE DIGITAL AGE OF THE 21ST CENTURY 187

10.1. Example 1: The Colours of the Heavenly Rainbow in Poetry..... 187

10.1.1. Colours in the Poetry of 18th-Century Poets 189

10.1.2. Colours of poets of the 19th century - the Golden Age of Russian poetry..... 191

10.1.3. Colours of Silver Age poets of Russian poetry..... 193

10.1.4. Colours of poets of the second half of the twentieth century 195

10.1.5. Colours of Russian poets of the XXI century 197

10.1.6. Colours of Russian-language foreign poets of the twenty-first century 198

10.1.7. Colour potentials of different epochs of Russian poetry 199

10.1.8. Language energy of different epochs of Russian poetry..... 200

10.1.9.1. Colour Holes in the Language of Poets 203

10.2. Example 2: Egocentrism in Poetic Language 205

10.2.1. Introduction: psychological and poetic egocentrism 206

10.2.2. Egocentrism in the language of eighteenth-century poets 208
10.2.3. Egocentrism in the language of poets of the Golden Age of Russian poetry 209
10.2.4. Egocentrism in A. Pushkin's poetic language before and after important events 210
10.2.5. Egocentrism in the Poetic Language of the Silver Age of Russian Poetry 212
10.2.6. An example of the effect of an event on the egocentrism of a Silver Age poet 213
10.2.7. Egocentrism in the language of poets of the second half of the twentieth century 214
10.2.8. Egocentrism in the language of XXI century Russian poets 215
10.2.9. Egocentrism in the language of Russian-speaking foreign poets of the XXI century 216
10.2.10. The Relationship Between Colours and the Egocentrism Coefficient in the Language of Poets from Different Epochs 218
10.3. Example 3: The Lemma “love” and Its Semantic Context in Poetic Language 220
10.3.1. The lemma “love” and its surroundings in poetic language 221
10.3.2. Law of Universal Gravitation in the Universe of Poetic Language (joke) 221
10.3.3. Semantic connections of the lemma “love” in the poems of poets 222
10.3.4. Nouns and other parts of speech of poets of different epochs in the neighbourhood of the lemma ‘love’ 223
10.3.5. Frequencies of use of the lemma ‘love’ by years of poets' lives .. 224
10.4. Example 4. Lexical norm of poetic language 227
AN AFTERWORD AND PROSPECTS FOR FURTHER RESEARCH 231
LIST OF REFERENCES 234
Terms and abbreviations 236