Computational Classics: Homer in Translation and the Science of Stories

A Thesis Presented

by

Tessa Lawler

to

The Faculty of the Graduate College

of

The University of Vermont

In Partial Fulfillment of the Requirements
for the Degree of Master of Science
Specializing in Complex Systems and Data Science

May, 2026

Defense Date: March 18th, 2026
Thesis Examination Committee:

Peter Sheridan Dodds, PhD, Advisor
Chris Danforth, PhD, Co-Advisor
Jacques A. Bailly, PhD, Chairperson
Holger Hoock, DPhil, Dean of the Graduate College

© Copyright by
Tessa Lawler
May 2026

Abstract

In the 8th century BCE, the Greek poet Homer is said to have composed the epic poems known as the *Iliad* and the *Odyssey*. Nearly 3,000 years later, audiences are still captivated by Homeric epic, in large part due to the classicists who have translated these texts into modern languages. In the 21st century, audiences engage with these texts almost exclusively through translation and adaptation, with translators serving as the intermediaries between the truth of the original stories and a contemporary audience's conception of them. Given the enduring significance of these ancient texts, it is critical to understand the approach translators take to their work and how popular translations of Homeric epic differ.

Recent advances in the scientific study of stories present opportunities for a computational approach to this work. Allotaxonometry, the study of the structure of complex systems through comparison of the prevalence of their component parts (Dodds et al., 2023) and hedonometry, the study of the emotional arcs of stories (Reagan et al., 2016) are two developments of particular relevance. In this work, we used these methods to compare six translators' versions of the Iliad and Odyssey. Our aim was to find out whether these were appropriate tools for this type of analysis and what they would reveal about the ways in which the translations in our sample differed.

The challenges of using these methods to compare texts composed over millennia of human history were considerable. To identify differences beyond the most superficial, orthographic level (e.g. the spelling of character names, the frequency of use of contractions, etc.), it was necessary to develop and implement a multi-step standardization process which temporarily suppressed these differences so the texts could be more deeply and holistically compared. In the end, we found that these methods can detect meaningful and interesting differences between translations of Homeric epic. The allotaxonometric analyses were particularly fruitful, with findings that both support and further existing work in the study of Homer in translation. Contextualized within broader classical scholarship, these results illuminate artistic, philosophical, linguistic, and political differences between translations of Homeric epic and demonstrate that computational tools can meaningfully contribute to this work.

This thesis is dedicated to my beloved grandfather, Edwin J. Lang (1937–2020), who told me to never stop learning.

Acknowledgements

I would first like to thank my committee. My advisors, Peter Dodds and Chris Danforth, provided endless support and encouragement as I pursued this project, and I am truly indebted to them. I'd like to also thank Jacques Bailly, my committee chair, for his willingness to be the classicist in the room for my defense.

I'm grateful for many current and former students at the University of Vermont. In particular, I'd like to acknowledge my primary collaborator on the work in this thesis, Sarah Phillips, whose contributions, insight, and friendship have been invaluable, and Lucy Hart, my collaborator on some upcoming work in this field. I'd also like to thank everyone at the Vermont Complex Systems Institute and the Computational Story Lab for their encouragement, feedback, and assistance as I learned to navigate the field of complex systems: specific shout-outs are due to Julia Zimmerman, Calla Beauregard, and Ashley Fehr.

While on the topic of UVM, I would be remiss if I did not mention that I was able to pursue this degree because of the University's tuition remission policy for full-time employees, which allowed me to take all the classes I needed for this program for free.

The support of friends and family during this process has been essential. I'm grateful to Reese, Timmy, Cari, and the wonderful friends I've made in Vermont: Betsy, Jon, Tuka, Emmanuel, and Bailey. A million thank you's to my family, who have always supported my education: my brother Logan, my parents, Deborah and Colin Lawler, Papa and Grandma Kay, Grandma Nancy and Grandpa Ivor, Aunt Michele and Uncle Andy, Aunt Sharon and Uncle Phil, my cousins Jake and Conor, and Grandma Allegray and Grandpa Bernie. I'd also like to shout out my many animal friends, past and present: Tank, Casimir, Destin, Annie, Freya, Bear, Pepe,

Mack, Leda, and all the other good individuals I've met at Fare Thee Well Farm.

I wish to acknowledge some of my former professors in the fields of antiquity. While at the University of California, Los Angeles, I took a course from Professor Kara Cooney on "Women & Power in Ancient Egypt" that I believe permanently shifted something in my brain and ignited my interest in the intersection of feminism and antiquity. I also want to thank Professor Jennifer Weintritt at Northwestern University. It is highly unlikely this project would ever have occurred to me if I hadn't taken her course, "Women of the Trojan War: Ancient and Modern Adaptations," in the Winter of 2020. Professor Weintritt was also the one to tell me that classics could be a part of my life regardless of what profession I ultimately landed in, for which I will always be grateful.

I want to acknowledge that this work would not exist without the efforts of scholars who, for millennia, have worked to preserve and study the texts on which this thesis is based. In particular, we owe a debt of gratitude to the classicists who have translated ancient texts into modern languages—they have opened up new worlds to so many of us.

Today, the field of classical studies is under serious threat due to the cumulative effects of decades of austerity. The continued survival of the field is attributable to the tireless efforts of the contemporary classicists who are out there fighting the good fight. We all owe them a great deal.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

> "Tell me about a complicated man. / Muse, tell me how he wandered and was lost / when he had wrecked the holy town of Troy, / and where he went, and who he met, the pain / he suffered in the storms at sea, and how / he worked to save his life and bring his men / back home. He failed to keep them safe; poor fools, / they ate the Sun God's cattle, and the god / kept them from home. **Now goddess, child of Zeus, / tell the old story for our modern times. / Find the beginning.**"
>
> *The first 10 lines of Emily Wilson's 2018 translation of* the Odyssey. (Wilson/Homer, 2018, p.105)

1.1 Introduction to Homeric Epic

Sometime around the 8th century BCE, a child called Homer is said to have been born. Little is known today about Homer's life; indeed, even the Greeks living mere centuries after Homer's death seem to have known very little. Homer was perhaps

born in Ionia, then an Ancient Greek territory in Western Anatolia, now a region in modern-day Türkiye. He was reported to have been blind. Most importantly, he is credited with the creation of the epic poems *the Iliad* and *the Odyssey*, two of the most enduring works of the Western literary canon. These stories are known today as the Homeric epics, and their profound significance to Western society mean that even today, Homer can be referenced in certain circles simply as "The Poet."

The Homeric epics are two distinct but related stories. *The Iliad* is a war story. It covers a roughly two-month period towards the bitter end of the Trojan War, a mythical conflict started over the beautiful Helen of Troy and fought by the Trojans, the Achaeans, and their respective allies. The story starts when a conflict within the conflict erupts: Agamemnon, a leader on the Achaean side, confiscates a woman named Briseis who had previously been given as a war-prize to Achilles, the most-skilled warrior on the Achaean side. In response to this grave insult to his honor and pride, Achilles refuses to fight with the rest of the Achaean troops, giving the Trojan side a tangible advantage in the war. Achilles' decision and its consequences for everyone in the war are the subject of the remainder of the poem. It is a rather brutal story about the devastation of war, the costs of pride, rage, and vengeance, and perhaps most of all, the futility of resistance in the face of fate.

The Odyssey is the story of one man's ten-year journey home from the war. The poem follows Odysseus, a Trojan War veteran of the Achaean side, as he desperately tries to return home to Ithaca despite countless obstacles set in his path: by angry gods, monsters, his own men, and, not infrequently, himself. Odysseus is a complicated character, and as such, so is his *Odyssey*. There are moments of humor and levity along with terrible violence and grief. At its core, it's a story about everything

one might sacrifice for the chance to come home.

There is a great deal of mystery surrounding the creation of *the Iliad* and *the Odyssey*. The degree of Homer's involvement itself is a topic of great debate: the so-called "Homeric Question" has occupied scholars since ancient times. Was Homer the sole author of the poems? If not, how involved was he in the creation of the Iliad and Odyssey? Did Homer, the man, even exist? We may never know the answers to these questions about the epics' authorship, but we do have a good deal of evidence of how the stories were transmitted. *The Iliad* and *Odyssey* are believed to be the product of an oral storytelling tradition. Our earliest extant reports of the epics indicate they were spread acros Ancient Greece by word of mouth for hundreds of years before ever being written down in full. The poems were performed by people known as *rhapsodes*, who memorized episodes from the texts (e.g. *Patrokleia*, "The Patroclus Episode" from the Iliad) (Tsagalis, 2020b) and performed these smaller sections of the poems for audiences due to the sheer length of the full poems.

Information on exactly when and how the Iliad and Odyssey came to be written down is hard to come by. It is unlikely complete written texts existed in full during these first few centuries. Presumably Homer himself would not have been able to write them down himself—as a blind man, he was almost certainly illiterate—but he would not have been the only one thus limited. The Greeks were relatively late to literacy in antiquity, and the standard Ancient Greek alphabet was just beginning to come into use at the time the Homeric epics are thought to have been composed.

Some scholars, buoyed by ancient primary sources like Cicero and Aelian, believe the tyrant ruler Pisistratus commissioned the compilation and written recording of the epics in their entirety in advance of their performance at one of Athens' Panathenaian

festivals in the 6th century BCE. While others dispute this specific theory, called the Pisistratean Recension (Finkelberg, 2017; Tsagalis, 2020a), it is likely the poems were written down for one reason or another around this time.

Despite this, the earliest extant written evidence of Homeric epic (often referred to as *Homeric papyri*) is only from the 3rd century BCE, around 300 years later. It is thought that stable written copies of the epics were generated around this time by Alexandrian scholars. These stable copies would have been used to make many more copies, which in turn would have been copied themselves. More than 1,000 Homeric manuscripts exist today, making the Iliad and Odyssey two of the most well-preserved stories of the Ancient World (The Hanna Holborn Gray Special Collections Research Center, 2014). Despite this, the oldest extant complete manuscript of the epics is only about a thousand years old. These large jumps in time—from when the epics were composed to when they were first written down, from when they were first written down to when the oldest extant Homeric papyri was produced, and from the date of that papyri to the oldest extant complete manuscript—define the history of these texts and contribute to the mystery that surrounds them.

1.1.1 Lost and Found: Homer in Translation

Many are surprised to learn that the Homeric epics were not translated in their entirety into modern languages until the middle of the 16th century CE. The English playwright George Chapman (1559–1634) was the first to translate the full *Iliad* and *Odyssey* into English in 1615. This act opened up the world of Homer to an audience who had previously not had access to it, and the invention of the printing press two centuries prior meant these texts could be more cheaply and rapidly produced than

was previously possible. More translators would follow Chapman's example. Alexander Pope, whose translations are likely the most enduring in English, published his translations of *the Iliad* and *the Odyssey* a century later. More complete translations of the Homeric epics emerged over the next century, and by the 18th century the business of Homer was booming: over the course of the last two hundred years, several scores of translators—possibly more than 100—have applied themselves to the task of producing English versions of the epics. (Wilson, 2023)

Translating any text is a complex, complicated task. But translating Homer into English is perhaps especially daunting. The temporal distance between the world of Homer and that of the modern translator is naturally a significant impediment. Translators are best served by having a deep understanding of the cultural context of the original work, and many see their goal as to reproduce for their audience the feeling the original evoked in its own. But Homer's audience and world are long gone. By the time George Chapman published his translations of the epics in 1615, roughly two-and-a-half millennia had passed since their creation, and any hope of understanding with certainty how Homer's words were heard and felt by the Ancient Greeks has passed with them.

There are also stylistic and structural features which make the Homeric epics especially difficult to translate. The poems were composed in a poetic meter called dactylic hexameter that is considered particularly difficult to render in English. Most translators try to replicate the effect of Homer's poetry by rendering it in some other kind of poetic verse: Alexander Pope famously wrote his 1715 and 1725 translations in rhyming heroic couplets, while Emily Wilson used iambic pentameter for her translations in 2023 and 2018. Choosing to translate in any form of poetic meter poses

a challenge, as the most "literal" translation of a Greek line in Homer is unlikely to fit neatly into the constraints of the corresponding metered line. Some translators circumvent this entirely by translating Homer in prose instead of poetry, though an important Homeric quality is undeniably lost in the process.

The selection of meter and the decision to translate in poetry or prose are just two of the countless choices translators of Homer must make in the course of their work. Though these may seem mostly practical or aesthetic in nature, all the decisions translators make are necessarily downstream of their beliefs about antiquity, their philosophy on translation, and even their politics. A translation of Homer is influenced by these qualities of the translator and the characteristics of the society in which they and their audience lived. Given this, it is perhaps unsurprising that two translations of the Homeric epics can be so dramatically different.

1.1.2 Homer's Relevance Today: Contemporary Homeric Reception

Today, the legacy of Homeric epic is apparent in every quarter of modern media: from cinema to manga, from board games to video games, and from stories on the online fanfiction repository Archive of Our Own (AO3) to traditionally-published literary adaptations (Margaret Atwood's *The Penelopiad*, Madeline Miller's *The Song of Achilles*, and Alice Oswald's *Memorial* are just a few examples in a genre which has exploded in popularity in recent years). (Jenkins, 2020a)

While Homeric epic may be benefiting from something of a 21st-century revival, its relevance to a modern society is not a recent development. The persistent influence

of Homer on the storytelling and artistic tradition of Western civilization cannot be understated. As Jorge Luis Borges wrote in his 1932 essay *Las Versiones Homéricas*: "Our first reading of famous books is really the second, since we already know them. The cliche 'rereading the classics' turns out to be an unwitting truth." (Borges & Levine, 1992) Homer's influence is such that people today who have never sat down to read the epics nevertheless know many of the major beats.

There are many reasons *the Iliad* and the *Odyssey* have enjoyed such longevity of relevance. The classicist Thomas E. Jenkins argues that the surge of interest in the epics in the twentieth century is attributable to the fact that people were living through the deadliest conflicts in human history and thus found a story about the eternal devastation of war highly relevant to their own lives. (Jenkins, 2020c) Indeed, connections to the two world wars of the twentieth century are common in Homeric reception of the time. On the cover of Stanley Lombardo's translation of the Iliad (2000) is the famous D-Day photograph "Into the Jaws of Death." Of this choice, Lombardo has written, "I have found that even today's high-school students recognize the photograph and understand its presence on the cover not as a bid for 'relevance' but as a symbol of the universality of war." (Lombardo, 2010, p. 227)

Jenkins expands on his thoughts on the universality of Homeric reception in a chapter for the Cambridge Guide to Homer titled, "Homer in Social Thought:"

> "A history of the reception of Homer is, at the same time, a history of nearly all social currents: receptions of Homer reflect (or contest) the whole gamut of Western social thought, including feminism, modernism, existentialism, and conservatism. Sometimes *the Odyssey* attracts receptions because its themes seem to run entirely contrary to contemporary

> social trends: an artist's new 'version' of Homer becomes an act of resistance, whether overt or covert." (Jenkins, 2020b, p. 526)

Classicist Emily Greenwood argues that the themes of *the Odyssey* are not necessarily out of line with "contemporary social trends". She writes that "postcolonial receptions of the Odyssey can be plotted between two poles. At one pole, Odysseus is construed as a protocolonizer who twists words and stories to legitimize lies, violent raids, piracy, and high-handed treatment of foreign peoples," but that

> "At the other pole, in postcolonial literature Odysseus has also been evoked as a figure for the wretched of the earth, the refugee, the exile, the homeless migrant, and the wily trickster figure of post- and anticolonial resistance. In particular, both fictional and nonfictional narratives have used the plot of the Odyssey to dramatize the contemporary refugee crisis, evoking Odysseus's wanderings as a paradigmatic example of involuntary migration." (Greenwood, 2020, pp. 532–533)

Homer's continued relevance, then, is entirely comprehensible, if still very impressive. These stories are very much ones for our modern times. While the above are explanations for *why* Homer's works are still relevant, they do not explain *how* it is so. Simply put, the mechanism by which Homer is still relevant today is translation. In a time where hardly anyone reads Ancient Greek and the institutions which train people to do so are being cut down left and right, translators have worked to democratize access to these texts and to keep them in the hearts and minds of a contemporary audience. By extension, the reliance on translation by contemporary audiences means understanding how these translations differ is critical. Translators are the intermediaries between the original Homeric texts and a modern audience's interpretation of

them, so it is critical to understand the ways in which the translations we read shape our understanding of Homer's world.

1.1.3 The Science of Stories: Computational Approaches

Translation studies is well-established within the broader classical studies field, and work which evaluates different translations of classical texts is abundant. Computational approaches to this work, however, are very much an emerging area of the field.

Previous work in applying computational methods to translation studies has utilized a variety of methodologies and linguistic measures. Lucic and Blake, for instance, used the Stanford Lexical Parser to create sentence-level tokens from two translations of *The Notebooks of Malte Laurids Brigge* by Rainer Maria Rilke, and found measurable differences between translators' syntax and semantics: namely, in their use of negation and prepositional modifiers, and adverb and adjective choice. (Lucic & Blake, 2011) In a comparison of the original Polish and two English translations of Henryk Sienkiewicz's novel series, *The Trilogy*, multivariate analysis of correlation matrices of relative frequencies of the most frequent words in dialogue was found to show that characters' idiolects were differentiated by nationality, social status, gender, and age, and that these patterns remained similar across all three texts. (Rybicki, 2006) Others have borrowed techniques from authorship analysis and evaluated translations by calculating similarity to the source text (Ryabko & Savina, 2024). There have even been instances of these approaches being applied to the study of Homeric

epic in translation: Cynthia Whissel (2004), who studies affective associations with word sounds, conducted an investigation of Alexander Pope's use of sound in his translation of *the Iliad* by scoring the phonemes which comprise each Book of text to ascertain their associated emotional character. Daniel Russo, in 2024, explored the representation of female characters in Samuel Butler's translation of the Odyssey using corpus-based methods, and found that female characters were prominently associated with domestic and relational modifiers, while male characters were associated with terms of power and authority.

Recent developments in the science of stories have resulted in new computational tools which have potential applications for the study of Homer in translation. Allotaxonometry, for instance, is a method of comparing two systems (e.g. economies, ecosystems, stories) based on rank-to-rank pairs of their components parts. (Dodds et al., 2023) Previously, allotaxonometry has been used to find words that were most topically and culturally relevant during the COVID-19 Pandemic, (Alshaabi et al., 2021), to track changes in the discussion of mental health on Twitter (Stupinski et al., 2022), and to explore the language used by incels on the social media platform Reddit (Gothard et al., 2021).

Another such tool is semantic analysis, or the study of the meaning of language through computational means. In a pioneering study in this field, Osgood *et al.* (1957) proposed three semantic differentials on which language could be scored: valence, arousal, and dominance (VAD). The first of these, valence, is an early example of *sentiment* analysis, or the computational determination of the emotional valence of language, generally conceived of as a Good/Bad, Positive/Negative, or Happy/Sad spectrum. Much of the subsequent work on semantic analysis was inspired by Os-

good *et al.*'s framing. The ANEW framework, for instance, produced a dictionary of words scored on the VAD paradigm (Bradley & Lang, 1999), which has been used by researchers to semantically analyze everything from song lyrics to State of the Union addresses. (Dodds & Danforth, 2010)

Many such semantic analysis dictionaries now exist (LIWC is another, somewhat notorious example (Pennebaker et al., 2001)), but their effectiveness for semantic analysis varies widely. Researchers at the University of Vermont conducted an evaluation of multiple dictionary-based semantic analysis methods and found that many dictionaries either contain too few words or have an insufficiently complex scoring system to really be of use. They concluded that these dictionaries are most effective when the words are scored on a continuum (i.e. on a 9-point scale where 1 is the least positive and 9 is the most positive) and there are enough words in the dictionary to represent a sufficiently large share of the words in the target text.

Of the dictionaries they tested, Reagan *et al.* (2017) found strong evidence to support the effectiveness of the labMT (language assessment by Mechanical Turk) dictionary (Dodds et al., 2015), which contains over 10,000 of the most common words in a variety of English-language corpora. Each word in the dictionary was scored 50 times by raters on Amazon's Mechanical Turk survey platform, resulting in an average score on a nine point sentiment scale for each word on the corpus. Negative words like *miserable* (score = 2.54), *killed* (1.56), and *argument* (3.14) have scores closer to the negative pole of 1, while positive words like *love* (8.42), *friendly* (7.66) and *advantage* (6.84) are closer to the positive pole of 9. The distribution of words over this spectrum was left-skewed, inspiring the title of the paper: "Human language reveals a universal positivity bias." While words skew positive in the labMT

dataset, the majority of words are closer to a middle point than to either extreme. Function words like *only* (4.92), *the* (4.98), and *to* (4.98) naturally appear towards the neutral midpoint of 5, but so do many other words: *questions, offices, boot,* and *pursued* all have a perfectly neutral rating of 5.00 in the labMT dataset.

The labMT dictionary has been applied to a variety of scientific inquiries, but of particular interest to the classical field is the plotting of the sentiment in a text over time. Researchers at the University of Vermont have used this method to identify temporal patterns of higher and lower sentiment on social media platforms (Dodds et al., 2011) and to identify the arc of sentiment through time in stories (Reagan et al., 2016). This latter approach was used in combination with a neural network to classify 1,327 stories into groups of story shapes, inspired by Kurt Vonnegut's failed thesis proposal on the topic. Vonnegut believed that most stories could be classified according to a few basic shapes which could be plotted graphically with good or ill fortune as a function of time. (Vonnegut, 1995) Because dictionary methods like labMT can be used to generate a time series of sentiment scores for a text, the emotional arcs of stories can be plotted in the vein of Vonnegut's theory.

1.1.4 This Study

In this study, we explore English-language translations of Homeric epic using tools and methods from the fields of natural language processing and the science of stories. With contextualization from broader classical scholarship, these analyses can shed light on philosophical, historical, and political approaches to translation, and reveal the suitability of computational tools to this type of inquiry. The purpose of this work is not to use these tools as a replacement to traditional modes of classical scholarship,

but rather to test these methods to see if they have anything to contribute to the fields of Homeric and translation studies.

Chapter 2

Allotaxonometry

2.1 Basics of Allotaxonometry

The Allotaxometer (Dodds et al., 2023), developed by researchers at the Vermont Complex Systems Institute, is a tunable instrument by which two systems can be compared. It takes as input a ranked list from each system with the following information about the prevalence of its component types:

1. The *type* or name of the object,
2. Its *frequency* within the system, operationalized as the count of how many times it appears *and* its rank based on that count, relative to the frequency of other elements in the system,
3. The *probability* of the object type occurring within the system, and
4. The *total number of types* of objects the system contains.

With this information, the Allotaxonometer produces a figure which contains a wealth of information on the divergence between the text as it pertains to the difference in the ranks of their component parts: this principle is called rank-turbulence divergence. The figure contains a diamond-shaped histogram on which each component is plotted according to its rank in both systems. In a comparison of two systems, System A and System B, a component which is the same rank in both systems will be positioned on the vertical center line. A component which is of higher rank in say, System A, will appear to the left of the vertical center, on what can be thought of as System A's side of the plot. This component's distance from the vertical center is proportional to how much more common it is in System A than B.

The unique orientation of this histogram has multiple benefits. First, it allows for the shape of the divergence between systems to be interpreted. If the components of two systems cluster along the vertical median, the systems are very similar, whereas two very dissimilar systems will result in the components pushing away from one another towards the outer edges of the histogram.

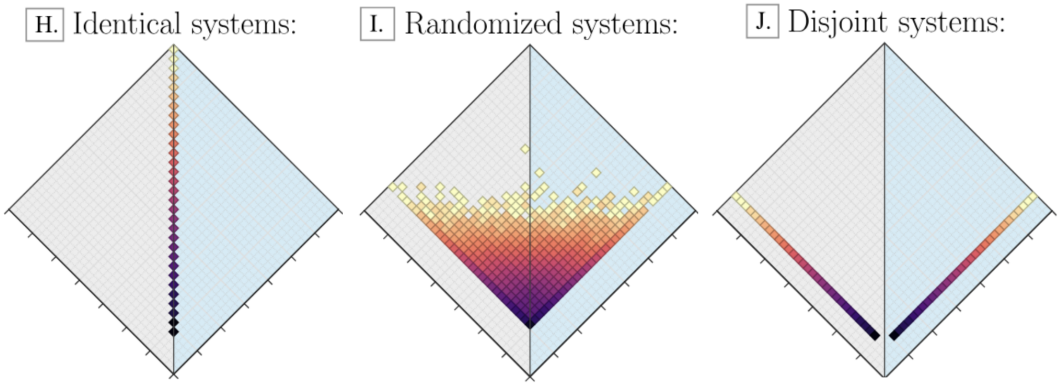

Figure 2.1: The Shapes of Three Types of Systems Comparisons (Dodds et al., 2023)

The second benefit to the orientation of this histogram is that it is easy to spot which components are more common in one system than another. Consider Figure 2.2, which compares two days of tweets on Twitter. The first day is November 9th, 2016, which was the day after the 2016 U.S. presidential election. The second is August 13th, 2017, which was the day after the white supremacist "Unite the Right" rally in Charlottesville, Virginia, USA, where a woman named Heather Heyer was murdered.

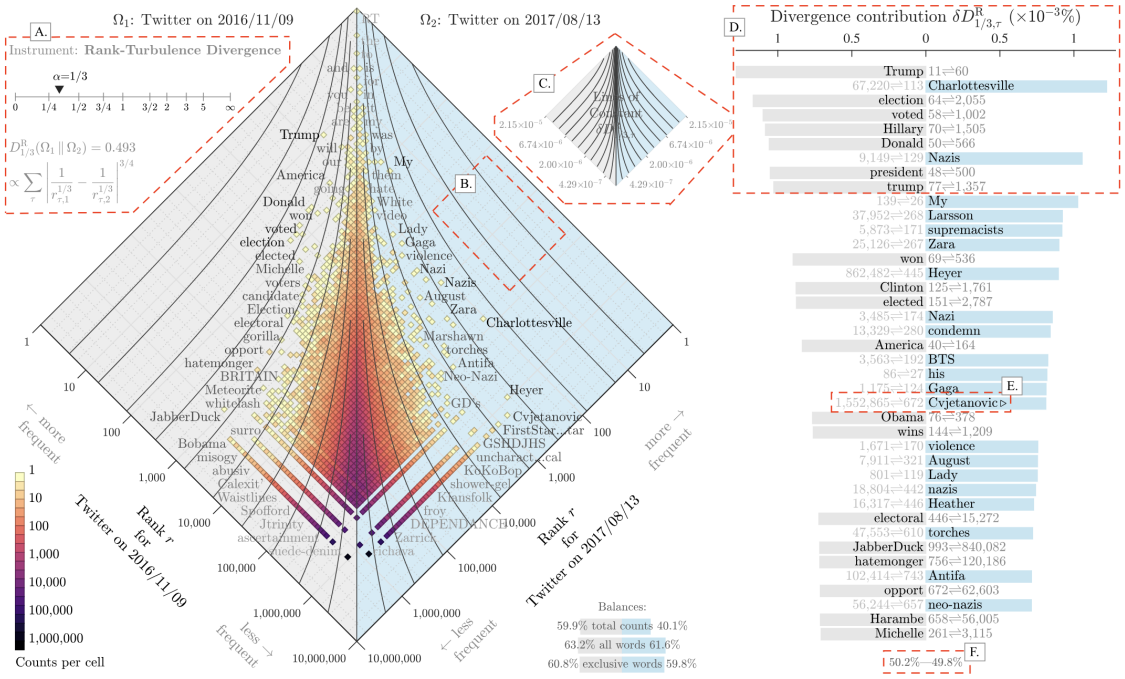

Figure 2.2: A Sample Allotaxonometry Plot (Dodds et al., 2023)

By looking at the left half of the histogram, we can see a number of words which were more common on November 9th, 2016 than August 13th, 2017: *Trump, America, Donald, won, voters, election*, etc. On the right half of the histogram, we see words that were more common on August 13th, 2017: *White, Nazi, violence, Charlottesville,*

and *Heyer*.

To the right of the histogram in Figure 2.2 is *the divergence contribution list*. It presents the 40 components which most significantly contribute to the divergence between the two systems. In other words, these are the components whose relative frequency is most different between the two systems. These two components of an allotaxonometric figure, the divergence contribution list and the histogram, will be the primary focus of analysis in this study.

Many kinds of systems can be compared using allotaxonometry. Two ecosystems could be compared based on the prevalence of the various plant and animal species they contain, or two days on Twitter based on which words were most commonly tweeted. For the purpose of this study, our systems will be individual translations of the Homeric epics.

2.2 Methods

2.2.1 Our Translations

There have been several scores of full translations of *the Iliad* and *Odyssey* into English since George Chapman's in 1615. Some of these translations are more well-known (and well-regarded) than others. We sought to select a sample of translations ranging from Homeric Epic's early days in English to present day. Our primary constraint was the availability of digital versions of the texts. In the end, we selected two older translations (Alexander Pope's and Samuel Butler's) and four from the mid-20th century to present day (Richmond Lattimore's, Robert Fagles, Stanley Lombardo's, and

Emily Wilson's). We sourced digital versions of each of the translations of *the Iliad* and *the Odyssey* from open access platforms like Project Gutenberg (gutenberg.org), Internet Archive (archive.org), and the Perseus Collection (perseus.tufts.edu). See the table below for further information on our translations.

Table 2.1: Translations in Our Sample

Translator Information			**Dates of Publication**	
Name	***Life***	***Nationality***	***Iliad***	***Odyssey***
Alexander Pope	1688–1744	British	1715	1725
Samuel Butler	1835–1902	British	1898	1900
Richmond Lattimore	1906–1984	American	1951	1965
Robert Fagles	1933–2008	American	1990	1996
Stanley Lombardo	1943–	American	1997	2000
Emily Wilson	1971–	British-American	2023	2017

2.2.2 Exploratory Phase

Basic Data Cleaning

After converting the digital files to text files (.txt format) where necessary, we cleaned the data in preparation for standardizing and analyzing. This primarily involved removing all non-Homeric text from the files, which included chapter titles and summaries, book titles, author information, page headings and footers (e.g. *BOOK IV*, *HOMER: THE ODYSSEY*, *Homer's Iliad*), page and line numbers, and any supplementary materials (e.g. translators' notes, introductions, glossaries, and appendices)

included with the translation.

Data Generation

After basic data cleaning, we built a script to turn all letters in the text into lowercase, remove punctuation, and finally to tokenize the text using the NLTK library in Python. To tokenize, in the context of natural-language processing, is to break down a text of any length into individual "tokens." In the field of natural-language processing (NLP), the branch of computer science which aims to make human language comprehensible to computers, tokenizing text is very common. When computers are working with text, they prefer for it to be organized into tokens because it makes the text easier to understand and work with. A token is often a complete word (e.g. *Homer* or *spear*) but not always. Punctuation marks, numerical digits, and even smaller word-components can all be tokens. For instance, the word *Hector's* would commonly be broken down into two tokens: *Hector* and *'s*.

We began by lowercasing all letters and removing punctuation from the text, a common first step in NLP, because tokenization may otherwise fail to recognize certain words as the same type of token based solely on their position in a sentence. Consider the example below:

Sentence #1: They hung the painting on the wall.

Token Type	They	hung	the	painting	on	wall.
Count	1	1	2	1	1	1

Sentence #2: The guests later stumbled upon the painting.

Token Type	The	guests	later	stumbled	upon	the	painting.
Count	1	1	1	1	1	1	1

The combined token type counts for the two sentences are then:

Token Type	guests	hung	later	on	painting	painting.
Count	1	1	1	1	1	1
	stumbled	The	the	They	upon	wall
	1	1	3	1	1	1

Without isolating punctuation, a tokenizer sees "painting" in sentence #1 and "painting." in sentence #2 as two different types of tokens, because the word "painting" in sentence #2 is followed by a period. Similarly, it sees the first instance of the word "the" in sentence #2 as a different type from the other three occurrences of the word, simply because that word starts the sentence and is thus capitalized. While there are some instances where these distinctions would be valuable, we are more

concerned with the fact that these words are *in the text at all* as opposed to where they fall in the grammatical structure of any given sentence. As such, all letters in the text must be lowercased and punctuation isolated and removed.

As previously stated, the Allotaxonometer requires as input not a text file with a long string of tokens but the previously mentioned pieces of information about how common those tokens are in the text in CSV format. The first few rows of a CSV file in our sample after the basic pre-processing described above will look something like this:

```
type,counts,prob,total
the,4923,0.050833281704974906,6711
and,4073,0.042056460772773266,6711
to,2561,0.026444045185139292,6711
i,1635,0.016882473204881978,6711
you,1618,0.016706936786237945,6711
```

The data above tells is that the most common token in the text in question is *the*, followed by *and* and *to*. We can also see that *the* appears 4,923 times in the text, that its probability is 0.0508, and that there are 6,711 unique tokens in the text, including this one. Once this information is extracted from the text, it can be fed into the Allotaxonometer for analysis.

2.2.3 On the Decision to Standardize the Texts

Early allotaxonometric comparisons between different translators' versions of Homeric epic revealed that orthographic differences were far and away the most significant

contributors to divergence between translations. Namely, translators differed with respect to the contemporaneity of function words, the use of contractions, and the spelling conventions of proper nouns and British versus American English. See Table 2.2 for examples.

Table 2.2: Description of Standardized Linguistic Differences in the Texts

Area of Divergence	**Definition**	**Example**
Character, place, and group names	There are many ways to anglicize names from Ancient Greek, so naming conventions and spelling differ between translations.	Older translations may use Roman names in place of the Greek: *Ulysses* instead of Odysseus, *Jupiter* for Zeus, *Minerva* for Athena, etc.
British vs. American English	Certain words are spelled differently in American versus British English.	Words like *armor* (British: armour), *favor* (British: favour), and *honor* (British: honour) appear frequently in the Iliad and Odyssey.
Antiquated language	Older translations may use more antiquated language than newer ones. This primarily affects function words with antiquated equivalents like *you*, *yourself*, *where*, and *before*, but certain antiquated spelling conventions can be applied to a much larger collection of words.	Older translations of Homeric Epic may use not only antiquated function words like *thou* (you), *hath* (has), and *art* (are), but also antiquated spellings of words such as *didst* (did) and *blacken'd* (blackened).
Contractions	Some translators rely very heavily on contractions, while others use them very seldomly or not at all.	Contractions include words like *you're* (you are), *I've* (I have), *won't* (will not), and *we're* (we are).

It is obviously true that orthographic differences between translations have significance. Alexander Pope and Samuel Butler's choices to use the Roman names for the Greek gods and some Greek heroes, for instance, is downstream of what Shannon Farley (2016) points out is the forcing of Greek narratives into Roman moulds, a decision that made the stories more palatable for more than just an Ancient Roman audience. She writes, "Greek narratives (were) told in Roman frameworks, with every Greek god given a Roman name and the Roman values at the forefront" of these texts." (p. 6) Further, the Roman versus Greek names for the gods in a translation create very different effects for a reader versed in the classics. As classicist Matthew Arnold said in his famous lectures *On Translating Homer* (1861), "The Latin names of the Greek deities raise in most cases the idea of quite distinct personages from the personages whose idea is raised by the Greek names. Hera and Juno are actually, to every scholar's imagination, two different people." (p. 71)

Antiquated diction, too, is another feature that merely be wiped away without due consideration. Alexander Pope's grandiose and antiquated language has significance not only as it betrays his philosophy on translating Homer but because it impacts a contemporary audience's understanding of what a translation of Homer should sound like, even three hundred years its after publication. Recent translations of Homer, among them Emily Wilson's and Stanley Lombardo's, are often criticized for sounding too "modern," a critique Wilson herself preempted in the Translator's Note for her *Odyssey* (2018). She writes, "Mild stylistic archaism is often accepted without question in translations of ancient texts and can be presented as if it were a mark of authenticity. But of course, the English of the nineteenth or early twentieth century is no closer to Homeric Greek than the language of today." (p. 87)

The use of language that is antiquated or more modern, then, is not an indicator simply of the age of the translation, but the philosophy of the translator. Wilson further clarifies her position on the issue as follows: "All modern translations are equally modern. The question facing translators and their readers is whether to try to disguise this fact, through stylistic tricks such as archaism and an elevated, artificially 'literary' register, or to underline it, and thereby encourage readers to be aware that the text exists in two different temporal and spacial moments at once." (p. 87)

Therefore, our decision to standardize the texts for this project must be seen not as a dismissal of the importance of these differences, but merely as an acknowledgment that these are layers of differences which *can* be standardized so more semantic and thematic differences can be unearthed. To demonstrate our commitment to this, we have conducted the standardization in a stepwise manner which allows for the contributions of each of these orthographic differences to be clearly seen. The stepwise approach therefore has the effect of peeling away layers of orthographic difference until the semantic divergence between the texts is revealed. Though only a sampling of the figures generated by these analyses will be included in the main text of this paper, all 6,048 allotaxonometric figures are available for viewing.

2.2.4 The Process of Standardizing the Texts

Mechanics of Standardizing

The method used to replace words in the standardization process utilized a Python data structure called a dictionary. Just like in a traditional, physical dictionary,

information in a Python dictionary is stored in pairs. We would use a traditional dictionary by searching for a word and reading the accompanying definition. In this way, we used the name of the word to retrieve its meaning. In a Python dictionary, information is stored in "key-value" pairs similar to this, where the "key" is like a word in the dictionary and the "value" is its definition. For coding purposes, dictionaries use these pairs to store information in a value that can be retrieved by its key.

We built bespoke Python dictionaries for each of the five phases of the standardization process. Our key-value pairs were pairings of nonstandard and standardized words. Using this dictionary in concert with Python's .replace() method, we were able to search our translations for unwanted text and either replace it with its standardized alternative (by making the key the nonstandard term and the value its standardized equivalent) or remove it from the text altogether (by making the value in the key-value pair empty space, the key would be removed and replaced with nothing).

After the replacements were made, we used the methods described in the Data Generation subsection above to generate data in allotaxonometric format. At the end of each phase of standardization, when replacements and tokenizing were completed, a new text file was produced, identical to earlier iterations but for the changes made in that phase. An accompanying Allotaxonometer-formatted CSV file would then be produced for each text file.

Phase 0: Contractions

We wanted to remove punctuation before the first phase of standardization for the reasons previously discussed. Before punctuation in the texts could be altered, though, a decision needed be made with respect to contractions (e.g. *I'll*, *we'd*, *let's*, etc.). Early

explorator analyses revealed that some translators in our sample rely heavily on contractions (Pope[1], Fagles) where others use them not at all or nearly so (Lattimore, Wilson). As a result, comparisons between, say, Fagles' and Wilson's translations yielded a divergence mostly accounted for by these contractions, most of which are function words and thus occur relatively frequently in the text. Because translators in our sample mostly elected not to use contractions, so chose to standardize instances of them to their un-contracted form.

The reason this step needed to be completed before punctuation was removed because certain contractions with their apostrophe removed could just as easily be other words. For example: if you remove the apostrophe from contracted *let's*, you're left with *lets*, which could just as easily be *let us* as the third-person singular form of the verb *to let*: e.g. "She *lets* him walk home alone after school." Other examples include *he'll* → *hell*, *I'll* → *ill*, and *we're* → *were*. Therefore, contractions must be removed first, before punctuation is manipulated and the other standardization phases are attempted.

Contractions were removed in two phases: first by hand and second using Python code we developed for this purpose. See Table 2.4 for all instances of the former and Table 2.3 for those of the latter. Some contractions needed to be removed by hand because their meaning is ambiguous out of context. For example, the contraction *she'd* can mean *she had* (e.g. "*She'd* finally done it.") or *she would* (e.g. "*She'd* need to take a look first."), and the only way to know which meaning was intended is to look at the words surrounding the contraction for context. Other contractions only ever have one meaning (*I'm* is always *I am*, *can't* is always *cannot*, etc.) and

[1]Pope in particular uses a lot of antiquated contractions: *'twere* = it were, *shan't* = shall not, *e'er* = ever, *dasn't* = dare not, etc.

can be replaced with their equivalent more easily. For the former category, we went into the text and examined each occurrence of the contractions individually, making replacements only after the intended uncontracted form was identified. For the latter, we replaced the contractions en masse using Phase 0 of the Standardizer Python code or using a simple find-and-replace method within the text file itself (as was necessary in some irregular cases like *can't*, which unlike other *n't* contractions expands to a single word, *cannot*).

Table 2.3: Contractions Standardized Using Code

Category	**Examples**
Compound contractions	I'd've, shouldn't've, you'dn't've)
Will/Shall ('ll)	he'll, we'll, they'll, that'll
Am ('m)	I'm
Not (n't)	can't, don't, isn't, hadn't, mayn't)
Are ('re)	we're, how're, they're, why're, you're)
Have ('ve)	I've, could've, ought've, so've, what've, you've
Special cases	let's (let us), dasn't (dare not), e'en (even), e'er (ever), 'twere (it were), shan't (shall not)

Table 2.4: Contractions Standardized by Hand

Contraction	**Standardized/Uncontracted form**
he'd	he had / he would
I'd	I had / I would
she'd	she had / she would
there'd	there had / there would
we'd	we had / we would
who'd	who would / who had
you'd	you had / you would
it'd	it would / it had
that'd	that would / that had
they'd	they would / they had
what'd	what did / what had
when'd	when did / when had
where'd	where did / where had
why'd	why did
he's	he has / he is
how's	how has / how is
it's	it has / it is
she's	she has / she is
this's	this has / this is
what's	what is / what does
where's	where has / where is
who's	who has / who is
why's	why has / why is
something's	something is / something has
everybody's	everybody is
everyone's	everyone is
everything's	everything is
here's	here is
nothing's	nothing is
somebody's	somebody is
someone's	someone is
so's	so is
that's	that is
there's	there is
when's	when is
which's	which is
one's	one is
can't	cannot
shan't	shall not
dasn't	dare not

There were three areas where compromises needed to be made with respect to contractions. The first concerns *'ll* contractions, which can be expanded to either *x will* or *x shall*. Unfortunately, because these words have the same meaning and share the same verb tense, there is no way to know with certainty whether the translator using the contraction meant *I'll* to mean *I will* or *I shall*. The *'ll* contraction most commonly expands to the former, however, especially in modern English, and the translations in our sample which most heavily used *'ll* contractions were published in the past 50 years. For these reasons, and further to ensure consistency by standardizing all contractions, we chose to standardize *'ll* contractions universally to *x will*. Fortunately, the translators most likely to use *shall* in place of *will* on our sample, Alexander Pope and Samuel Butler, only used *'ll* contractions a handful of times between them, so it is likely any suppression of true *shall*'s in the texts was minimal.

The second area where compromise was necessary concerned *'s*, which is sometimes a contraction and sometimes a possesive article. While an *'s* following a word like *how*, *it*, or *that* necessarily indicates the *'s* is a contraction which expands to *x is* or *x has*, an *'s* after a character or place name (e.g. *Troy's*, *Hector's*) could just as easily be a contraction (e.g. "Penelope*'s* going to bed.") as the possessive form ("Penelope*'s* bed was made.") outside of context. As a result, these cases cannot be standardized automatically. It would be possible (and even ideal) to standardize these individually by hand, but this would be a very time intensive process. Every translator in our sample used *'s* after character names in some capacity at least hundreds of times, and some used it several thousand times in a single translation. Due to time constraints, we only standardized *'s* suffixes where they were definitively contractions, as in the aforementioned cases, and did not standardize them after character names. After the

standardization of contractions in Phase 0 of data processing, remaining *'s* suffixes were removed from the text altogether.

Finally, certain contractions which are unique to the older translations in our sample (e.g. *o'er*, *'twas*, *shan't*, *'tis*, and *'d* contractions like *unnumber'd* or *plunder'd*) could have been justifiably standardized either in this phase of processing or in Phase 5, where we "modernize" function words like *thou*, *art*, and *hither*. We chose to standardize them in this phase with all other contractions.

Phase 1: Isolating Punctuation

The next phase of standardization was a more straightforward one. Isolating punctuation is a common pre-processing step in natural language processing, not only for the reasons illustrated in the opening paragraphs of the description of Phase 0, but because punctuation is often of interest in its own right. If one translator uses a plethora of exclamation points, for instance, while another uses none at all, that is a stylistic choice which influences the reader's experience of the text and is therefore of interest to our analysis. Now that contractions have been dealt with, we can safely isolate punctuation to see what differences emerge between texts, if any.

Python's NLTK tokenizer, which is used in every phase of standardization, performs some manipulation of punctuation as it breaks words into tokens. For example, after tokenizing, Line 1 of Fagles' Iliad is transformed from

"Rage—Goddess, sing the rage of Peleus' son Achilles," to

"Rage—Goddess , sing the rage of Peleus ' son Achilles ,"

Note how new space appears before the commas and the apostrophe on Peleus, while the emdash between Rage and Goddess remains. In order to fully isolate punctu-

ation, we wrote code in Python which inserts spaces between all punctuation marks in the text, with one exception, to ensure they would be read by the tokenizer as their own tokens. The aforementioned exception is hyphens, which play a uniquely important role in these texts. Epithets, descriptive adjectives or phrases which accompany and express key information about paired-nouns, are a key feature of Homeric Epic. Famous examples include *swift-footed* Achilles, *short-lived* Hector, and *gray-eyed* Athena. The hyphens within these epithets must be treated differently than other punctuation marks not merely for Homeric reasons, but for natural language processing ones: they are not separating two distinct tokens but rather creating one new, connected token, and therefore must be preserved.

In order to accomplish this, we identified the unicode characters corresponding to epithet- and compound-word-separating hyphens in the text (U+002D and U+2013) and made sure not to isolate them unless they were bordered on either side by a space. This was necessary because at least one translator in our sample used the hyphen character U+002D not just to separate epithets, as a hyphen, but also as an emdash and sentence-breaking device. Consider these exerpt from Book I of Samuel Butler's *Iliad*:

> "For nine whole days he shot his arrows among the people, but upon the tenth day Achilles called them in assembly- moved thereto by Juno, who saw the Achaeans in their death-throes and had compassion upon them." (Butler/Homer, 1898)

The first instance of the U+002D hyphen, *assembly-*, is used as an emdash, while the second, in *death-throes*, is used to make a compound word. By removing these hyphens only when they are immediately preceded or followed by empty space (either

a newline character or a space), we isolate the hyphen in the first case while preserving the compound word in the second. Outside of this exception, all punctuation marks in the texts were isolated.

Phase 2: Removing Punctuation

Punctuation marks tend to be most common tokens in any text. Because there are so many of them and they are used so frequently, they have a tendency to fill up the divergence contribution lists on any allotaxonometry plot where they're present in the text, at the expense of word tokens, which are pushed off the list as less significant contributors to divergence. While differences in punctuation usage can be interesting, we are most interested in words for this study, and consequently removed all punctuation but hyphens (see Phase 1 comments on epithets) from the text in Phase 2.

Phase 3: Name Standardizing

Early allotaxonometric analysis revealed that different spelling conventions for character, group, and place names contributed to a large amount of divergence between texts, which makes sense—names tend to be among the most frequently occurring words in stories. Naming discrepancies resulted primarily from different approaches to anglicizing Homeric Greek names. For instance, Fagles, Lombardo, and Wilson spell the name of the Trojan prince and warrior as *Hector*, whereas Lattimore spells it *Hektor*. Similarly, the Greek warrior most commonly known as *Ajax* (Pope, Butler, Fagles, Lombardo, Wilson) is called *Aias* in Lattimore's translation. More examples can be seen in Table 2.5 below. In order to peel away this orthographic layer of

divergence, we created a list of standard spellings for as many of the more than 1,000 names in *the Iliad* and *Odyssey* as we could.

Table 2.5: Some Examples of Divergence in Name Spellings

Pope	**Butler**	**Lattimore**	**Fagles**	**Lombardo**	**Wilson**
Aeneas	Aeneas	Aineias	Aeneas	Aeneas	Aeneas
Deiphobus	Deiphobus	Deïphobos	Deiphobus	Deïphobus	Deiphobus
Minerva	Minerva	Athene	Athena	Athena	Athena
Ulysses	Ulysses	Odysseus	Odysseus	Odysseus	Odysseus
Vulcan	Vulcan	Hephaistos	Hephaestus	Hephaestus	Hephaestus

Using glossaries from the translations in our sample for which they were available (the translations of Richmond Lattimore, Robert Fagles, Stanley Lombardo, and Emily Wilson) we manually matched character names using spelling similarity, glossary definitions, and other available information (e.g. line number indicating first appearance in the text). Glossaries were not available for any of the epics translated by Alexander Pope or Samuel Butler. Fortunately, the most significant divergence in naming conventions in their translations pertains to the names of the Greek gods featured in the *Iliad* and *Odyssey*. Pope and Butler chose to use the Roman names for the gods (e.g. *Jupiter* or *Jove* instead of *Zeus*, and *Neptune* instead of *Poseidon*. We standardized all names of deities to their anglicized Greek version (e.g. *Zeus*, *Poseidon*) as this was the most common choice across translations in our sample.

In choosing the standard spelling for a name, we defaulted to Emily Wilson's chosen spelling for two reasons. First, her glossary was the most complete. While no translator in our sample included every name mentioned in the combined four glossaries, Wilson's came the closest, and choosing hers as the standard meant fewer standard spellings needed to be arbitrarily decided outside of this standard. Second,

Wilson's spelling conventions were more common in our sample than those of the only other translator with a similarly complete glossary, Richmond Lattimore.

One area where special consideration had to be taken was in the standardization of group names. There are spelling differences in group names just as with other names in Homeric Epic: *Danaan* is sometimes spelled *Danaän*, *Achaean* may be spelled *Achaian*, and *Dardanians* are sometimes referred to simply as *Dardans*. But beyond this, translators may use different names may be used to refer to the same group of people.

In the Trojan War, the non-Trojan side on which Achilles, Agamemnon, and Ajax fought may be called the *Greeks*, the *Achaeans*, the *Argives*, or the *Danaans* in an English translation. The first of these terms is the only one not to appear in the original text. This is because, as Emily Wilson states in her introduction to the Odyssey:

> "Greece," as a unified entity, is an invention of the classical age; in the sixth and especially the fifth centuries BCE, Greek-speaking people began to define themselves as Hellenes, in contrast to the "barbarian" (meaning "non-Greek-speaking") peoples of other civilizations, such as the Persians and the Egyptians. But in Homer, as Thucydides points out, there is no single term for all Greek people. Those who sail to attack Troy from places that would later be defined as "Greek" are categorized by names for smaller ethnic tribes, or as the followers of individual leaders: the Danaeans, the Achaeans, the Myrmidons, and so on.

And yet, it is not always the case that these "names for smaller ethnic tribes, or ... the followers of individual leaders" are used to actually refer to that specific

subgroup. *Danaans*, for instance, is used in the Iliad only to refer to the Greeks as a whole. Similarly, Greeks not from Argos or Achaea are referred to in the poem as *Argives* and *Achaeans*.

Despite this, it is not the case that they are entirely interchangeable. As classicist Androniki Oikonomaki (2018) points out, "although it is not always easy to discern their different use and function within the text ... it seems that Achaioi, Danaoi, and Argeioi are not identical terms and the poet sometimes intentionally ascribes to each term different qualities in a different context in order to produce meaning."

This is perhaps especially clear in the Odyssey, where "the more general unmarked term Achaioi denotes not only the warriors of the Iliadic past, but also the population of specific Greek regions of the Odyssean present, as Ithaka, Argos, Iason Argos, Zakynthos, and Crete. In the 'peaceful' world of Odyssey, where the war is a memory, Danaoi and Argeioi signify the besiegers of Troy, since they are never acting characters, but rather the people of a distant past." (Oikonomaki, 2018)

Presumably these terms would have meant something to an ancient audience. To a modern English-speaking one, however, they are generally unfamiliar and the distinctions between them opaque. Resultantly, some translators choose to translate instances of *Achaioi*, *Danaoi*, and *Argeioi* all as *Greeks*.

The Trojan side is mostly free of this dilemma. The Trojan side as a collective entity is generally just referred to as "the Trojans and their allies", though Oikonomaki notes that "sometimes Trojans and Dardanians appear to be synonymous terms." (Oikonomaki, 2018) We found little discrepancy in the frequency of use of *Dardans*/*Dardanians* between translators, but we did find occurrences of *Ilian* and *Phrygian* in Alexander Pope's Iliad where other translators used *Trojan*.

Ultimately, where we found evidence of translators in our sample using a specific sub-term to refer to an entire group, we standardized all occurrences of that sub-type to the larger group name. So, *Argive*, *Danaan*, and *Greek* all became *Achaean*, and *Phrygian* and *Ilian* became *Trojan*. Those interested in how often different names were used by the translators in our sample, consult Table 2.6. Some translators appear to have stuck closely to mentions of *Achaioi*, *Danaoi*, and *Argeioi* in the original text, which Oikonomaki reports to exist at a ratio of 65:16:19 (Butler's Iliad was 68:15:17; Lattimore's was 67:16:18) while others were quite off.

Table 2.6: Group Mentions in the Iliad by Translator

Term	**Pope**	**Butl.**	**Latt.**	**Fagl.**	**Lomb.**	**Wils.**
Greek(s)/Grecian	407	0	0	0	553	845
Achaean(s)/Achaian(s)	16	574	635	276	104	1
Danaan(s)/DanaĂ¤n(s)	0	125	148	10	24	0
Argive(s)	26	145	169	395	38	14
Myrmidon(s)	17	39	40	41	38	42
Trojan(s)	303	622	627	637	553	666
Ilian(s)	4	0	0	0	0	0
Phrygian(s)	31	2	3	3	1	3
Lycian(s)/Lykian(s)	51	46	50	34	45	48
Dardanian(s)/Dardan(s)	33	34	38	35	26	33
Thracian(s)	15	17	18	18	17	20

Phase 4: British to American English

British and American English differ most noticeably with respect to vocabulary and spelling. The British refer to fried potatoes as *chips*, Americans call them *fries*. The

cargo space in the back of a car is called the *boot* in British English and the *trunk* in American English. The world's most popular sport is called *soccer* in the U.S. and *football* in Britain (and in most other places.) Though these vocabulary differences loom large in public perception of the differences between these dialects, spelling differences are also highly relevant when working with the written word.

The words spelled differently in British versus American English largely fall into a set of defined categories. For example, the British suffix *-our* (e.g. *armour* and *valour*) is generally *-or* in American English (*armor*, *valor*), the British *-ise* (e.g. *practise*, *characterise*) is usually *-ice* or *-ize* in American English (*practice*, *characterize*), and words ending in *-re* in British English (e.g. *centre*, *calibre*) may end in *-er* in American English (*center*, *caliber*).

Our sample of translations was split between British and American English spelling, which led to significant orthographic contributions to divergence between texts. For the fourth phase of standardization, we standardized the spelling of these words. American English was chosen as the "standard" for two reasons. First, the the majority of translations in our sample were written in American English. Second, the LabMT dataset, which was used to calculate the sentiment of our translations, is biased towards American English spelling. Therefore, standardizing British English spellings to American English ensured our British English translations would not be disadvantaged in our hedonometric analysis.

We identified the opportunity for British to American English standardization in the exploratory stages of allotaxonometric analysis. The words *armour*/*armor* and *honour*/*honor* came up repeatedly as significant contributors to divergence in comparisons between translations written in British and American English. We searched

the texts in our sample for instances of the *-our/-or* ending and found other examples: *valour/valor*, *favour/favor*, *clamour/clamor*, *vigour/vigor*. The quantity of examples we discovered prompted us to approach the identification of spelling differences in a more efficient way. We consulted various lists of words which differ in spelling between the two dialects and further searched the texts for certain areas of divergence (e.g. *-ise* and *-er* endings as discussed above) to assemble a list of words to replace, which we did entirely using Python code. See Table 2.7 for a list of all the standardized British to American English words.

Table 2.7: Spelling Differences Between British and American English

Category	**British English**	**American English (Standard)**
ou/u	colour, valour, armour, mould	color, valour, armour, mold
-ise/-yse	practise, recognise, analyse	practice, recognize, analyze
ll/l	travelling, councillor, woolly	traveling, councilor, wooly
l/ll	skilful, fulfil, appal	skillful, fulfill, appall
-ence/-ense	defence, offence	defense, offence
-re/-er	sceptre, centre, meagre	scepter, center, meager
-e	gramme, annexe, tonne	gram, annex, ton
oe/e	oesophagus, manoeuvre	esophagus, maneuver
Misc.	grey, plough, towards, sceptical	gray, plow, toward, skeptical

While this method may not have produced a completely exhaustive list of words

in the text with different spellings, this was unlikely to cause significant issues for our analysis. The aim of this method was to catch the most frequently occurring problem words, because words which occur frequently in one text and do not appear at all in another contribute the most to divergence between the texts as defined by our methods. Words we missed because they occurred very infrequently in the texts are likely not going to significantly contribute to divergence between texts. The purpose of this exercise was not to create perfectly standardized texts as much as it was to create texts that were reasonably standardized to make them more suitable for allotaxonometric analysis.

Phase 5: Modernizing

The final step in our standardization process was a modernization of the texts. In early allotaxonometric analysis, we noticed the older translators in our sample used antiquated forms of certain words: *thou* instead of *you*, *thine* instead of *your/yours*. A closer examination of these texts revealed a number of other antiquated versions of common verbs and function words were in use. Their relative frequency in the texts resulted in their dominating the divergence contribution lists, so we choose to standardize them.

In order to identify the words we needed to replace, we conducted a survey of the Pope and Butler translations to find antiquated words with modern equivalents. In the end, we compiled a list of roughly 70 terms which we believe constituted the most commonly used antiquated words in the text. A number of these words needed to be standardized by hand, either because not all instances of that word should be replaced or because the antiquated word had two possible modern equivalents, and

the appropriate standardized replacement could only be ascertained by looking at the context in which the original was used. In the latter case, take the word *thine*, which can be used to mean *your* or *yours*. Similarly, *whence* can mean *where* or *from where*, so it's necessary to standardize these words manually to ensure their replacement is truly their equivalent in function.

For an example of the former case, where not all instances of a word should be replaced, consider the two pairs of lines below, from Book I of Alexander Pope's Iliad:

> Great as thou art, and like a god in fight,
> Think not to rob me of a soldier's right.
>
> A maid, unmatch'd in manners as in face,
> Skill'd in each art, and crown'd with every grace;

Alexander Pope commonly used *art* as an antiquated conjugation of the verb *to be*. He also, and about as frequently, used *art* as a noun, to refer to a person's work or skill in a particular area. Because of this, we had to standardize each case of the former *art* by hand so as not to accidentally and incorrectly modify the *art*s in the latter group. The bulk of antiquated words we identified for replacement could be replaced automatically using code. See Table 2.8 below for the list.

A final note on what we did not standardize. There were some words in our translations that were simply more common in older texts as opposed to exclusive to them (e.g. *thus*, *moreover*, *lest*, *amidst*). We elected not to standardize these words. Additionally, due to the evolution of the English language over the past several centuries, there were certain grammatical practices more common in older than newer translations. For instance, when constructing superlative adjectives, Alexander

Pope commonly uses the *-est* suffix (e.g. *choicest*, *noblest*, *purest*, *tenderest*) where modern/contemporary translators would instead use *most* before the word (e.g. *most noble*, *most tender*). Similarly, Pope was more likely to hyphenate some words than other, more contemporary translators (e.g. *eye-balls*, *horse-hair*, *first-born*). We elected not to modify these words for the reasons stated above.

Table 2.8: Antiquated Words in Our Texts and Their Modern Equivalents

Punctuation-dependent (Modernized in Phase 0)	
Antiquated	***Modern***
-'d suffix (e.g. heap'd, employ'd, enjoy'd)	-ed suffix (heaped, employed, enjoyed)
e'er	ever
howe'er	however
o'er	over
ne'er	never
tis	it is
twas	it was
whate'er	whatever
whatsoe'er	whatsoever
whene'er	whenever
where'er	wherever

Function Words	
Antiquated	***Modern***
aforetime	formerly
betwitxt	between
ere	before
gainst	against
hither	to here
hitherto	until now
o	oh
oft	often
thee	you
thence	from there
thine	your/yours
thither	to there
thitherto	until then
thou	you
thy	your
thyself	yourself
to-morrow	tomorrow
to-night	tonight
twixt	between
whereon	on which
whither	to where
whitherto	to where
ye	you

-st/-est Suffixes	
Antiquated	***Modern***
arrivest	arrives
bathest	bathes
blest	blessed
camest	came
canst	can
comest	come
couldst	could
darest	dare
didst	did
dost	do
drovest	drove
durst	dared
dwellest	dwells
fliest	fly
givest	give
hadst	had
hast	have
liest	lie
mayest	may
mayst	may
mightst	might
owest	owe
protectest	protect
rulest	rule
seest	see
shakest	shake
shouldst	should
whilst	while
wouldst	would

Miscellaneous	
Antiquated	***Modern***
art	are
hath	has
scaped	escaped
scatheless	unscathed
stedfast	steadfast
unburthen	unburden
wast	were

2.2.5 Results of Standardization

In Figure 2.5, we can see the results of two comparisons between Alexander Pope and Emily Wilson's *Iliads*: the top plot was generated after Phase 2 of standardization and the bottom plot after the fifth and final phase. At this stage, words have been tokenized and punctuation and contractions have been removed. We can see at once three types of orthographic differences. The first is differences in character and group name spellings, driven in large part in this instance by Pope's decision to use the Roman names for the Greek pantheon. Pope uses *Jove* in place of the Greek name *Zeus*—these two tokens appear in the third and fourth places of the divergence contribution list, at ranks 35 and 39 in their respective texts. Further down the list, we can see the name *Athena* appears only in Wilson's text. This is because Pope uses her Roman name, *Minerva*. Not all names are inconsistent. A look at the histogram shows the tokens *greeks* and *trojans* appear in both texts, though their position to the right of center indicates they are more common in Wilson's *Iliad* than Pope's.

The second type of orthographic difference visible in the top plot is antiquated versus modern English. The divergence contribution list indicates that the antiquated words *thy*, *thou*, *thee*, and *ye* appear only in Pope's translation. The modern *you*, while not exclusive to Wilson's text, is significantly more common in her translation (rank 7) than Pope's (rank 172). The final type of orthographic difference makes a small appearance near the bottom of the divergence contribution list: the word *armor* is ranked 141 in Wilson's *Iliad*'s tokens at this stage, but apparently does not occur at all in Pope's translation. This is true, in a sense; Pope doesn't use the word *armor*, but he does use the British equivalent (*armour*) several dozen times.

In the bottom plot, the texts have been fully standardized and the aforementioned problem words have fallen out of the plot altogether. In their absence, more thematic words (e.g. *fighters, lance, race, son, sire, bold,* and *foe* emerge. We can also observe that, with an exception of the shortening of the bottom-most lines branching out on either side of the vertical center as words only used in one text were standardized, the shape of the divergence is largely preserved from the second to fifth phase of standardization.

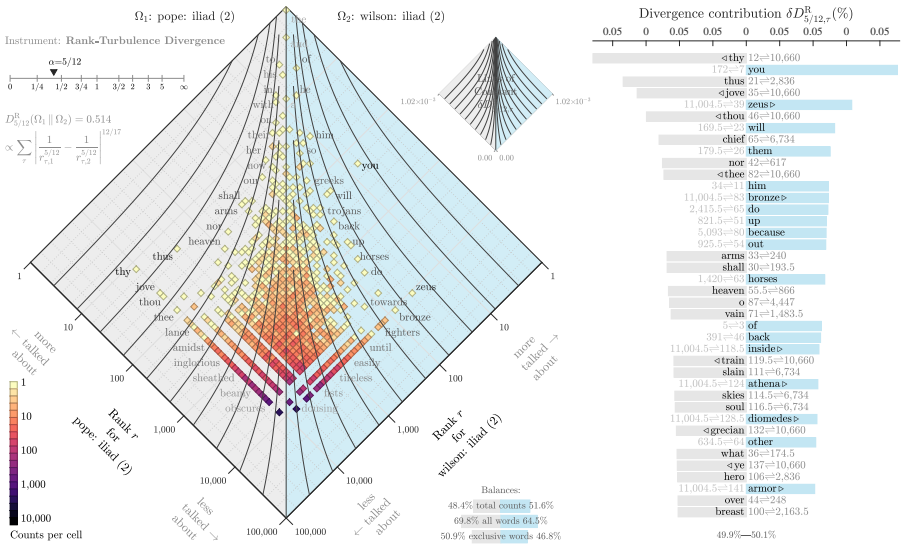

Figure 2.3: Comparison of Pope's Iliad and Wilson's Iliad at Phase 2

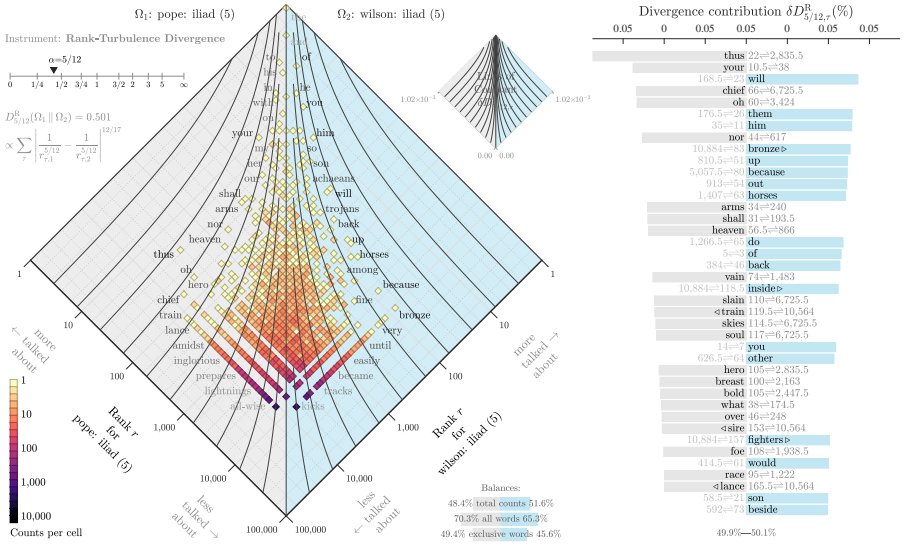

Figure 2.4: Comparison of Pope's Iliad and Wilson's Iliad at Phase 5

Figure 2.5: Comparisons of Pope's Iliad (pub. 1715) and Wilson's Iliad (pub. 2023) at Phases 2 and 5 of Standardization

In Figure 2.8 we can see spelling discrepancies of a different kind. Lattimore is known for anglicizing Greek names more directly than many of his fellow translators,

who often latinize the Greek names as an intermediate step before anglicizing them. Lattimore preferred to go straight from the Greek to English, so he writes *Menelaos* instead of *Menelaus*, *Achaians* instead of *Achaeans*, *Achilleus* instead of *Achilles*, and so on. [2]

The vast majority of the nearly 1,000 names in Lattimore's glossary were spelled differently than our standard. Because character names generally appear very frequently compared to other tokens in stories, a non-standardized comparison between Lattimore and any other translator results in something like Figure 2.6, where the majority of words in the divergence contribution list are just names with different spellings. Once standardization is complete (Figure 2.7), thematic words like *great-hearted*, *blameless*, and *henchman* appear. As with Figure 2.5, standardization shortens the bottom prongs of tokens as the list of words present only in one text is dramatically reduced in length.

[2] A small note on the bolded phrase towards the center of the divergence contribution list (the jumbled phrase containing "triangleleft" and "danaäns"): so-called "special characters" like ä cause errors in MATLAB's rendering of allotaxonometry plots. That item should simply read *danaäns* with a triangle to the left of the image to indicate it only appears in Lattimore's *Iliad*.

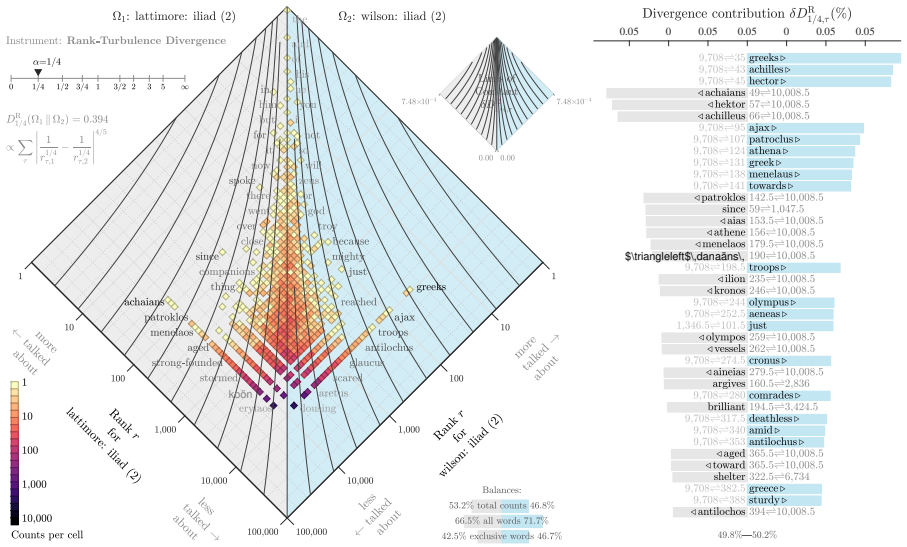

Figure 2.6: Comparison of Lattimore's Iliad and Wilson's Iliad at Phase 2

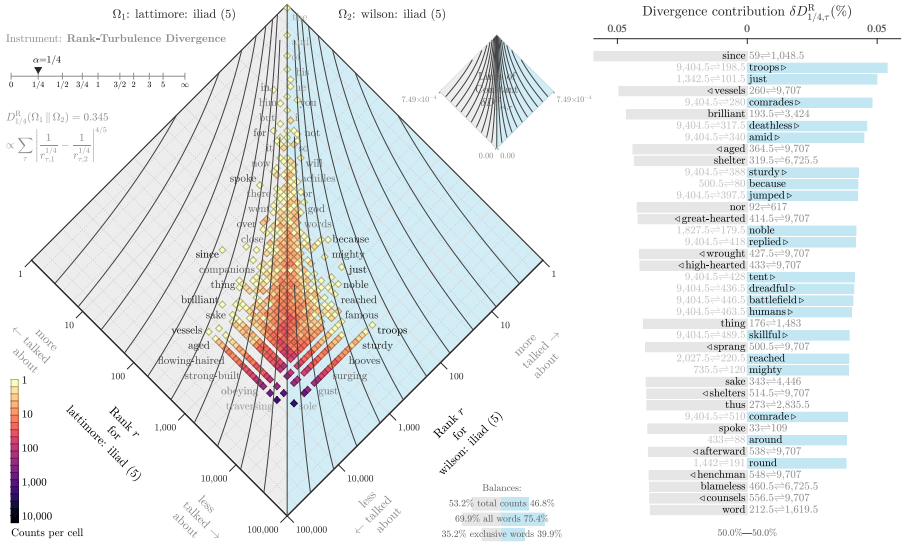

Figure 2.7: Comparison of Lattimore's Iliad and Wilson's Iliad at Phase 5

Figure 2.8: Comparisons of Lattimore's Iliad (pub. 1951) and Wilson's Iliad (pub. 2023) at Phases 2 and 5 of Standardization

In order to demonstrate what changes in a text after each phase of standardization, we have run comparisons of each translation against itself after each subsequent round

of standardization. In Figure 2.9, we can see that the isolation of punctuation breaks tokens like —*the* into two parts: — and *the*. In Figure 2.10, punctuation is removed, and the bottom-most and outermost line on the left side of the plot shows that all of those tokens no longer appear in Pope's *Odyssey* after Phase 2. The punctuation marks were the most common tokens after Phase 1, which is typical for English-language texts. Once they are removed, we can see that the most common word tokens in the text shift up to occupy higher rankings in the divergence contribution list. See, for example, how *the* moves from rank 2 to rank 1, *of* from 7 to 4, and so on.

In Figure 2.11, we see the effects of Phase 3, where names were standardized. *Ulysses* is replaced by *Odysseus*, *Jove* by *Zeus*, and so on. In Figure 2.12, British English is standardized to American English, so *labours* becomes *labors*, *honoured* becomes *honored*, etc. The fact that the bulk of the words in the plot are matched on that median line demonstrate that this phase of standardization was very targeted and did not effect most words in the text. The fifth and last phase of standardization modernized antiquated language. We can see this phase of standardization is mostly working properly, but also that at least a few errors have occurred. First, *yough* is apparently tied for rank 490.5 in Pope's Odyssey (Phase 5). The word does not appear at all in Phase 4, and in fact shouldn't be there in Phase 5 either—it seems as though the code has changed some instances of *though* to *yough* because it's misidentifying the letters *thou* as the *word thou* and substituting them with the letters *you*. It's not clear what in our code caused this to happen, as not all instances of *though* in the text were changed, though it did happen over 30 times. Additionally, the token *vain-for* at the bottom of the plot is clearly meant to contain an emdash and not a

hyphen. This is likely due to an error in the text file we sourced which has led to that token not being broken up as it should have been in Phase 2.

In the final figure of this set (2.14), we compare Pope's Odyssey at the first and final phases of standardization. The plot reveals a sampling of all the different types of tokens we standardized during this process.

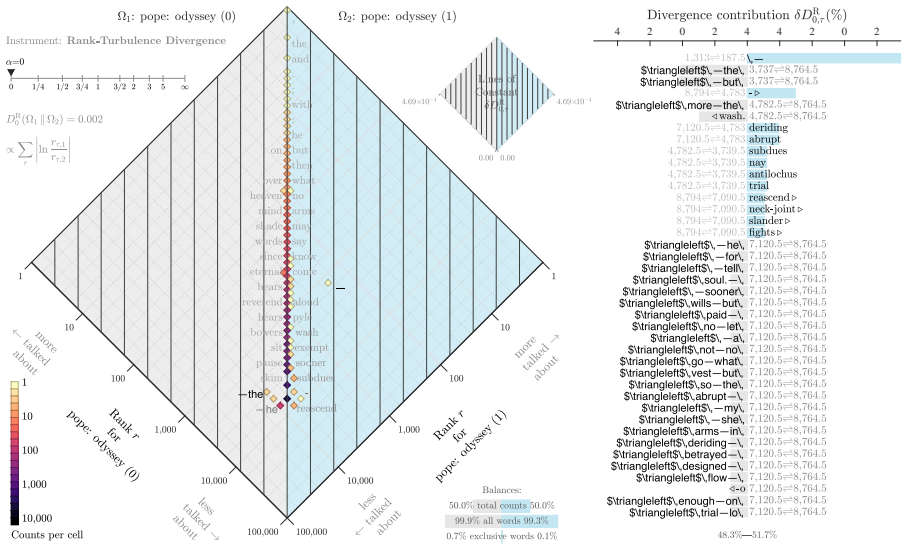

Figure 2.9: Comparison of Alexander Pope's Odyssey at Phases 0 and 1 of Standardization

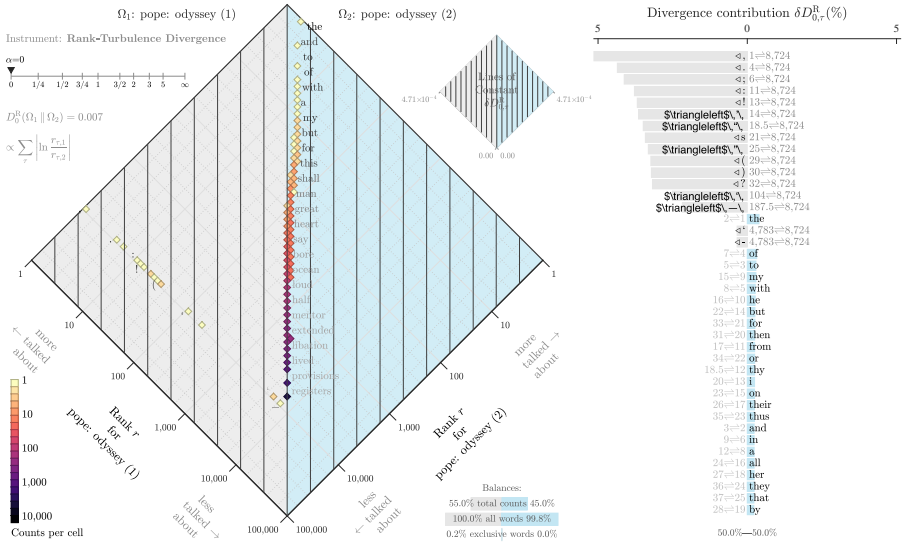

Figure 2.10: Comparison of Alexander Pope's Odyssey at Phases 1 and 2 of Standardization

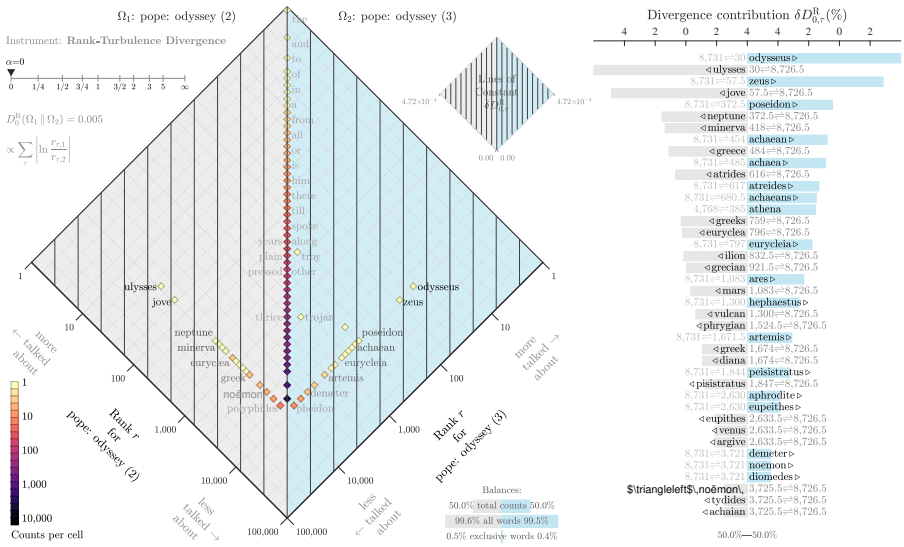

Figure 2.11: Comparison of Alexander Pope's Odyssey at Phases 2 and 3 of Standardization

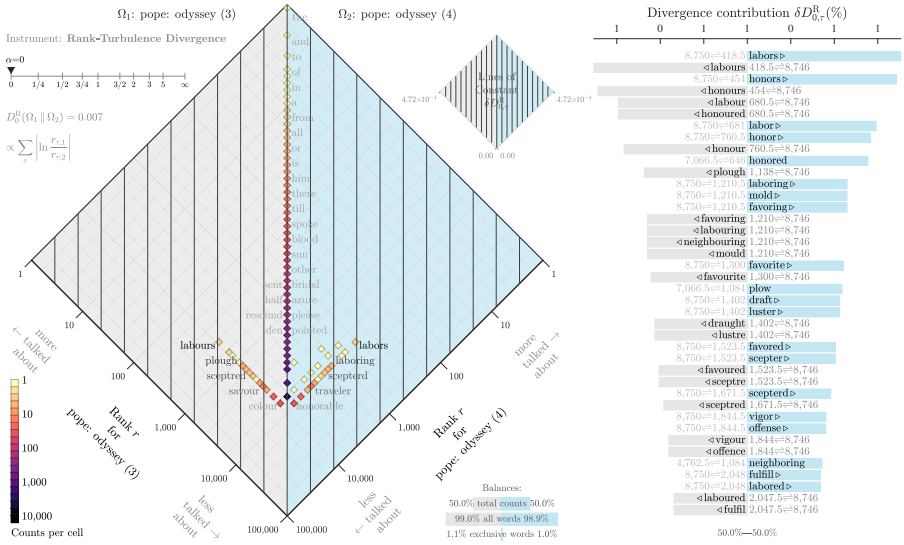

Figure 2.12: Comparison of Alexander Pope's Odyssey at Phases 3 and 4 of Standardization

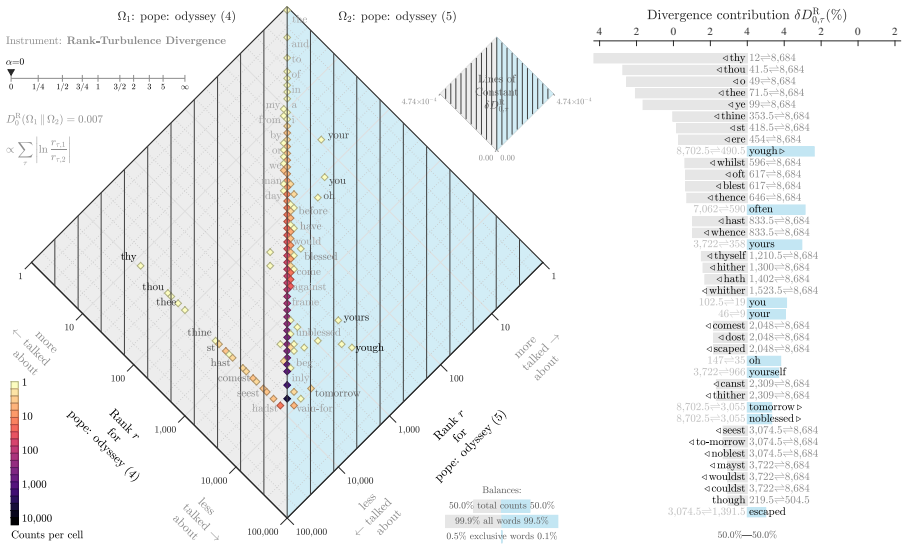

Figure 2.13: Comparison of Alexander Pope's Odyssey at Phases 4 and 5 of Standardization

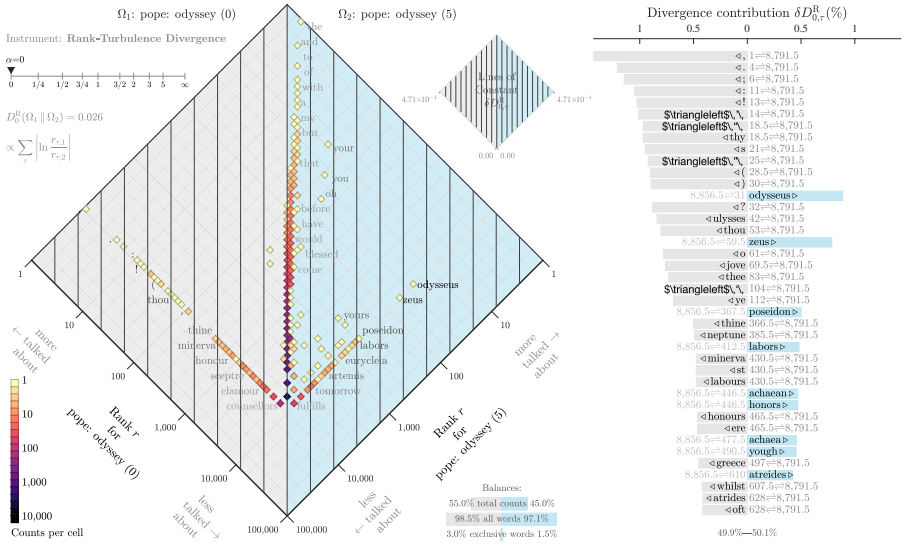

Figure 2.14: Comparison of Alexander Pope's Odyssey at the First and Last Phases of Standardization

2.3 Results

The products of our standardization data generation, then, were 72 text files and 72 CSV files. This is a set of text and accompanying CSV files for each phase of standardization for each translation in our sample: 6 phases $*$ 6 translators $*$ 2 translations per translator (*the Iliad* and *the Odyssey*) = 72. We used the CSV files to run the following analyses using the Allotaxonometer:

1. A comparison of two translators' versions of *the Iliad* or *the Odyssey* at the same level of standardization, such as:
   - Translator A's *Iliad* (Phase 0) x Translator B's *Iliad* (Phase 0)
   - Translator A's *Iliad* (Phase 1) x Translator B's *Iliad* (Phase 1)
   - ...
   - Translator A's *Iliad* (Phase 5) x Translator B's *Iliad* (Phase 5)
2. A comparison of a translator's *Iliad* or *Odyssey* at a given phase of standardization with the phase that immediately follows it, such as:
   - Translator A's *Odyssey* (Phase 0) x Translator A's *Odyssey* (Phase 1)
   - Translator A's *Odyssey* (Phase 1) x Translator A's *Odyssey* (Phase 2)
   - ...
   - Translator A's *Odyssey* (Phase 4) x Translator A's *Odyssey* (Phase 5)
3. A comparison of a translator's *Iliad* or *Odyssey* at the first and last phases of standardization.

- Translator A's *Iliad* (Phase 0) x Translator A's *Iliad* (Phase 5)
- Translator B's *Iliad* (Phase 0) x Translator B's *Iliad* (Phase 5)
- ...
- Translator F's *Iliad* (Phase 0) x Translator F's *Iliad* (Phase 5)

We did not run any analyses which compared the texts of two translators at different levels of standardization (e.g. Translator A's *Iliad* (Phase 0) x Translator B's *Iliad* (Phase 3)) or analyses which compared *the Iliad* and *the Odyssey* directly to one another (e.g. Translator A's *Iliad* x Translator A's *Odyssey*). The former would not have produced particularly interesting results and the latter was outside the scope of this study.

We were most interested in the comparisons of translations after standardization had been completed. While the results of comparisons at Phases 0–4 were used to validate the standardization process and are available for viewing in their entirety, in this paper we will mostly discuss findings from Phase 5 allotaxonometric comparisons.

2.3.1 Evolving Divergence Values Through Standardization

We will begin with two tables which demonstrate the results of our standardization process. In addition to the histogram and accompanying divergence contribution list, each allotaxonometric comparison results in a single divergence value for those two systems at a given value of the tunable parameter α. This value is a single number that tells us how divergent two systems are based on the principles of rank-turbulence divergence, with values closer to zero indicating a less divergent system.

We collected the divergence values for all comparisons of our *Iliad*s and *Odyssey*s at three levels of standardization: Phase 0, where contractions had been removed, Phase 2, where contractions and punctuation had been removed, and Phase 5, where contractions and punctuation had been removed and names, British-English words, and antiquated words had been standardized. In order to meaningfully compare the divergence values of these comparisons, we needed to choose a single value for α that was a reasonably good fit for the majority of comparisons. In this case, that value was $\alpha = \frac{1}{3}$.

Tables 2.9 and 2.10 show the divergence values for comparisons at Phase 0, Phase 2, and Phase 5 in the leftmost tables, and the differences between [Phase 2 and Phase 0] and [Phase 5 and Phase 2] in the two rightmost tables. These tables show us two things: first, whether and to what degree texts are becoming more standardized/similar to each other throughout the standardization process, and second, which texts in our sample are the most and least divergent according to this metric.

Regarding the former point, we can observe that standardization consistently minimizes the divergence between texts. We see bigger jumps for certain translators than others, but these differences are consistent with our expectations. For example, comparisons with Robert Fagles' translations of the Iliad experience a bigger jump than others towards convergence after Phases 1 and 2, which standardized the many contractions Fagles uses in his translations. Similarly, comparisons with Richmond Lattimore's translations show a greater movement towards divergence because the standardization of character, group, and place name spelling had a great impact on his translations.

On the latter point, Alexander Pope's translations are consistently the most di-

vergent from other texts in our sample. While divergence is minimized across the board as standardization occurs, it remains true that his translations at each phase are still the most divergent in our sample.

Table 2.9: Divergence Values for Comparisons of the Iliad at Three Levels of Standardization

Iliad (Phase 0)						
	AP	SB	RL	RF	SL	EW
AP		0.479	0.532	0.522	0.508	0.511
SB	0.479		0.422	0.447	0.394	0.411
RL	0.532	0.422		0.440	0.403	0.422
RF	0.522	0.447	0.440		0.391	0.406
SL	0.508	0.394	0.403	0.391		0.353
EW	0.511	0.411	0.422	0.406	0.353	

Iliad (Phase 2)						
	AP	SB	RL	RF	SL	EW
AP		0.472	0.531	0.496	0.503	0.499
SB	0.472		0.414	0.413	0.378	0.388
RL	0.531	0.414		0.418	0.399	0.409
RF	0.496	0.413	0.418		0.356	0.370
SL	0.503	0.378	0.399	0.356		0.335
EW	0.499	0.388	0.409	0.370	0.335	

Iliad ($D_{Phase2} - D_{Phase0}$)						
	AP	SB	RL	RF	SL	EW
AP		-0.007	-0.001	-0.026	-0.005	-0.012
SB	-0.007		-0.008	-0.034	-0.016	-0.023
RL	-0.001	-0.008		-0.022	-0.003	-0.013
RF	-0.026	-0.034	-0.022		-0.035	-0.036
SL	-0.005	-0.016	-0.004	-0.035		-0.018
EW	-0.012	-0.023	-0.013	-0.036	-0.018	

Iliad (Phase 5)						
	AP	SB	RL	RF	SL	EW
AP		0.467	0.495	0.483	0.491	0.485
SB	0.467		0.358	0.401	0.363	0.372
RL	0.495	0.358		0.375	0.359	0.361
RF	0.483	0.401	0.375		0.350	0.362
SL	0.491	0.363	0.359	0.350		0.328
EW	0.485	0.372	0.361	0.362	0.328	

Iliad ($D_{Phase5} - D_{Phase2}$)						
	AP	SB	RL	RF	SL	EW
AP		-0.005	-0.036	-0.013	-0.012	-0.014
SB	-0.005		-0.056	-0.012	-0.015	-0.016
RL	-0.036	-0.056		-0.043	-0.040	-0.048
RF	-0.013	-0.012	-0.043		-0.006	-0.008
SL	-0.012	-0.015	-0.040	-0.006		-0.007
EW	-0.014	-0.016	-0.048	-0.008	-0.007	

Table 2.10: Divergence Values for Comparisons of the Odyssey at Three Levels of Standardization

Odyssey (Phase 0)						
	AP	SB	RL	RF	SL	EW
AP		0.520	0.534	0.519	0.538	0.534
SB	0.520		0.399	0.428	0.376	0.393
RL	0.534	0.399		0.408	0.366	0.389
RF	0.519	0.428	0.408		0.364	0.380
SL	0.538	0.376	0.366	0.364		0.333
EW	0.534	0.393	0.389	0.380	0.333	

Odyssey (Phase 2)						
	AP	SB	RL	RF	SL	EW
AP		0.512	0.534	0.517	0.528	0.528
SB	0.512		0.394	0.413	0.362	0.378
RL	0.534	0.394		0.402	0.354	0.382
RF	0.517	0.413	0.402		0.345	0.367
SL	0.528	0.362	0.354	0.345		0.314
EW	0.528	0.378	0.382	0.367	0.314	

Odyssey ($D_{Phase2} - D_{Phase0}$)						
	AP	SB	RL	RF	SL	EW
AP		-0.008	-0.000	-0.002	-0.010	-0.006
SB	-0.008		-0.005	-0.015	-0.014	-0.015
RL	-0.000	-0.005		-0.006	-0.012	-0.007
RF	-0.002	-0.015	-0.006		-0.019	-0.013
SL	-0.010	-0.014	-0.012	-0.019		-0.019
EW	-0.006	-0.015	-0.007	-0.013	-0.019	

Odyssey (Phase 5)						
	AP	SB	RL	RF	SL	EW
AP		0.509	0.514	0.507	0.518	0.519
SB	0.509		0.362	0.403	0.350	0.366
RL	0.514	0.362		0.381	0.331	0.358
RF	0.507	0.403	0.381		0.343	0.364
SL	0.518	0.350	0.331	0.343		0.312
EW	0.519	0.366	0.358	0.364	0.312	

Odyssey ($D_{Phase5} - D_{Phase2}$)						
	AP	SB	RL	RF	SL	EW
AP		-0.003	-0.020	-0.010	-0.010	-0.009
SB	-0.003		-0.032	-0.010	-0.012	-0.012
RL	-0.020	-0.032		-0.021	-0.023	-0.024
RF	-0.010	-0.010	-0.021		-0.002	-0.003
SL	-0.010	-0.012	-0.023	-0.002		-0.002
EW	-0.009	-0.012	-0.024	-0.003	-0.002	

2.3.2 Rhyme Scheme in Pope

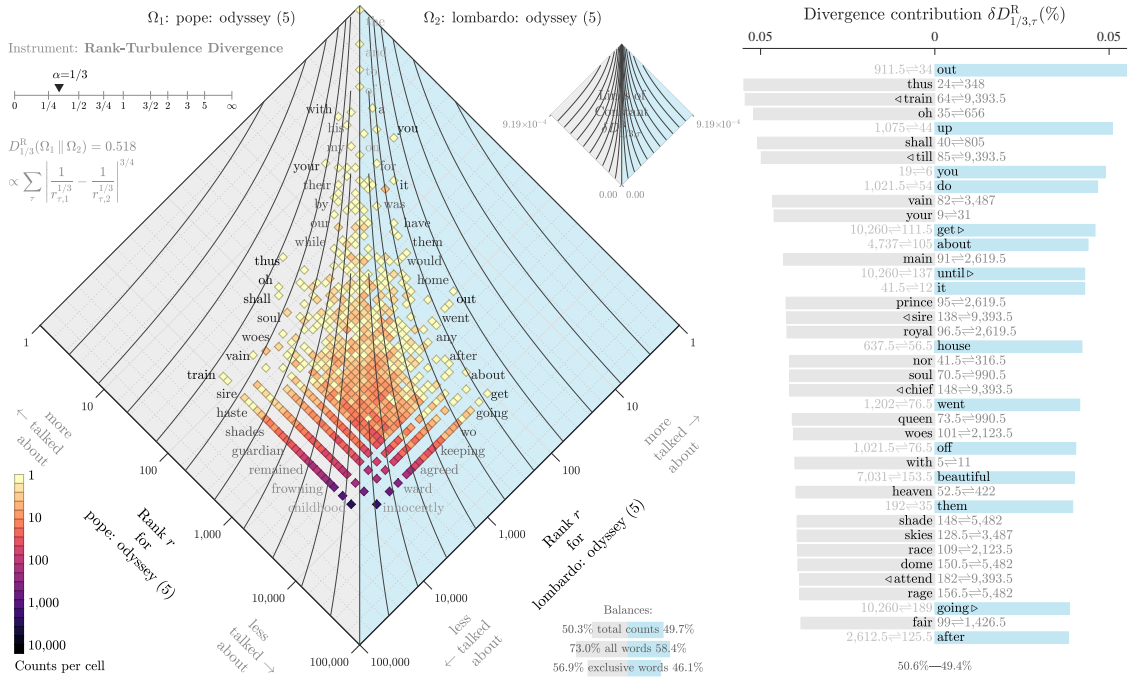

Figure 2.15: Pope's Odyssey (Phase 5) vs. Lombardo's Odyssey (Phase 5) at $\alpha = \frac{1}{3}$

In Figure 2.15, we have a comparison between *the Odysseys* of Alexander Pope (1725) and Stanley Lombardo (2000) after standardization. In the divergence contribution list for this plot, we can see sets of words which are more common in Alexander Pope's Odyssey, but also have something *in common* with one another. Take *train*, *vain*, and *main*: these words are particular favorites of Pope in large part because they rhyme, and the rhyming couplets Pope used as the structure of his translation necessitate that each line rhyme with its immediate successor. In fact, many of the words on Pope's side of the divergence contribution list share this easily-rhymeable quality. *fair*, *race*, *skies*, *woes*, and *shade* are all frequently used at the end of Pope's

lines as part of a rhymed-pair.

2.3.3 Matching Sets: Word Pairs

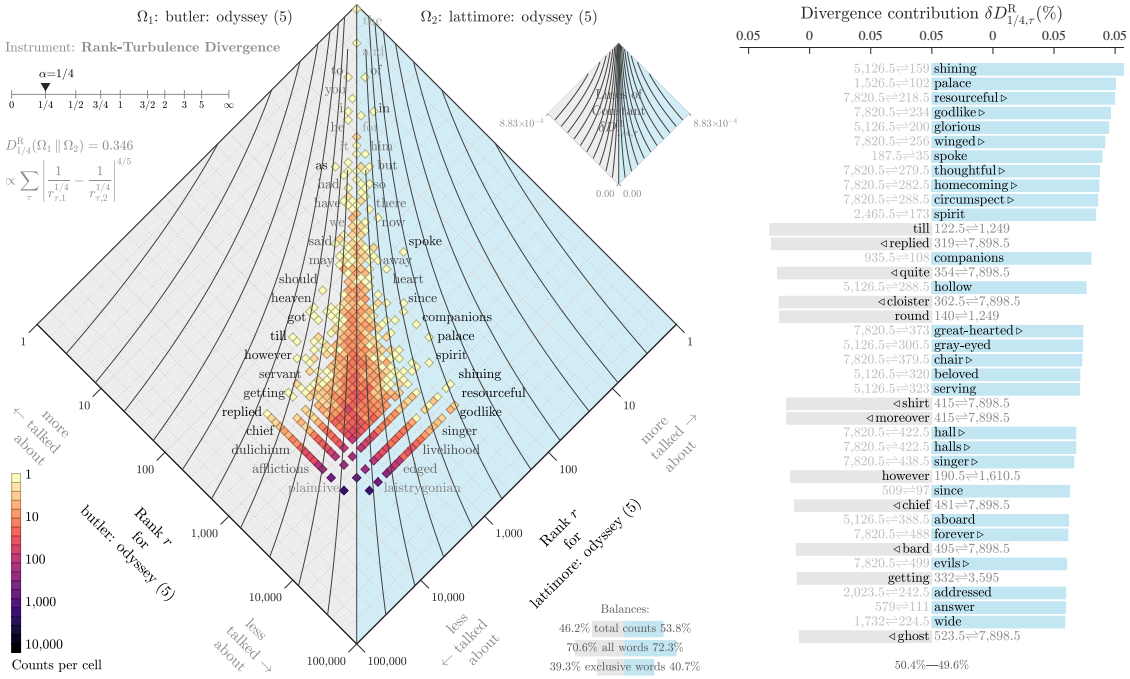

Figure 2.16: Butler's Odyssey (Phase 5) vs. Lattimore's Odyssey (Phase 5) at $\alpha = \frac{1}{4}$

Figure 2.16 illustrates a different potential use for allotaxonometric analysis in the context of translations: identifying word pairs from the same root word in the untranslated text. Our analyses frequently revealed words appearing on either side of the divergence contribution list which have the same basic meaning; consulting the texts revealed these words were often used in place of one another in the two translations being compared. For instance, Butler uses *cloister* where Lattimore uses *hall*, *halls*, or *palace*. Butler uses *bard* where Lattimore uses *singer*. Other word pairs include Butler's *ghost* and Lattimore's *spirit*; Butler's *replied* and Lattimore's *spoke*.

In Figure 2.17, we can see that Lattimore's *shelter* corresponds to Wilson's *tent*, his *spoke* to her *replied* or *said*, and his *immortals* to her *deathless* (e.g. *the deathless gods*). Similarly, the word from which Lattimore derives the epithets *great-hearted* or *high-hearted* is sometimes translated as *noble* in Wilson's *Iliad*.

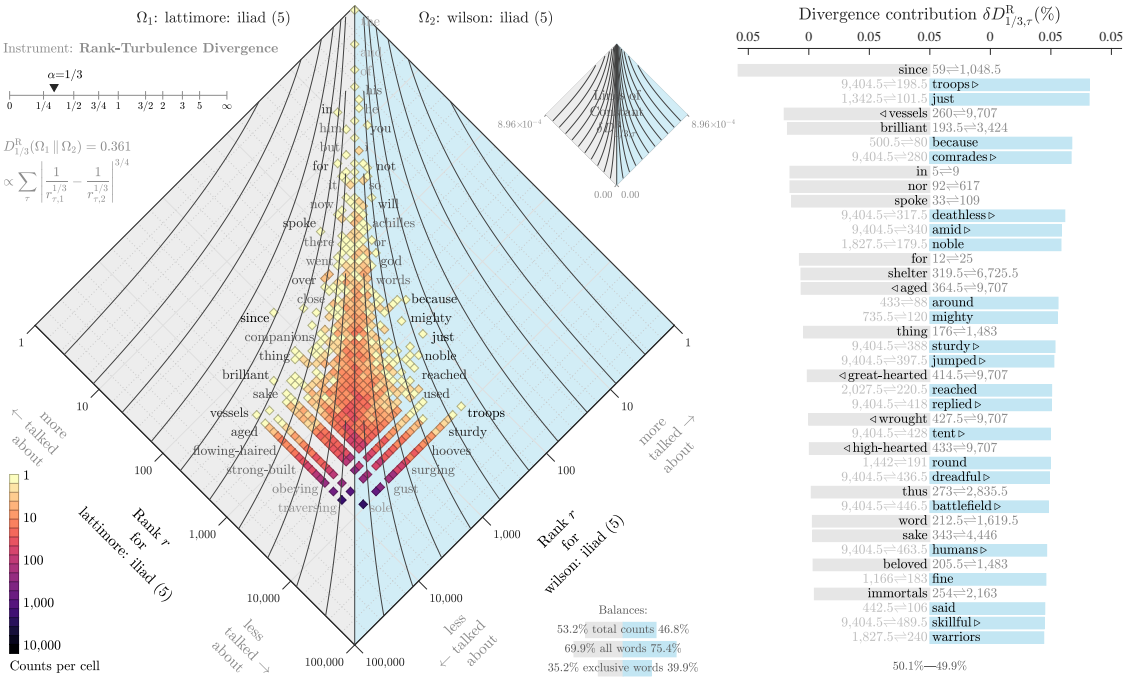

Figure 2.17: Lattimore's Iliad (Phase 5) vs. Wilson's Iliad (Phase 5) at $\alpha = \frac{1}{3}$

2.3.4 Epithets

On the matter of epithets, a look at that comparison of Lattimore's and Wilson's translations of *the Iliad* (this time at a lower value of α to better capture these less commonly used words) shows differences in how the two translators approach this most essential component of Homeric epic. The epithet Lattimore translates as *great-hearted*, for instance, is often translated as *noble* in Wilson's *Iliad*. This

example turns out to be evidence of a broader trend: Lattimore uses significantly more compound adjectives like *flowing-haired* and *strong-greaved* when translating epithets than Wilson does. His translation of *the Iliad* is full of these compound-adjectives: there are 535 unique hyphenated token types, the vast majority of them epithets, with those tokens used a total of 1405 times in the text. Wilson's translation, by contrast, only contains 312 unique hyphenated types, occurring 842 times in the text.

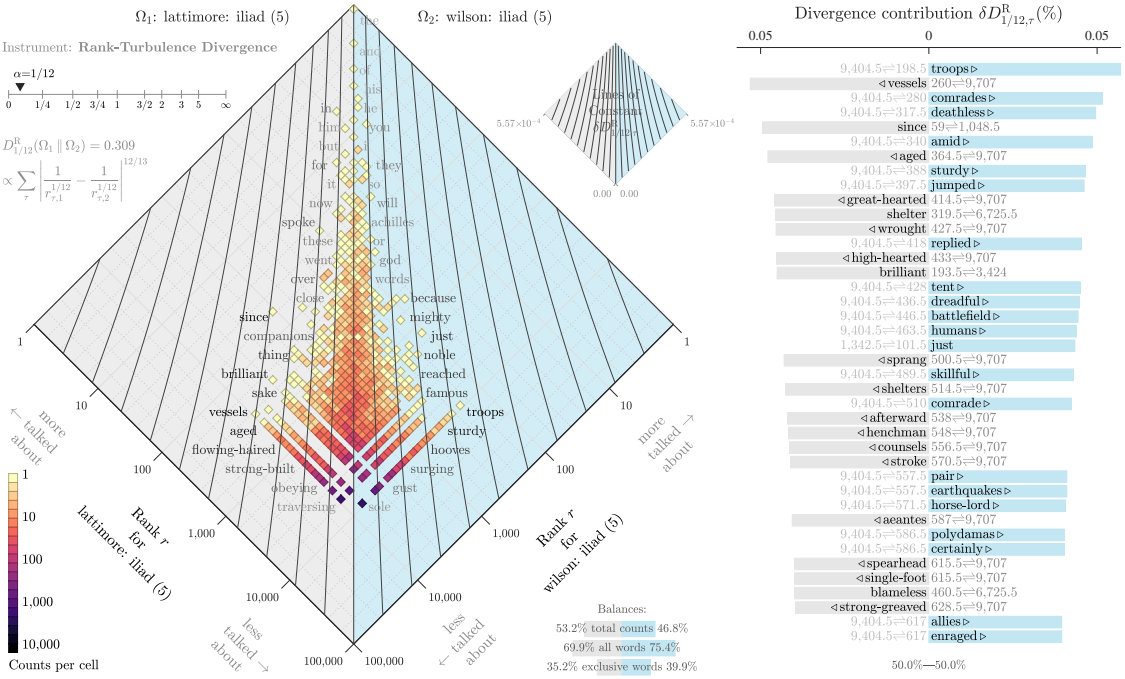

Figure 2.18: Lattimore's Iliad (Phase 5) vs. Wilson's Iliad (Phase 5) at $\alpha = \frac{1}{12}$

2.3.5 Grandiosity of Diction

Some comparions seemed to indicate differences in the grandiosity of the diction of translations in our sample. Consider Figure 2.19, where the tokens that are more

common in Wilson's translation of *the Odyssey* seem to be of a far more direct and even simple quality than those in Fagles translation. Certain terms are of a more elevated style in Fagles' translation: where Wilson uses *cows*, Fagles is more likely to use *oxen* or *cattle*. Wilson uses *ship* and *boat* more often than the more elevated *craft*, which is used much more often in Fagles' translation. Fagles uses the rather ornate *herald* in places where other translators might use servant or, in Wilson's case, *slave*. Fagles' side of the divergence contribution list in this plot also shows an evocative selection of words that are absent in Wilson's translation: *gallant, lustrous, glistening, anguish*.

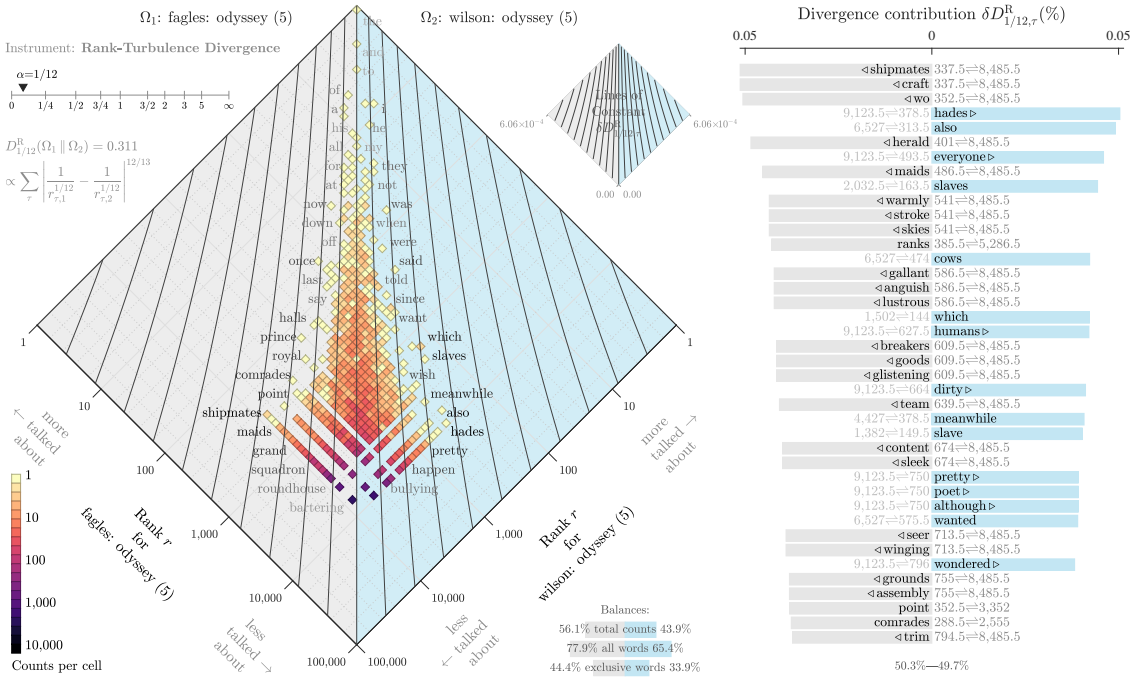

Figure 2.19: Fagles' Odyssey (Phase 5) vs. Wilson's Odyssey (Phase 5) at $\alpha = \frac{1}{12}$

This pattern, whereby an allotaxonometric comparison of two texts seemed to indicate more florid language in one translation than another, was something we

observed in multiple comparisons. But the observable nature of that effect seemed somewhat fluid—in other words, the fact that Fagles' diction looked more grandiose when compared to Wilson's didn't mean it would continue to when compared to other translations. See Figure 2.20, for instance, where, when compared to Alexander Pope's *Odyssey*, Fagles' diction seems more functional than ornate. Likewise, Wilson's more direct diction looks less basic when compared to Stanley Lombardo's, as in 2.21, with whom she is perhaps better matched in this particular respect.

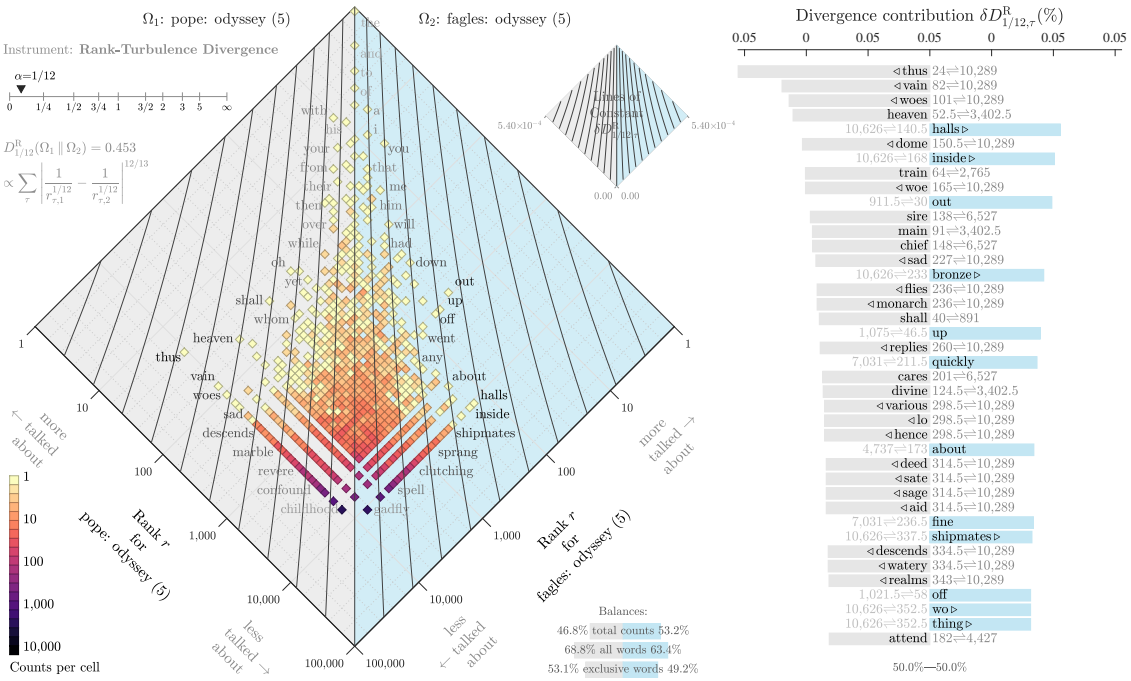

Figure 2.20: Pope's Odyssey (Phase 5) vs. Fagles' Odyssey (Phase 5) at $\alpha = \frac{1}{12}$

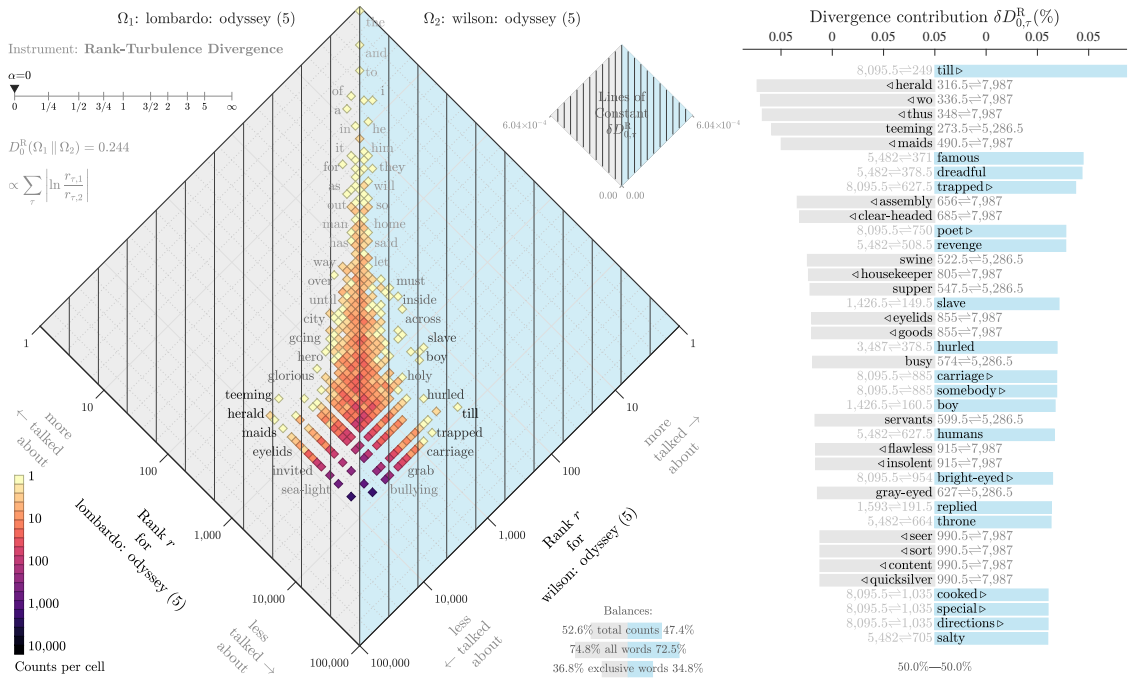

Figure 2.21: Lombardo's Odyssey (Phase 5) vs. Wilson's Odyssey (Phase 5) at $\alpha = 0$

2.3.6 The Visibility of Slavery in *the Odyssey*

In several of the previous examples, we may have noticed something else: any comparison with Emily Wilson's Odyssey tends to produce the words *slave* or *slaves* as major contributors to divergence between her translation and others'. In Figure 2.21, for instance, the rank for slave is 149.5 in Wilson's translation of the Odyssey and rank 1,426.5 in Lombardo's. Some word pairs can be identified here: the words *maids*, *housekeeper*, *servants*, and *herald* are words Wilson common translates so as to indicate the enslaved status of the characters. This phenomenon is also observed in Figures 2.22 and 2.23, where Butler's and Fagles' translations of the Odyssey are compared to Wilson's (the latter this time at $\alpha = \frac{1}{3}$ to reveal different words on

the divergence contribution list). The words *boy* and *girls* are also more common in Wilson's translation per these plots; these terms are often used to refer to enslaved characters in her translation.

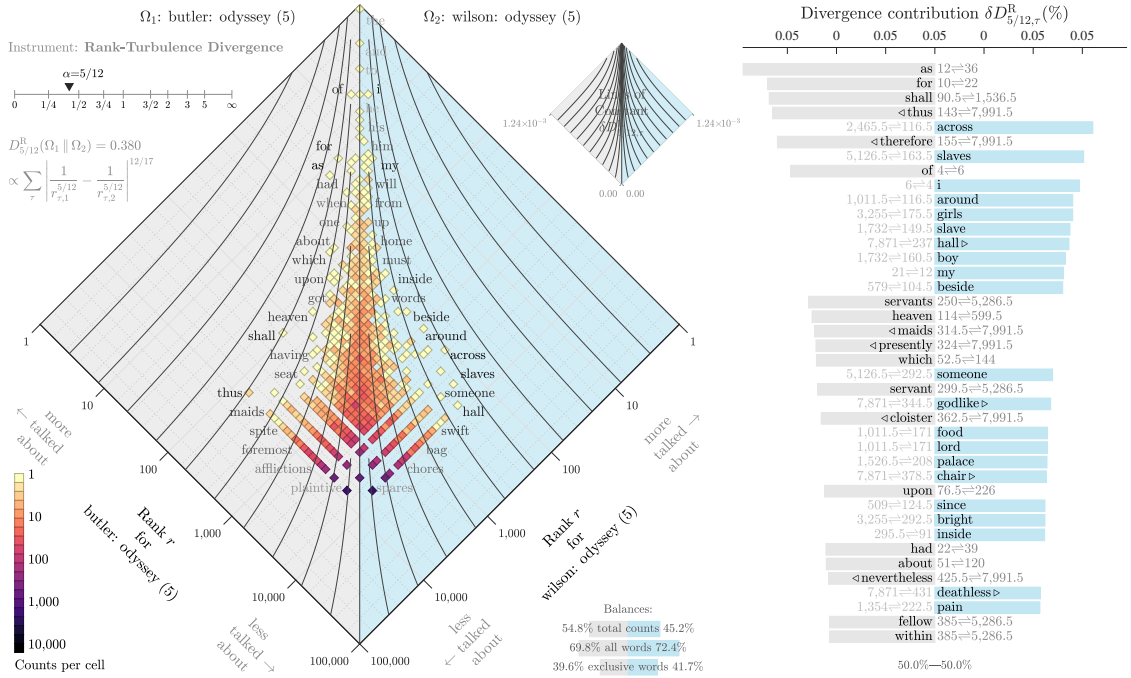

Figure 2.22: Butler's Odyssey (Phase 5) vs. Wilson's Odyssey (Phase 5) at $\alpha = \frac{5}{12}$

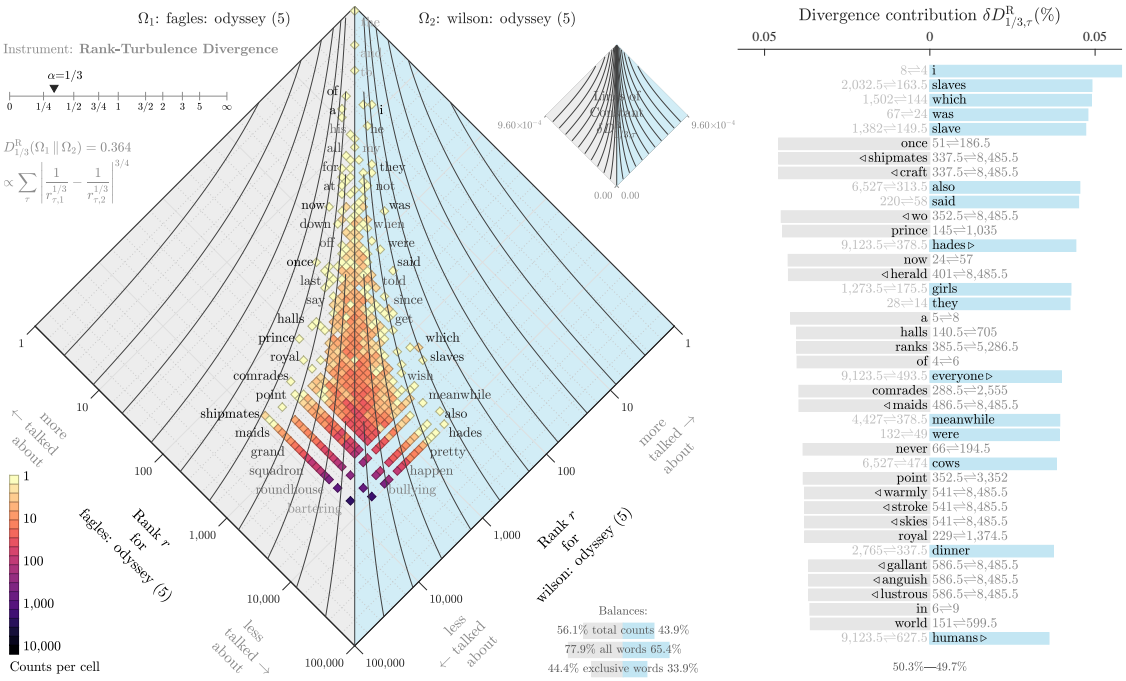

Figure 2.23: Fagles' Odyssey (Phase 5) vs. Wilson's Odyssey (Phase 5) at $\alpha = \frac{1}{3}$

2.3.7 Royalty and Titles

Some translators in our sample seem more inclined than others to use language that evokes the British monarchy and perhaps the older medieval era of Europe: see prince, herald, and royal in Fagles' translation of *the Odyssey* in Figure 2.23. Even the word *gallant* calls to mind the heroes of Arthurian legend. In Figure 2.24, Butler uses the words *knight* and *squire*, and in this company, *steeds* and *peer* also read as distinctly English.

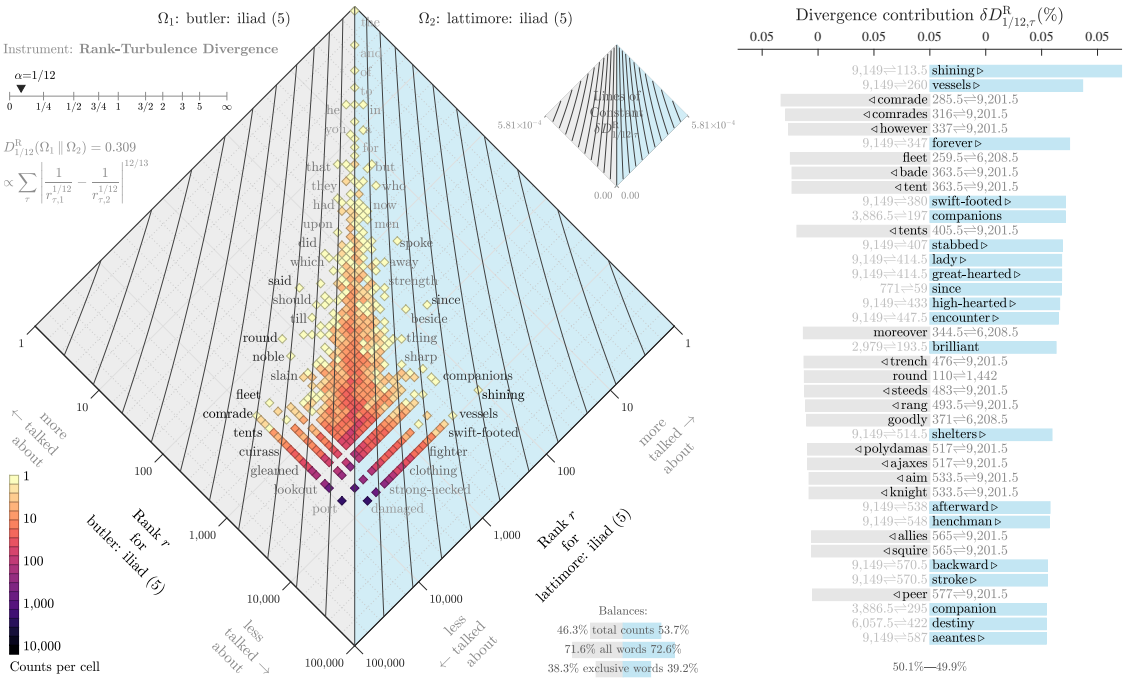

Figure 2.24: Butler' Iliad (Phase 5) vs. Lattimore's Iliad (Phase 5) at $\alpha = \frac{1}{12}$

2.3.8 Transition Words

Comparisons of Butler's translations of Homeric epic tended to reveal he used certain transition words more commonly than his fellow translators. In Figure 2.25, we see that words like *thus, therefore, however, moreover, further,* and *presently* all appear much more often in Butler's than Fagles' translation of *the Odyssey.*

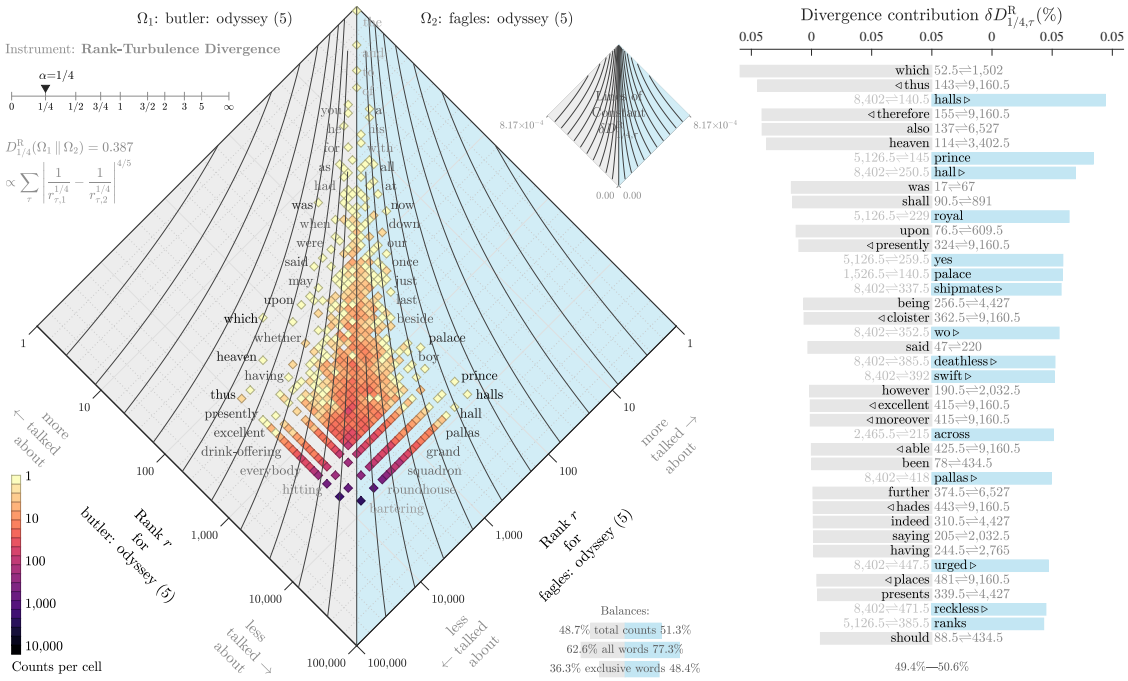

Figure 2.25: Butler's Odyssey (Phase 5) vs. Fagles' Odyssey (Phase 5) at $\alpha = \frac{1}{4}$

2.3.9 The Shapes of Our Comparisons

The allotaxonometric histogram can be visually evaluated to assess the degree of divergence between two texts. As explained in Section 2.1, The Basics of Allotaxonometry, more similar systems converge towards the vertical center of the histogram, and more divergent systems spread out and away from each other towards the outer-bottom edges of the plot. As a result, comparing multiple allotaxonometry plots can indicate at a glance which sets systems are more divergent from one another. Consider Figure 2.28, where two allotaxonometric plots are stacked to allow for visual comparison. The upper plot compares translations of *the Odyssey* by Alexander Pope and Emily Wilson, the most temporally distant translators in our sample. The

bottom plot compares Emily Wilson's translation of the text with that of her closest chronological peer, Stanley Lombardo. The difference in the shapes of the two plots is immediately apparent; the Lombardo/Wilson comparison produces a tight allotaxonometric shape with a small flare towards the bottom of the plot, while the data on the Pope/Wilson are much more spread out, and don't converge towards the top of the plot nearly as well as does the Lombardo/Wilson plot.

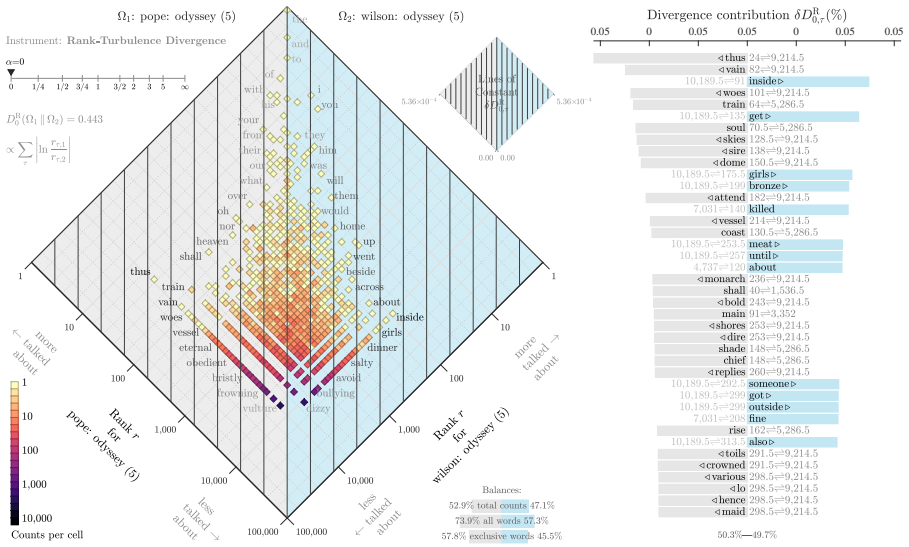

Figure 2.26: Comparison of Pope's and Wilson's Odysseys (Phase 5)

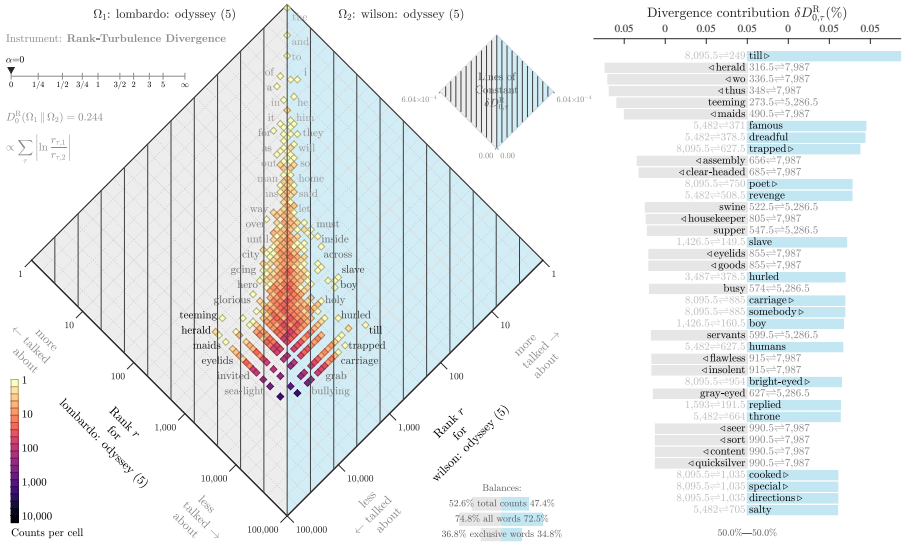

Figure 2.27: Comparison of Lombardo's and Wilson's Odysseys (Phase 5)

Figure 2.28: The Shapes of Two Comparisons of the Odyssey: Pope's vs. Wilson's (Top, 292 Years Apart) and Lombardo's vs. Wilson's (Bottom, 17 Years Apart)

2.4 Discussion

The results of our exploratory analysis indicate that scientific, distant-reading methods like allotaxonometry can identify meaningful differences between translations of Homeric epic. With the joining of these methods with close-reading approaches, these results take on particular significance.

Take the findings on Alexander Pope's rhyme scheme, for instance. Because Pope's translation has perhaps been the most well-known and enduring translation of the past three centuries, we have a wealth of classical scholarship on this and other aspects of his translations. Matthew Arnold, in his 1861 lectures, *On Translating Homer*, spoke thus of his opposition to rhyming Homer: "rhyme inevitably tends to pair lines which in the original are independent, and thus the movement of the poem is changed ... the moment the word chance strikes our ear, we are irresistibly carried back to advance and to the whole previous line, which, according to Homer's own feeling, we ought to have left behind us entirely, and to be moving farther and farther away from." (p. 16) Arnold believed there were four essential characteristics of Homeric epic which must be present in a successful translation, and the first of these was *rapidity of movement*. By forcing lines to rhyme with their predecessors, a weight is essentially attached to the ankles of every other line in the text as it's tethered backwards to its paired line. Arnold felt this resulted in a betrayal of this essential Homeric quality.

Arnold felt Pope's choice of rhyming couplets was ill-thought; others think it also betrays Pope's political orientation towards the task of translating Homer. As Shannon Farley puts it in her piece of transformative rewriting:

> "The appeal of the heroic couplet derived from the fact that that it was

> a traditional form of English poetry, dating back to Chaucer. The heroic couplet is highly ordered, and marks the poet who uses it as one who is not lured by Miltonic blank verse and, by metonymy, Milton's politics. Read aloud, it gives the feeling of soldiers on the march. Creating order out of chaos is a major theme in Pope's era, not just in form but also in content." (Farley, 2016, p. 9)

So then, the rhyme scheme of Pope must be understood to possess not only a poetic or literary quality but also a political one, as it is evidence of the conservative philosophy which shaped the form of Pope's translations, and thereby also influenced non-Greek readers' perception of the Homeric epics for hundreds of years.

Recall the line from this piece of Farley's we cited earlier: "Greek narratives (were) told in Roman frameworks, with every Greek god given a Roman name and the Roman values at the forefront" of these texts. The Roman mold is not the only one the Homeric epics have been forced into—just as the epics can be Romanized, they can be Anglicized. Our findings on diction regarding royalty and titles further relate to this. The choice to use terms like *knight* and *squire* puts the sometimes very strange Ancient Greek system into terms more immediately comprehensible to modern audience, and thereby fulfills what translation studies scholar Lawrence Venuti refers to as the platonic ideal of a translation in Anglo-American society: the text is "immediately recognizable and intelligble, 'familiarised,' domesticated, (and) not 'disconcerting[ly]' foreign." (Venuti, 2008, p. 5)

But in truth, these texts *are* disconcertingly foreign, at least at certain points if not consistently throughout. In using language which calls to mind the systems and stories of the translator and audience's own culture (e.g. the British system of

monarchy, the Arthurian legends), the translator may lead the reader to conclude that the world of Ancient Greece was in no way meaningfully different from our own—a position which simultaneously tames the undesirable features of the foreign society and elevates one's own to the idealized classical level in one fell swoop.

This idealization of the texts and of antiquity more broadly is manifest in the insistence by some translators to use grand, florid, and even antiquated language in their translations. The results of our analyses indicated that Alexander Pope's diction was particularly grand, and indeed, Pope is often held up as a particular example of this phenomenon by classical scholars. Some even credit Pope as the originator of this approach to Homer, as Wilson appears to: "The notion that Homeric epic must be rendered in grand, ornate, rhetorically elevated English has been with us since the time of Alexander Pope." (Wilson/Homer, 2018, p. 83) Pope in particular has been criticized for this by many scholars: Matthew Arnold (1861) famously said Pope "is artificial both in his style and in his words" (p. 65) and that, in Pope's translations, "Homer's thought has passed through a literary and rhetorical crucible, and come out highly intellectualised; come out in a form which strongly impresses us, indeed, but which no longer impresses us in the same way as when it was uttered by Homer." (p. 18)

Classicists are not united in this matter, of course. This approach to Homer is praised by many, evident not only in the continues relevance of Pope's translations but in the fact that he is so often emulated; in Bernard Knox's introduction to Robert Fagles' Iliad, he calls Pope's translation of the Iliad "the finest ever made." (Knox/Fagles/Homer, 1990) A translation which is more ornate, which feels more elevated and literary by the standards of its audience, is often seen as inherently superior. Indeed,

when Wilson's translations are criticized, it is often because readers feel her diction is affectedly plain, a conclusion that seems to be supported by our allotaxonometric findings. This is something Wilson owns, though. In her *Odyssey*translator's note, she writes that she, in fact, "*aimed* for a certain level of simplicity, often using fairly ordinary, straightforward, and readable English. In using language that is largely simple, my goal is not to make Homer sounds 'primitive,' but to mark the fact that stylistic pomposity is entirely un-Homeric." (Wilson/Homer, 2018, p.83)

Another point of divergence evident in both traditional classical scholarship and our own analyses is the question of how to translate Homeric Greek epithets into English. As seen in Figure 2.18 and discussed in Section 2.3.5, some translators in our sample were more inclined to use hyphenated, compound adjectives for what Arnold calls Homer's "double epithets"—words like *glaukopis*, often translated into English as *bright-eyed* or *gray-eyed*, which is a combination of the ancient Greek words *glaukós* (gleaming, gray) and *ṓps* (eyes)—than others in our sample. This should not be taken to mean, however, that Wilson's translation does not use compound-adjectives for epithets. Rather, her unique approach to translating epithets means she uses them as frequently but is often less directly literal (a famous characterization of Lattimore's translation) in her translation of them. She describes her approach for the reader in the translator's note which accompanies her *Odyssey*:

> "The formulaic elements in Homer, especially the repeated epithets, pose a particular challenge. The epithets applied to Dawn, Athena, Hermes, Zeus, Penelope, Telemachus, Odysseus, and the suitors repeat over and over in the original. But in my version, I have chosen deliberately to interpret these epithets in several different ways, depending on the demands of

> the scene at hand. ... In an oral or semiliterate culture, repeated epithets give a listener an anchor in a quick-moving story. In a highly literate society such as our own, repetitions are likely to feel like moments to skip. ... I have used the opportunity offered by the repetitions to explore the multiple different connotations of each epithet." (Wilson/Homer, 2018, p. 84)

Consequently, the word which Lattimore translates as *great-hearted* is *noble* or sometimes *valiant* in Wilson's translation. Similarly, Penelope is most often *circumspect* in Lattimore's translation of the *Iliad*; in Wilson's she is *shrewd, intelligent, careful,* and *wise*. We are also likely observing in Figure 2.18 Wilson's preference for using single instead of compound adjectives, which may make her text more readable and render her diction more direct. Arnold believed the invention of "strange, unfamiliar adjective(s) ... such was as 'voice-dividing' for $\mu\acute{\epsilon}\rho o\psi$" was to be avoided in Homer for this reason. He says these compound adjectives result in "an improper share of the reader's attention is necessarily diverted to this ancillary word, to this word which Homer never intended should receive so much notice; and a total effect quite different from Homer's is thus produced." (Arnold, 1861, p. 91)

The discussion of epithets naturally ties into our results regarding word pairs, as differences in the translation of epithets naturally produce these pairs. But, as with the *noble:great-hearted* pair in Figure 2.18, the word pairs we identified in Figures 2.16 and 2.17 (*cloister:halls, bard:singer, ghost:spirit, replied:addressed, shelter:tent, immortals:deathless*, etc.) needed to be confirmed by close-reading methods of the texts before we could definitively say we were seeing what we thought we were. While the Allotaxonometer brought many word pairs to the surface, we couldn't be sure they

were really the result of specific instances of divergence in translation without going into the text to confirm it. There are also instances where the second half of the word pair doesn't appear on the list—*vessels* is on Lattimore's half of the divergence contribution list, but Wilson's equivalent (*ships*, usually) is not. Nevertheless, the allotaxometer points us to an area and we can learn more by digging into the text.

This brings us to a reflection on the necessity of joint distant- and close-reading methods in this kind of work. Allotaxonometry identifies areas where systems diverge, and sometimes identifies things that we would not through a close-reading approach, but the significance of that divergence is often a question best-addressed by the latter than the former. The appearance of any given word in these plots means little outside of context—words can have many meanings, for instance, and knowing simply that translator *does* use a word in their text tells you nothing about *how*, *why*, and *where* they use it.

Therefore, our ability to interpret the results—to understand what we were seeing in the allotaxonometric plots, in other words—was dependent on both our prior knowledge of the texts and close-reading investigations performed based on the results of allotaxonometric analysis. The findings regarding Pope's rhyme scheme, for instance, may not have registered at all had we not been already aware that his translation was written in rhyming couplets. Three words in a plot which can be rhymed is not unheard of, and it could have been dismissed as mere coincidence. Similarly, we can propose an explanation for the preponderance of transition words in Samuel Butler's translation of the *Odyssey* because we know that his is the only prose translation in our sample, and these function words tend to be more common in prose than poetry.

But it is not just in the identification of these results that we benefit from close-reading approaches. A connection to broader classical scholarship helps us understand the *significance* of these divergences in translations. In Section 2.3.6, we call attention to our findings that differences in how translators in our sample referred to enslaved characters in *the Odyssey* contribute significantly to divergence in allotaxonometric comparisons. The continued political relevance of these texts makes these difference especially important to understand. The observation from the allotaxonometric results that Emily Wilson was more inclined to use *slave* than *herald, maid, servant,* or *attendant* in her translations calls attention to the enormous question of how to deal with mentions of slavery in the Homeric epics, a debate of great significance to contemporary classical scholarship. The Ancient Greeks were a slave-owning society, and *the Odyssey* in particular features many characters who are described using one of the many Ancient Greek words for enslaved people. When discussing her choice to call these characters *slaves*, Wilson frequently mentions the example of *dmoe*, a word that refers specifically to a female enslaved person who works in the home. Where other translators might use a word like *maid* to translate *dmoe*, Wilson intentionally uses language that conveys the character's enslaved status at the expense of the specific nature of their work, believing that "(t)he need to acknowledge the fact and the horror of slavery, and to mark the fact that the idealized society depicted in the poem is one where slavery is shockingly taken for granted, seems ... to outweigh the need to specify, in every instance, the type of slave." (Wilson/Homer, 2018, pp. 88-89)

While the word *slave* undeniably evokes a different image to a modern British or American audience (who is more familiar with the chattel slavery of the modern period than the model of slavery in Ancient Greece) than an ancient one, this is

likely true of every word in the poem, and a modern audience seeing the Greeks owned "slaves" and interpreting that as a great evil is hardly missing the mark, in any case. In his lectures, Matthew Arnold critiqued the comments of Homeric translator Francis William Newman in saying: "But I advise the translator not to try 'to rear on the basis of the Iliad, a poem that shall affect our countrymen as the original may be conceived to have affected its natural hearers;' and for this simple reason, that we cannot possibly tell how the Iliad 'affected its natural hearers."' (Arnold, 1861, p. 3) Given this, the question of whether to obfuscate the enslaved status of many characters in *the Odyssey* surely must be settled on different grounds.

2.4.1 Limitations

One finding that was consistent in every stage of standardization of the texts, from development to implementation, was that our efforts inevitably resulted in errors. On the one hand, allotaxonometric analysis was particularly well-suited to identifying errors in our methods. Errors often percolate to the top of the divergence contribution list—particularly when they are present in only one of the two texts being compared. For instance, if a mistake in our Phase 3 code turned Lattimore's *achaians* into *achaans*, it would still be a frequently used word that appeared only in one text, which naturally means it would percolate to the top of any divergence contribution list.

While we spent a great deal of time trying to fix any errors we identified, the nature of allotaxonometric analysis is that the removal of one error often just reveals another slightly less common error, and it is difficult to identify these errors outside of allotaxonometric analysis due to the sheer volume of tokens in each text. Because

of this, some errors were identified even after the standardization process was fully refined. For instance, the words *yon* and *lo* from Pope's translation of *the Iliad*, visible in 2.29, probably should have been standardized in Phase 5 with other antiquated words. These two words did not appear in the exploratory allotaxonometric analysis we used to identify which antiquated words needed to be standardized because they were much less common than the antiquated words we did standardize (e.g. *thy*, *o'er*, *ye*). Likewise, *goodly, forthwith,* and *thereon* from Butler's translation of *the Iliad* (Figure 2.30) should have been standardized but were not.

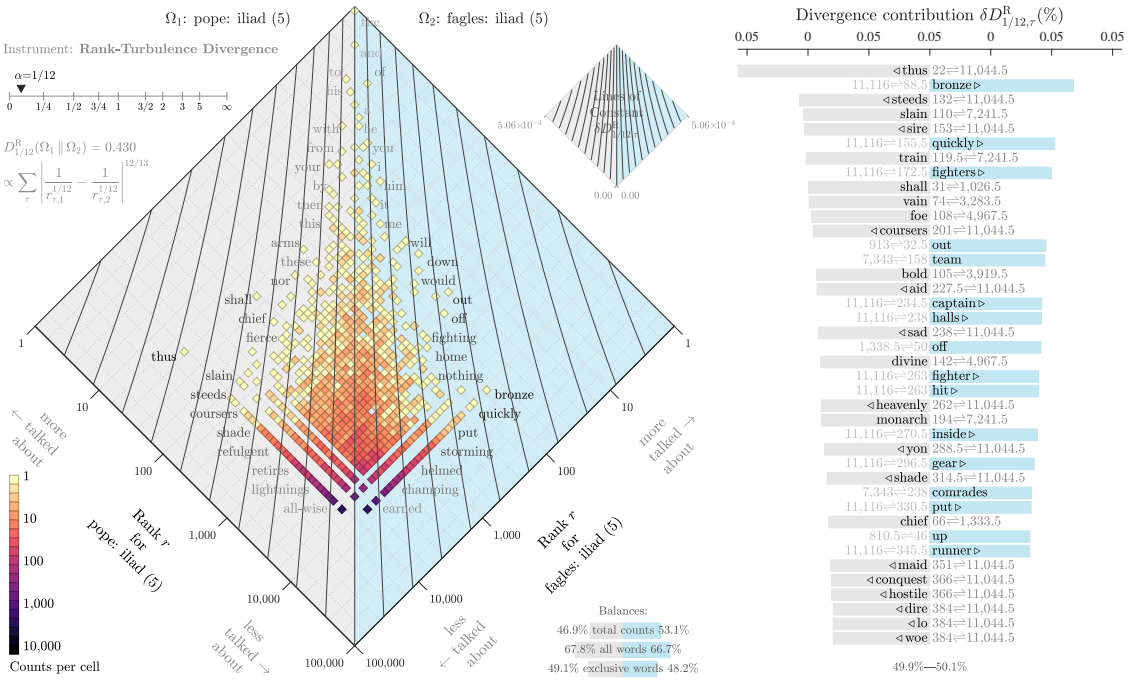

Figure 2.29: Pope's Iliad (Phase 5) vs. Fagles' Iliad (Phase 5) at $\alpha = \frac{1}{12}$

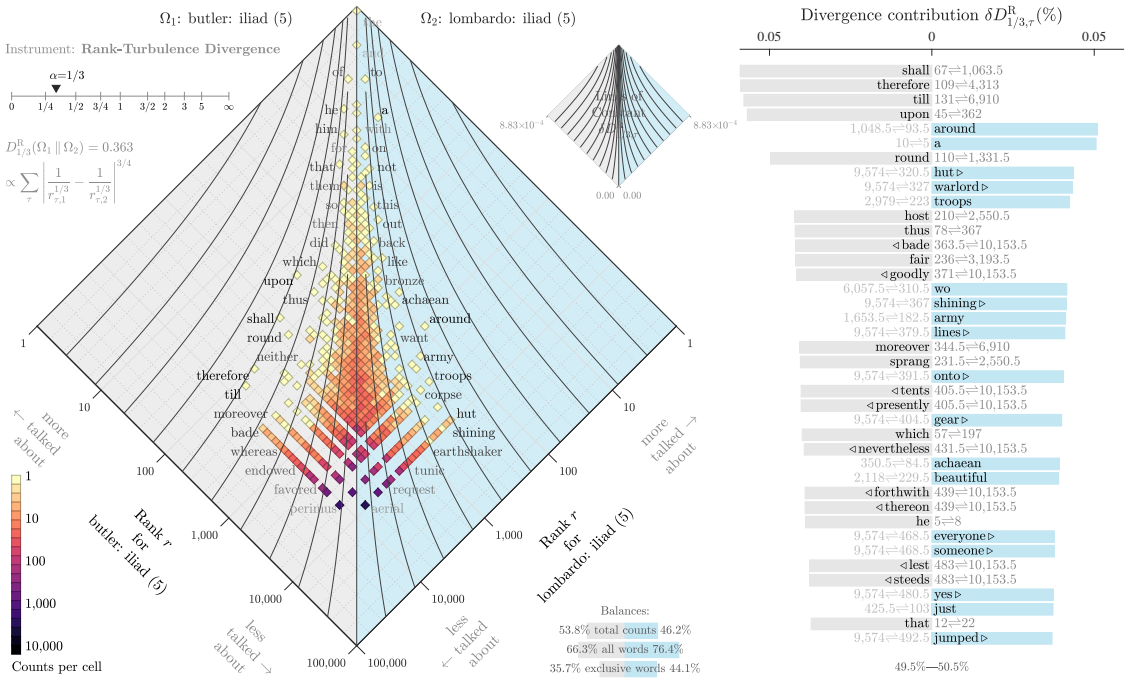

Figure 2.30: Butler's Iliad (Phase 5) vs. Lombardo's Iliad (Phase 5) at $\alpha = \frac{1}{3}$

Another more substantial limitation is not due to error but rather to the nature of allotaxonometry, which defines *significance* in divergence mathematically by identifying differences the relative frequency of items in two systems. Allotaxonometric analysis is therefore biased towards words of higher rank, and words that occur only a couple of times in one text or another are unlikely to appear in an allotaxonometric plot. As any student of stories knows, this is problematic. Sometimes, the most important moments in a story will hinge on a single word, and this is just not something this manner of analysis is suited to identify. There are moments where the divergence between the texts in our sample is extremely significant that the allotaxonometer is unable to pick up on.

Take these lines from Book 22 of *the Odyssey*, for instance. In this scene, Odysseus orders his son to execute the slave women who are accused of having sexual intercourse

with the suitors invading Odysseus' home in Ithaca. Telemachus decides to kill the women in a particularly brutal way—by hanging. Observe how Telemachus speaks of the women in each of the translations in our sample:

> "'To these shall we afford A fate so pure as by the martial sword? To these, the nightly prostitutes to shame, And base revilers of our house and name?"' (Pope/Homer, 1725)

> "'I shall not let these women die a clean death, for they were insolent to me and my mother, and used to sleep with the suitors."' (Butler/Homer, 1900)

> "'I would not take away the lives of these creatures by any clean death, for they have showered abuse on the head of my mother, and on my own head too, and they have slept with the suitors."' (Lattimore/Homer, 1965)

> "'No clean death for the likes of them, by god! Not from me—they showered abuse on my head, my mother's too! You sluts—the suitors' whores!"' (Fagles/Homer, 1996)

> "'I won't allow a clean death for these women—The suitors' sluts—who have heaped reproaches Upon my own head and upon my mother's."' (Lombardo/Homer, 2000)

> "'I refuse to grant these girls a clean death, since they poured down shame on me and Mother, when they lay beside the suitors."' (Wilson/Homer, 2018)

Emily Wilson writes about her approach to translating this scene in her 2017 essay for the New Yorker:

> "Contemporary translators and commentators often present the massacre of these women as if it were quite ordinary, and entirely justified. The murdered slaves are routinely described in contemporary American English translations as 'disobedient maids,' and are labelled as 'sluts' or 'whores'—a level of verbal abuse that finds absolutely no analogue in the Greek ... In the poem's original language, Telemachus refers to them only with *hai*, the feminine article—'those female people who ... slept beside the suitors.' In my translation, I call them 'these girls,' and hope to convey the scene in both its gruesome inhumanity and its pathos: 'their heads all in a row, / were strung up with the noose around their necks / to make their death an agony. They gasped, / feet twitching for a while, but not for long."' (Wilson, 2017)

The manner in which Pope, Fagles, and Lombardo make these embellishments of Telemachus' words are extremely significant, perhaps most especially in our modern context. But because none of these terms (*prostitutes, sluts, whore*) appear more than a handful of times in the text, they are not considered significant in an allotaxonometric context.

Chapter 3

Hedonometry

3.1 Methods

3.1.1 The Basics of Hedonometry

Hedonometry refers to the measurement of pleasure, happiness, or positive sentiment in language. For the purpose of this study, we are measuring the sentiment of text temporally using the labMT dictionary developed by researchers at the University of Vermont, which contains over 10,000 English-language words from a variety of corpora. (Dodds et al., 2015) Each of those words has a score on a 9-point scale of sentiment, with a score of 1 representing the most negative sentiment and a score of 9 representing the most positive sentiment.

The labMT dictionary can be used to plot the shape, or emotional arcs, of stories by generating a time series of the text. This involves searching the text for words which also appear in the dictionary, associating them with their sentiment score, and then using those scores to plot a computed average of sentiment over time for the

entirety of the text.

The specific methodology of hedonometry as proposed by Reagan et al. (2016) involves the selection of values for two tunable parameters. The first is the lexical lens, represented in this study with the variable δ or *delta*. The lexical lens determines which words will be included in the analysis based on where in the 1–9 sentiment continuum they fall. Generally, the lexical lens will be parameterized with a value of δ between 0 and 4, applied so as to exclude some share of words around the median, neutral point of the continuum. In other words, a lexical lens of $\delta = 1.5$ would mean words within 1.5 points on either side of neutral (5 on the continuum) would be excluded from analysis, and only words with a sentiment score between 1–3.5 and 6.5–9 would be included.

The lexical lens is implemented because the volume of words in a text with a near-neutral score can suppress the less-common, non-neutral words in the text. The words which are generally most common in English-language texts, function words (e.g. *the, and, of to, a, but*), all fall very near neutral in their sentiment scores. To say that these words are most common is perhaps understating it, as the frequency of words as a function of their rank will generally follow a power law distribution, whereby the most common words in the text are *extremely* common, and most other words appear only one or two times. The implication of this is that these neutral function words appear so many times that they must be suppressed if any change in sentiment over time is to be visualized.

When selecting a lexical lens for our analysis, we wanted to select a value that allowed for visualization of distinct moments in the text where sentiment was relatively higher or lower without making the detection of those moments so sensitive that the

plot was too noisy to interpret. Additionally, because so many words fall towards the neutral point, making the lexical lens too large may mean you exclude too many words from the text to get a clear picture of its emotional arc. Consider Figure 3.1 to see what share of words are excluded from James Joyce's *Ulysses* with a δ of only 1.5.

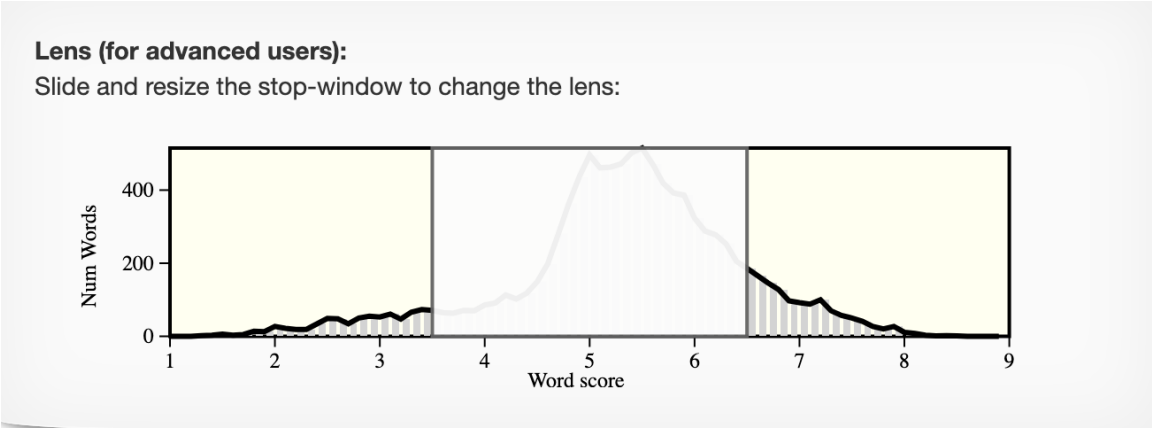

Figure 3.1: An Example of the Lexical Lens from Hedonometer.org

The selection of a δ value for this study was thus determined experimentally. Figure 3.2 shows the emotional arc time series plots for Emily Wilson's translation of *the Iliad*, generated at 3 different values for δ: 0.5, 1.5, and 2.5.

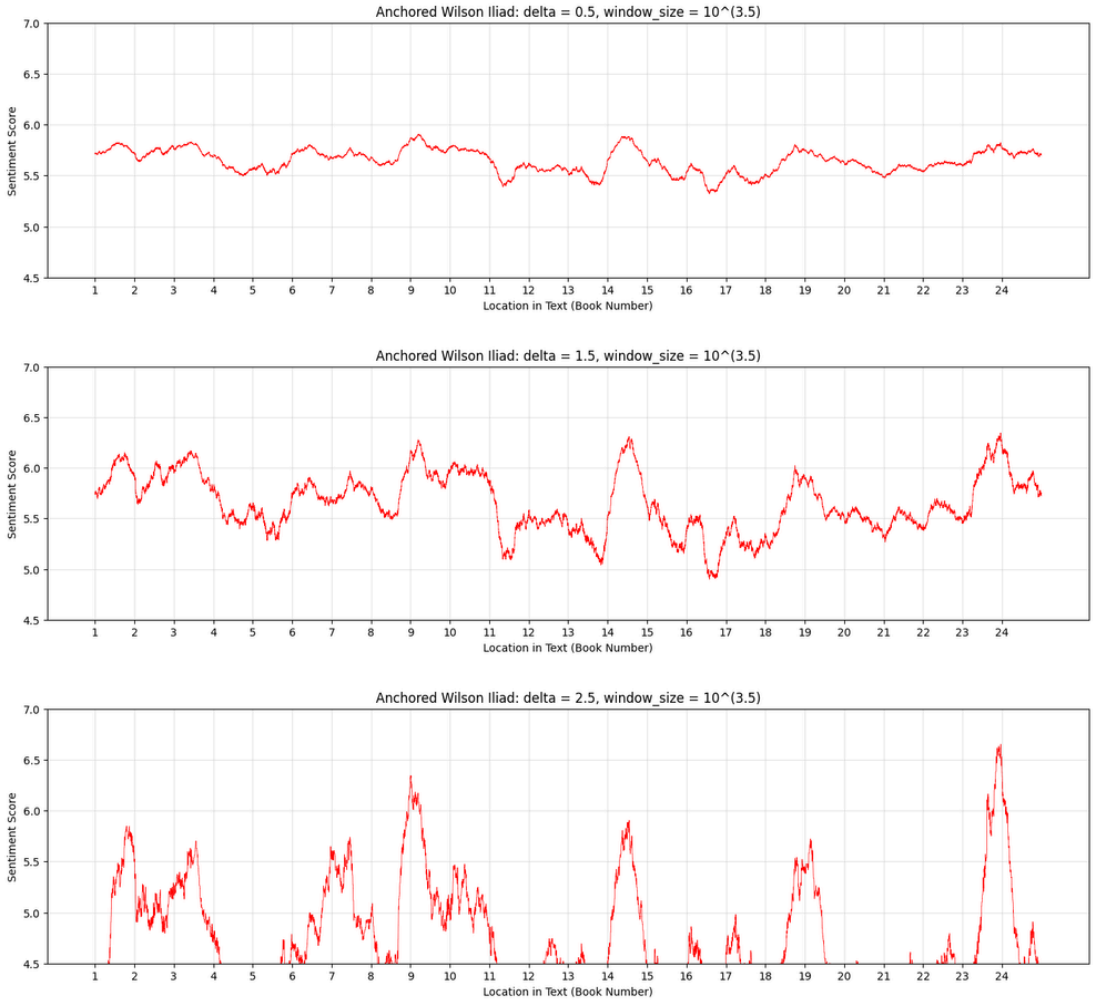

Figure 3.2: The Effect of Manipulating the Lexical Lens

$\delta = 0.5$ resulted in a plot that was too flattened. Too few neutral words were removed, which made identifying the parts of the text which were relatively high or low in their average sentiment score difficult. On the other side of the spectrum, $\delta = 2.5$ excluded too many neutral words and produced a plot that noisily swung from one extreme to another. The result is a story that never seems to have any neutral baseline from which it comes and goes in more intensely sentimental moments. $\delta = 1.5$

fit nicely in the middle, allowing for an easily interpretable plot with moments that are more positive, negative, and neutral in sentiment visible over the course of the story. This was the lexical lens we selected for all analyses in this study.

The second parameter for which a value needs to be chosen is the window size. The time series for the emotional arc plot is not generated by simply plotting a point at the sentiment value for each word as it appears in the text, because this would result in an unreadable, jaggedly-shaped plot whereby each of the tens of thousand of plot points in the text were on opposite poles of the sentiment range. In order to plot the sentiment of the text as a readable arc, it is necessary to plot points which average the sentiment values of many words in a given section of the text. This means, for example, the sentiment value plotted at the location in the time series which corresponds to the 50,000th word in the text is actually the average sentiment of that 50,000th word and a certain number of words which follow it. A visual example of this, from the 2016 Reagan *et al.* story shapes paper, is shown in Figure 3.3.

Figure 1 Schematic of how we compute emotional arcs. The indicated uniform length segments (gap between samples) taken from the text form the sample with fixed window size set at N_W = 10,000 words. The segment length is thus $N_S = (N - (N_W + 1))/n$ for N the length of the book in words, and n the number of points in the time series. Sliding this fixed size window through the book, we generate n sentiment scores with the Hedonometer, which comprise the emotional arc [22].

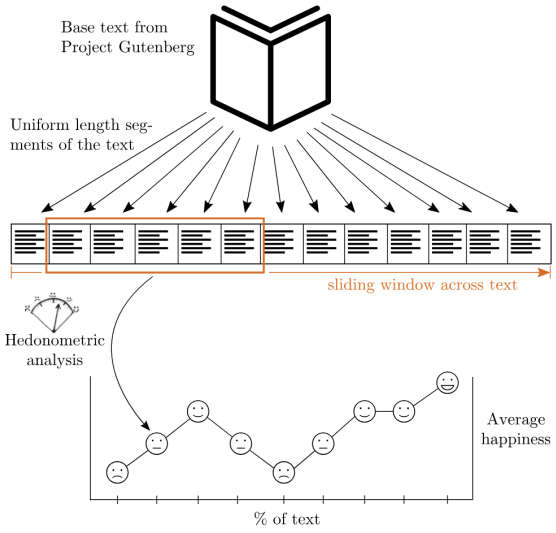

Figure 3.3: An Illustration of Window Sizes in Hedonometric Analysis (from Reagan et al., 2016)

By taking these averages, and computing each subsequent point in the time series by shifting the window one word to the right and taking the average of that new set of words, the generated story arc is smoothed out enough to be interpretable. This approach also makes logical sense, as our perception of sentiment is not dependent on any single word we read in a story but rather the collection of words which comprise scenes of particular emotional significance.

In choosing a window size, we need to balance two concerns. The first is that the window size cannot be too big. A window that is the same size as the entire text, for instance, will result in just one sentiment value, the average sentiment of the book. This would result in us losing the time series and story arc element altogether. The window size therefore needs to be small enough to allow us to see the fluctuations in sentiment in individual moments in the text. However, making the window too small will overpower the larger structure of the story in favor of dramatic fluctuations caused by the presence of individual highly sentimental words.

As with the lexical lens, window size was chosen experimentally. In Figure 3.4, we show the effect of three window sizes on the shape of the emotional arc of Emily Wilson's translation of *the Iliad*.

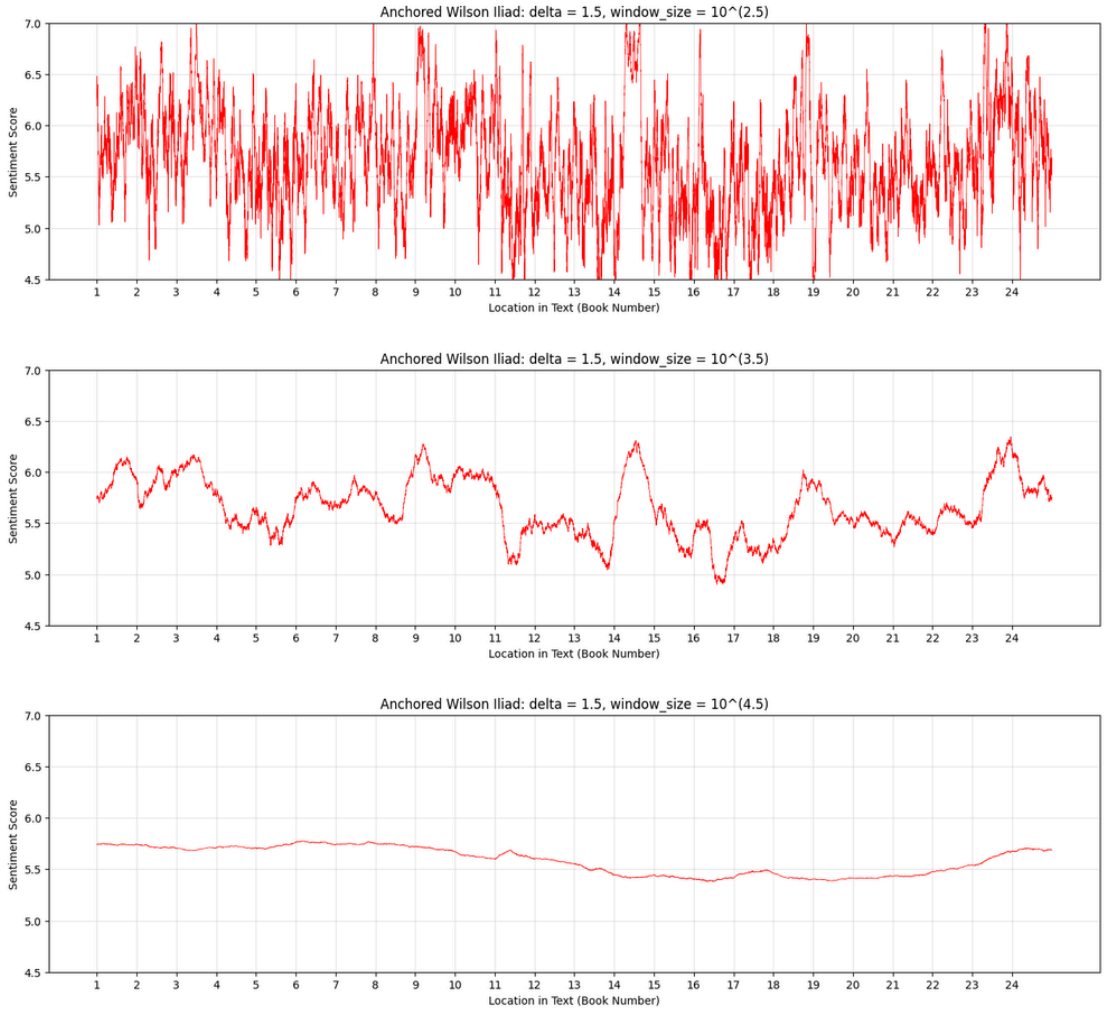

Figure 3.4: The Effect of Manipulating the Window Size

We can see in this figure that a window size of $10^{2.5}$ is too small. The plot swings from one extreme to another, and the story shape is lost and overpowered by words of high sentimental significance. A window size of $10^{4.5}$, however, is clearly too large. There are two few data points in the plot, and too many words being averaged to generate each point such that the sentiment score of each is roughly the same. A window size of $10^{3.5}$ strikes a good balance between these two extremes.

3.1.2 This Study

The purpose of this study was to plot and compare the emotional arcs for all of the translations of *the Iliad* and *the Odyssey* in our sample. The versions of the texts we used, sourced from open access platforms like Project Gutenberg (gutenberg.org), Internet Archive (archive.org), and the Perseus Collection (perseus.tufts.edu) underwent the complete, six-phase standardization process described in the Methods section of Chapter 2.

We chose to use tokenized and completely standardized files for this project because our standardization process broadly resulted in texts whose words were more likely to occur in the labMT dictionary. The American English *practice*, for instance, occurs in the labMT dictionary, but the British English *practise* does not. labMT is generally biased towards American English spelling, and we didn't want words in our translations to be omitted for analysis simply because they were spelled differently than the version of that word in the labMT dictionary. labMT is also biased towards modern English, but almost all of the words we modernized in Phase 5 are function words that would be excluded by the lexical lens anyway, so this was less of a concern.

To produce the plots, we fed the standardized text files of the translations into a Python script that did the following:

1. Searched the text for words that also existed in the labMT dictionary,
2. If found, matched the word to its sentiment score in the labMT dictionary,
3. Constructed a time series by computing the score of the first window (loc[0] + window size) and then every subsequent point by shifting the window one to the right and calculating the average of the words in that window repeatedly.

4. Built into the time series 23 "breaks" where a new chapter (called *books* in the Homeric epics) began so that translations could be aligned on the same time series plot.

Two small notes about this process. The first pertains to item three, where windows were identified and their average sentiment computed. The nature of this manner of analysis is that not all words in our texts will also appear in the labMT dictionary. This is especially true of translations of Homeric epic, which are in many ways more different than the texts in the corpora from which labMT's words are sourced. Additionally, many words in the text that *are* in the labMT dictionary are excluded from analysis due to them falling outside of the lexical lens of interest to us. The implications of this are that a window which contains, say, words 1–3,163 of the text will not actually contain 3,163 sentiment values to be added up for the computation of the sentiment score of that window, but significantly fewer than that. This is an inevitable part of textual sentiment analysis using a dictionary-based method, but we were reassured by what was generally good coverage of our texts by the dictionary. See Table 3.1 for information on what share of total words and word types were in both our translations and labMT dictionary.

Table 3.1: Share of Tokens and Types in Our Translations That Also Appear in the labMT Dictionary

Share of Tokens in Text Also in labMT		
Text	**Iliad**	**Odyssey**
Alexander Pope	0.8045	0.8107
Samuel Butler	0.8595	0.9053
Richmond Lattimore	0.8644	0.892
Robert Fagles	0.8237	0.8614
Stanley Lombardo	0.8464	0.8846
Emily Wilson	0.8461	0.8886

Share of Types in Text Also in labMT		
Text	**Iliad**	**Odyssey**
Alexander Pope	0.3521	0.3691
Samuel Butler	0.3804	0.469
Richmond Lattimore	0.3848	0.4529
Robert Fagles	0.353	0.4009
Stanley Lombardo	0.3913	0.4555
Emily Wilson	0.4034	0.4741

The second note is regards item four. In early attempts to compare the emotional arcs of different translations of the Homeric epics, we observed something of a lag effect for certain translations. While our initial efforts standardized the time series so all arcs began and ended at the same place (slightly shrinking or stretching the arcs of certain

translators whose translations were longer or shorter than average, respectively), we noticed that translations were still not fully aligned because the amount of time each translator spent in each chapter/book of the story was not consistent across our sample. In other words, Book 5 of Richmond Lattimore's Iliad might comprise 7% of the total length of his translation, whereas in Stanley Lombardo's it is only 5% of the total length. To address this issue, we chose to align translations not just at the beginning of Book 1 and end of Book 24, but at the beginning and end of each book in the texts.

3.2 Hedonometry Results

Before we examined the emotional arcs of individual translators in our sample, we computed average arcs for each of the Homeric epics by averaging the arcs of all the individual translations of that text. This allowed us to produce figures like 3.5, where something like a general arc of the Iliad and Odyssey are plotted against one another to reveal that *the Iliad* is generally less positive in sentiment than *the Odyssey* throughout the progression of the texts.

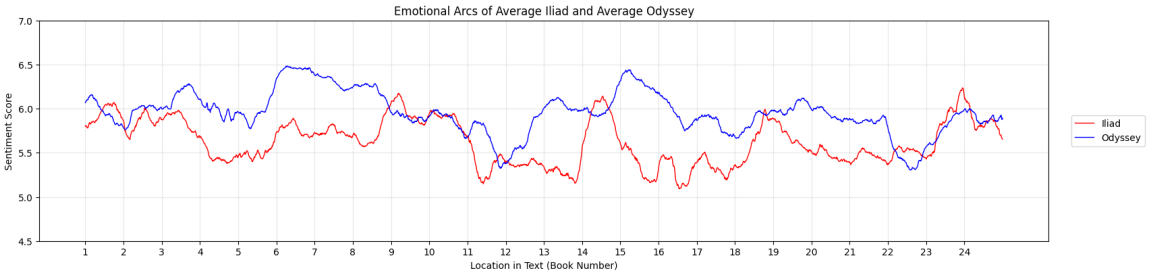

Figure 3.5: The Average Emotional Arcs of the Iliad and the Odyssey

The average arcs were also used in the creation of Figures 3.6 and 3.7, where the

emotional arcs of each text are annotated with important plot points which occur at that chronological location in the text.

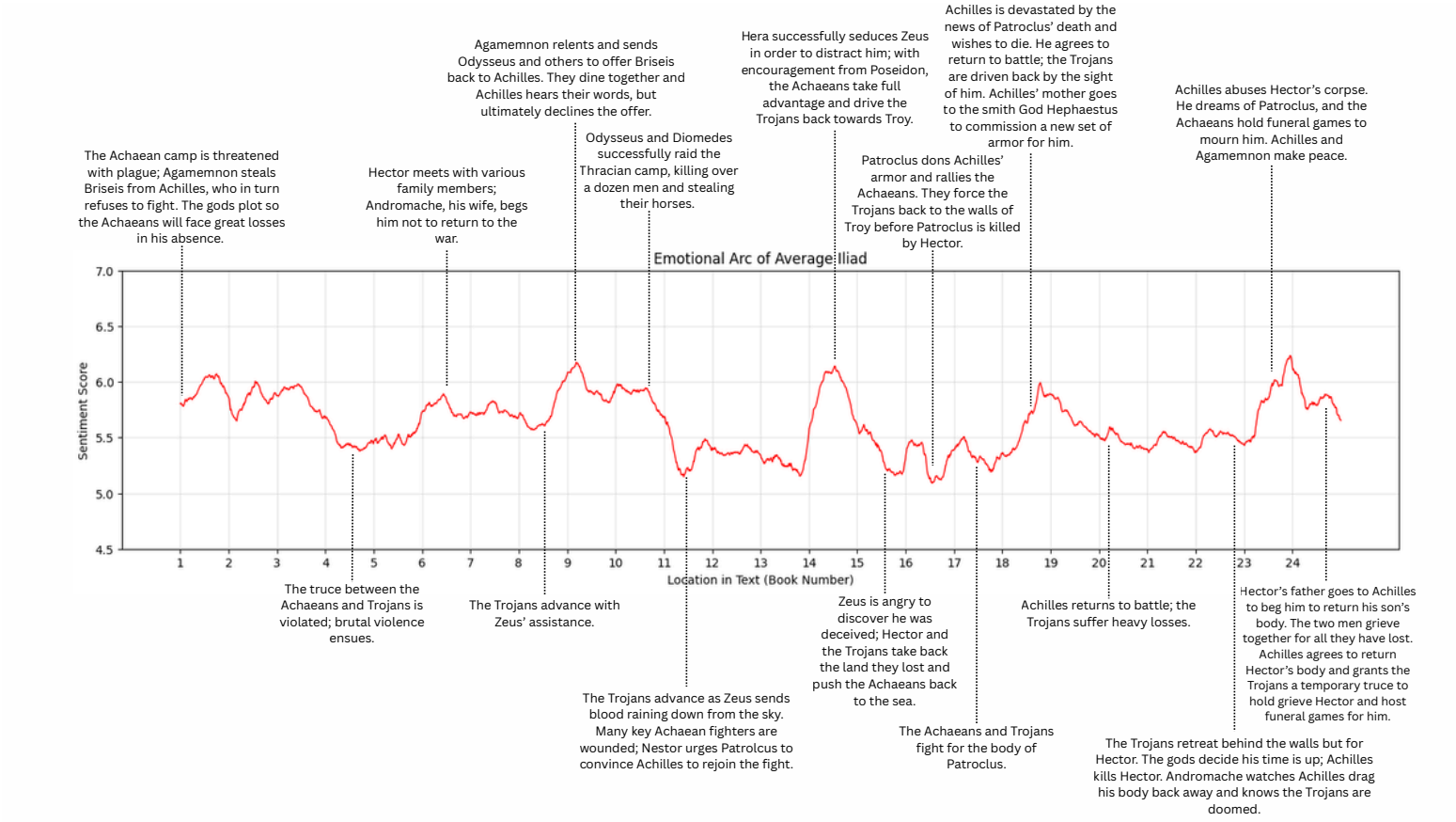

Figure 3.6: The Average Emotional Arc of the Iliad, Annotated with Important Plot Points

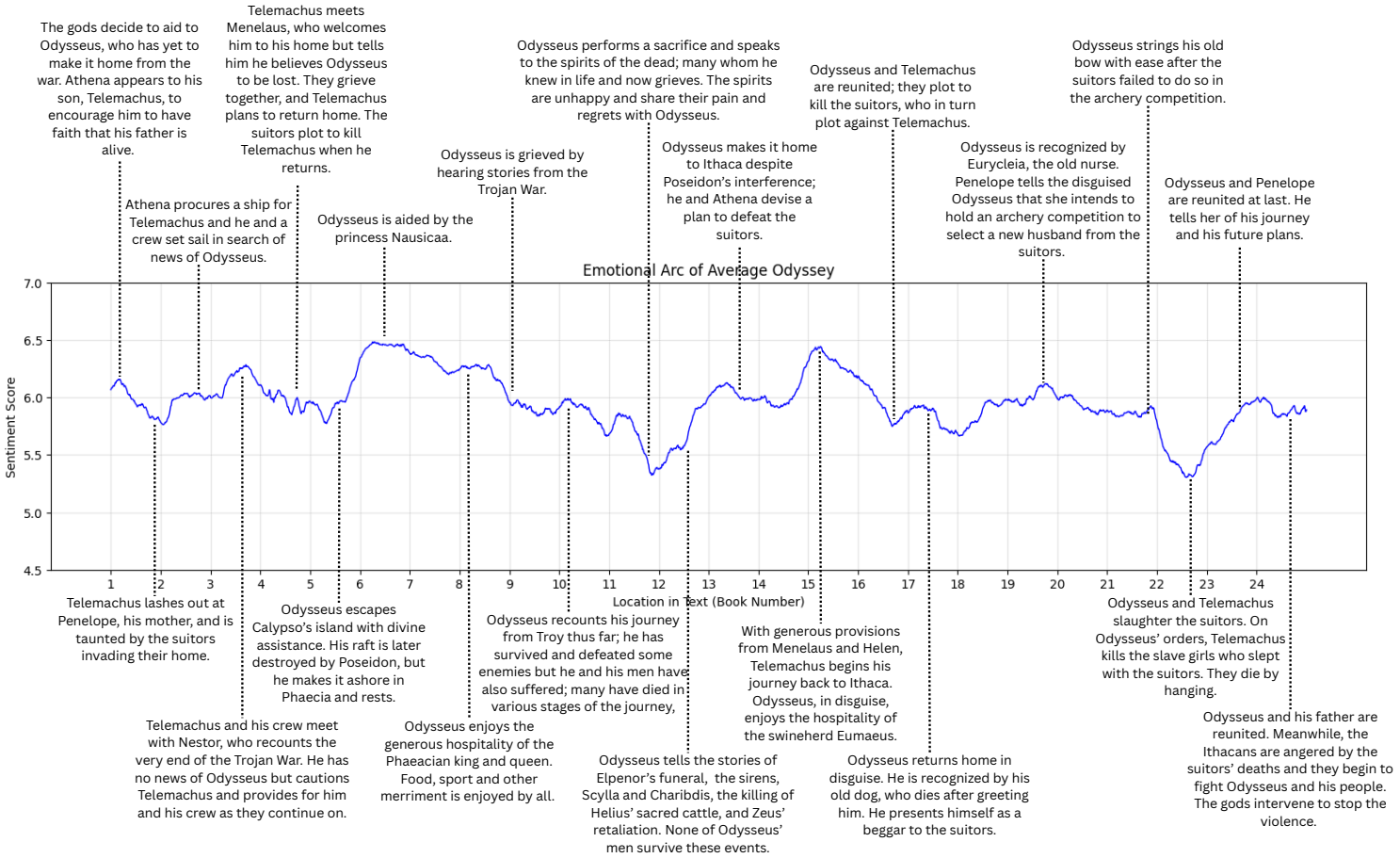

Figure 3.7: The Average Emotional Arc of the Odyssey, Annotated with Important Plot Points

The individual arcs of our translations of *the Iliad* and *the Odyssey* (e.g. the emotional arc of Stanley Lombardo's *Iliad* or Samuel Butler's *Odyssey*) are plotted against one another in Figures 3.8 and 3.9 to allow for an immediate comparison of their alignment with one another.

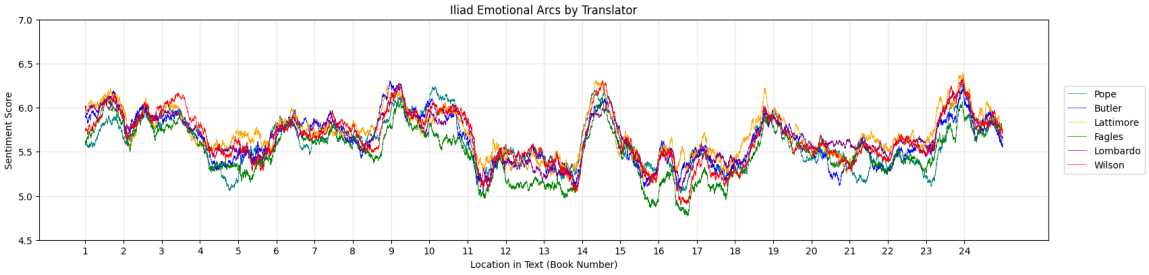

Figure 3.8: Emotional Arcs of All Translations of the Iliad in Our Sample

In Figure 3.8, we can see that translations are quite aligned in sentiment. Certain translators seem to be sitting below or above average at various points (Fagles and Lattimore, respectively), while other translators are clustered quite tightly throughout. We can also observe certain moments in the text in which translators are collectively more or less aligned: Books 4–6, 11–13, 16–17, and 20–23 show the greatest discrepancies in sentiment, while Books 2, 14, 18, and 24 appear more strongly aligned.

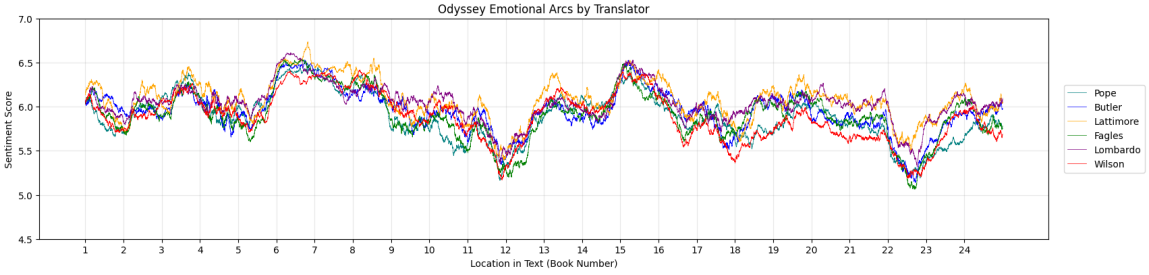

Figure 3.9: Emotional Arcs of All Translations of the Odyssey in Our Sample

In Figure 3.9, similar phenomena are observed. Lattimore again sits above the

rest in average sentiment, Fagles below. There is more sentimental convergence in Books 11–15, with a great deal of difference in the last quarter of the Book, where it almost appears as though 2 or 3 different groups emerge with their own arc regime.

The degree to which certain translations differ from the average arc of that text was naturally of interest to us based on these results. In order to more clearly visualize that relationship, we plotted each translation of *the Iliad* (Figure 3.10) and *the Odyssey* (Figure 3.12) against the average arc of those texts.

Figure 3.10 shows that, while all translations have individual moments where they are above or below the average curve, some translations are more consistently above or below average sentiment than most. The arc of Richmond Lattimore's translation of the Iliad, for instance, consistently sits above the average arc, while that of Robert Fagles is consistently below.

To see this illustrated in another way, also consider 3.11, where the average emotional arc of *the Iliad* is completely flattened to allow for better visualization of the distance between an individual translation's arc and the average arc of that text.

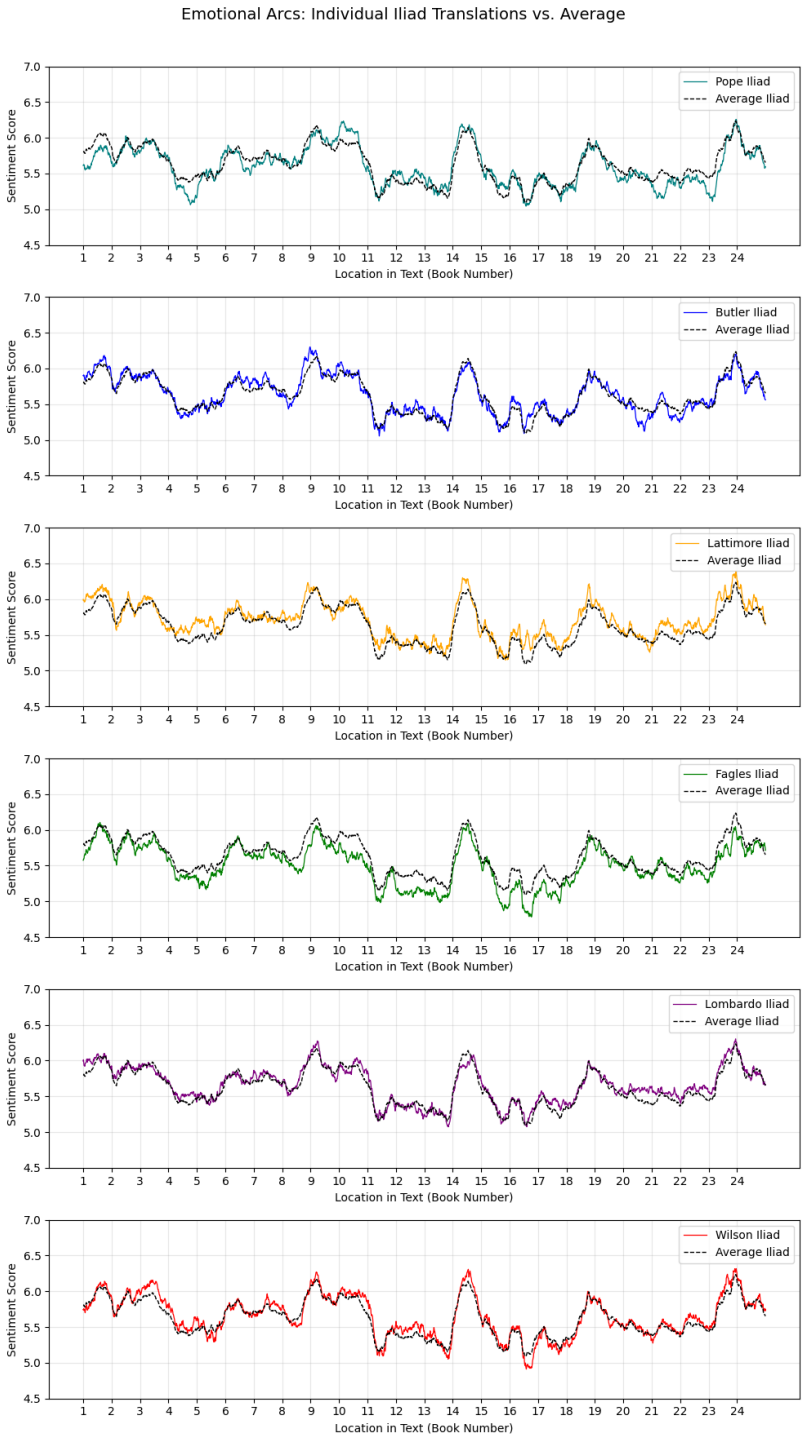

Figure 3.10: The Iliad Emotional Arc Comparisons: Translator vs. Average

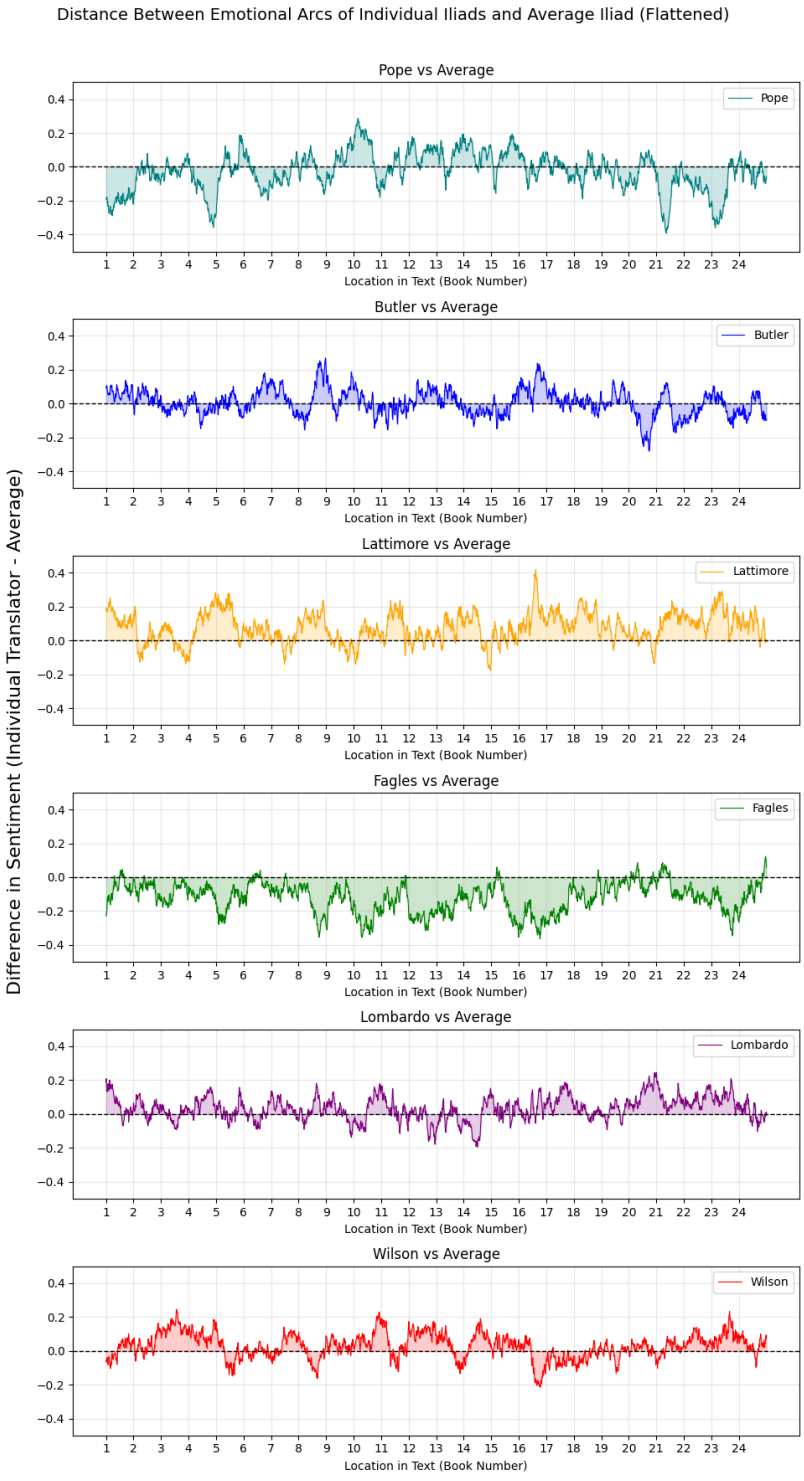

Figure 3.11: Flattened Emotional Arc Comparisons of the Iliad: Translator vs. Average

The same analyses were performed for translations of *the Odyssey*. See Figure 3.12, below for the first set of comparisons between the emotional arcs of individual translators and that of the average for the text. Once again, Richmond Lattimore's translation of the text is consistently above average in sentiment, while Robert Fagles' is below. We can also observe the more intense disagreement of the last quarter of the text which we identified in Figure 3.9: Lattimore and Lombardo are well above average sentiment in this section, with Emily Wilson notably below, as is Alexander Pope (in the last two books of the text, in particular). These findings may be even easier to see in Figure 3.13, where the average emotional arc of *the Odyssey* is flattened.

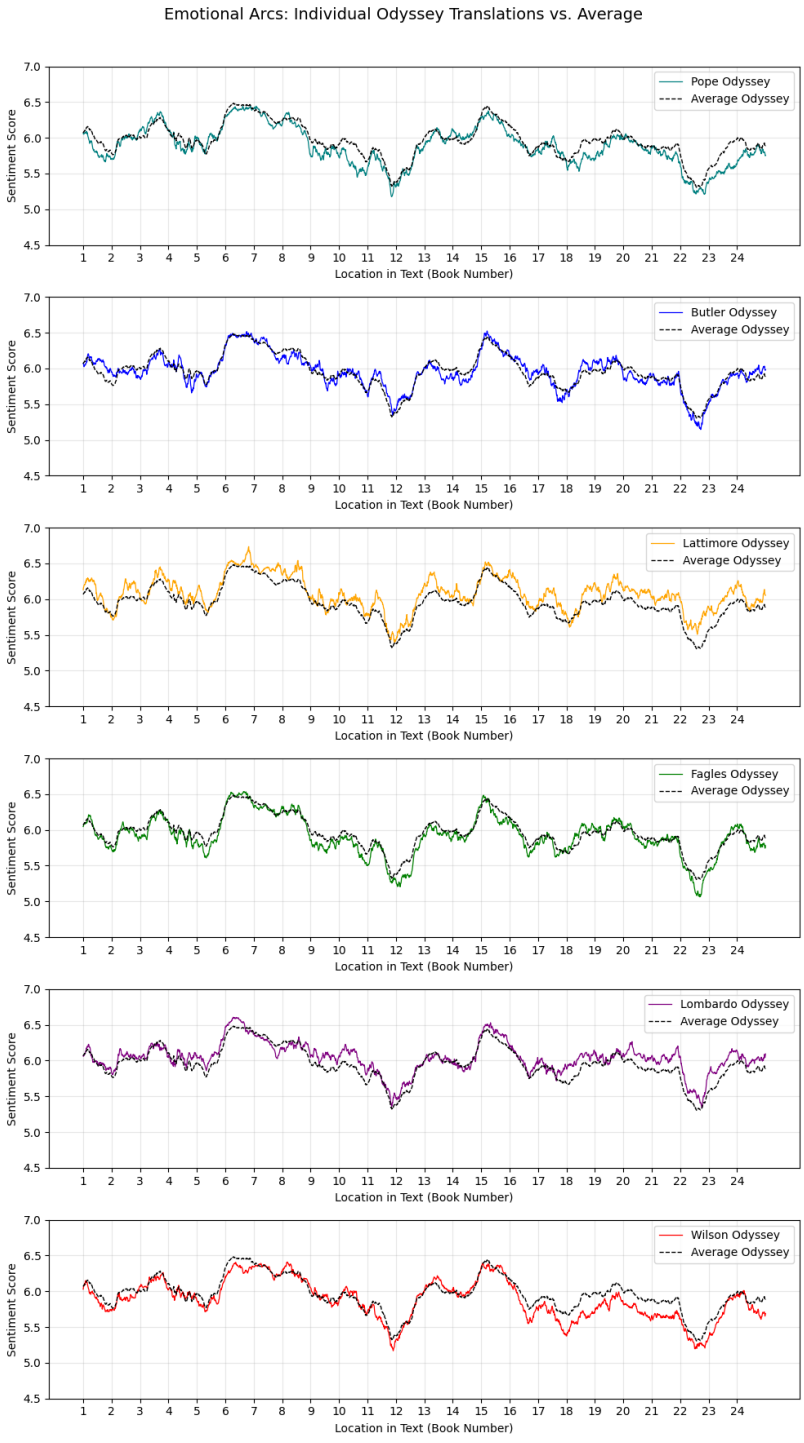

Figure 3.12: The Odyssey Emotional Arc Comparisons: Translator vs. Average

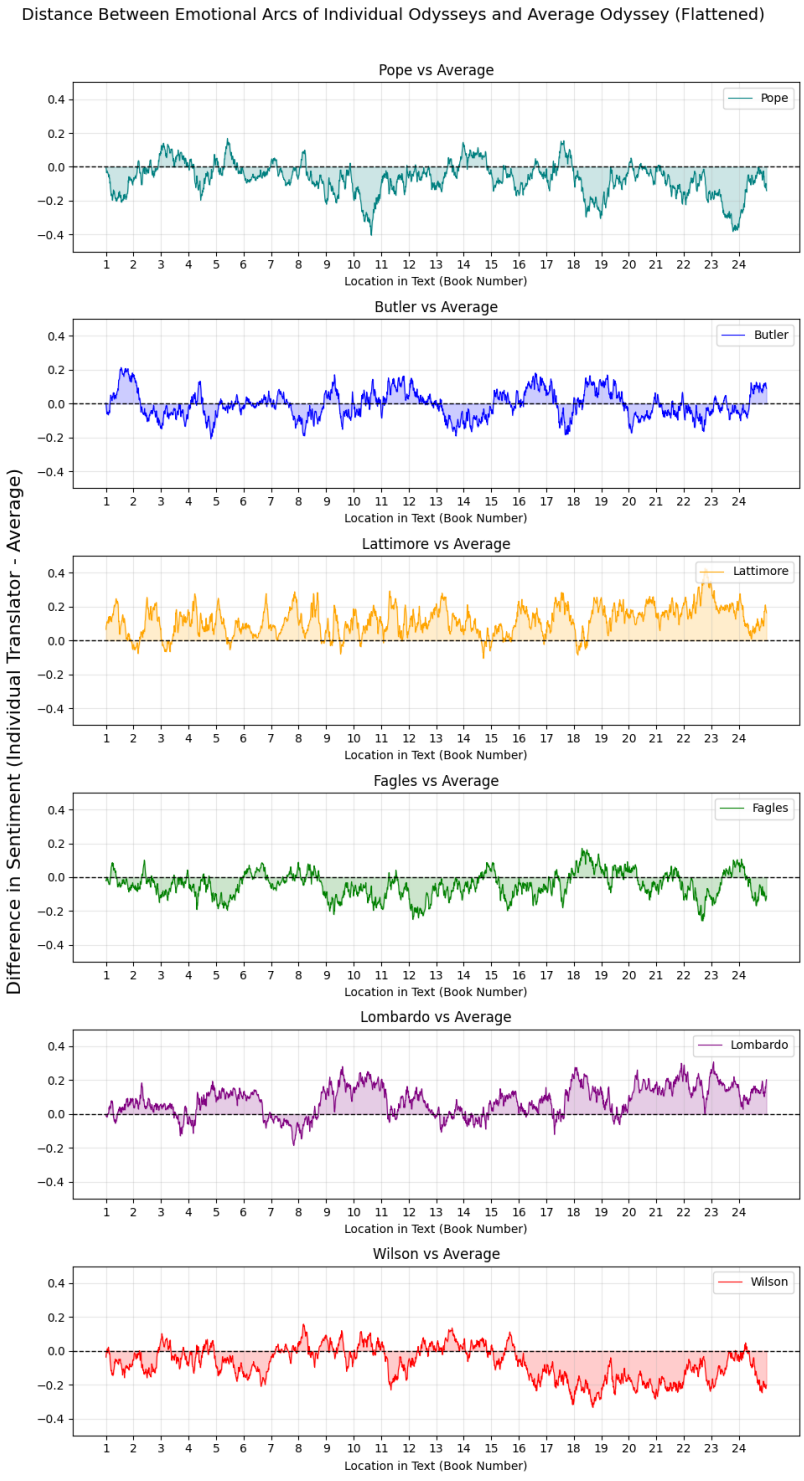

Figure 3.13: Flattened Emotional Arc Comparisons of the Odyssey: Translator vs. Average

While the below plots are too dense to easily follow the course of an individual translator's arc, and those translations which hew more closely to the average arc are harder to spot, it may be useful to compare the degree of the distance between certain translations and the average arc. For instance, in Figure 3.14 we observe that Alexander Pope's translation has moments that differ more greatly from average than is typical, and that Robert Fagles' translation has multiple stretches where it is quite far below average sentiment relative to other translations. In Figure 3.15, the degree of divergence with the average arc in Books 18–24 is unlike anything in the rest of the text, and we can see that Alexander Pope's dramatic downward spike in Book 23 is equal in magnitude to Richmond Lattimore's upward spike in Book 22.

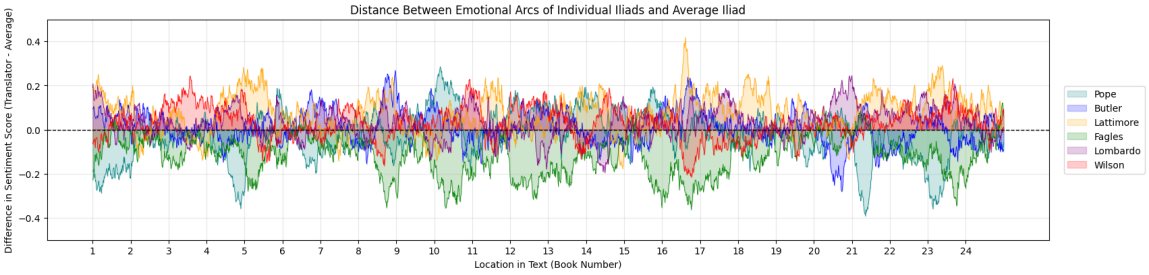

Figure 3.14: Flattened Iliad Emotional Arc Comparisons (Stacked)

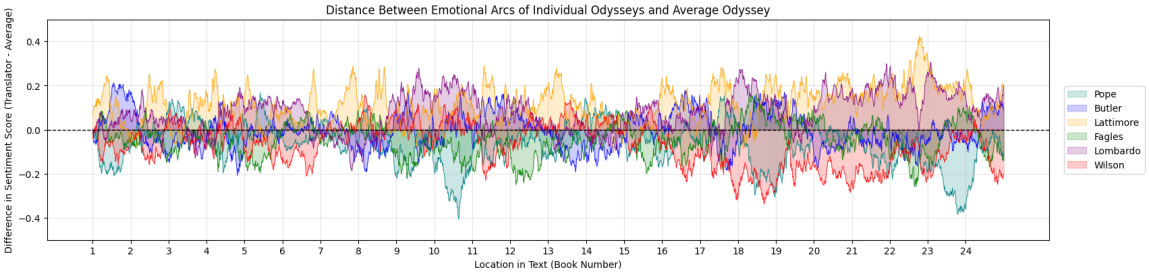

Figure 3.15: Flattened Odyssey Emotional Arc Comparisons (Stacked)

3.3 Hedonometry Discussion

3.3.1 The Shape of Homeric Epic

Our finding that the sentiment of *the Iliad* was consistently lower than that of *the Odyssey* is consistent with many classicists' conception of the poems. Consider the below, from Emily Wilson:

> "*The Odyssey* has a sense of playfulness, whereas *the Iliad* just has this sort of relentless focus on a very claustrophobic and often very dark world where people are having terrible, violent deaths and intense rage and grief and you never really get out of that. You have moments of respite during the occasional comic scene, like Hera's seduction of Zeus or the funeral games and slipping in the dung, but it doesn't have nearly as many domestic scenes and almost funny scenes as *the Odyssey* does." (Hilden, 2023)

However, classicists also speak of *the Odyssey* as being more tonally varied than the Iliad. See the below from classicist Lillian Doherty:

> "The tone of *the Odyssey*, I submit, is *not* sustained like that of *the Iliad* but various: now serious, now humorous, now gently ironic, now dignified—and even, occasionally, noble. The Odyssey requires a translator whose touch is light enough to capture the humor, but who can rise to the true Iliadic strain in the right places (such as Odysseus' response to the taunts of the young Phaeacians or his speeches to the suitors as he takes his revenge)." (Doherty, 1985, p. 164)

We might expect to see this tonal variety in the comparison of the average arcs of the Iliad and Odyssey (see Figure 3.5), but we don't seem to see this in the results of our analysis. The difference between the minimum and maximum sentiment score for each arc is roughly equivalent, and the frequency with which the arcs swing between these high and low points does not look to be greater for *the Odyssey* than *the Iliad*.

Regarding the shape of the epic's emotional arcs individually, it is unsurprising that neither epic neatly fits into one of the neat shapes proposed by Vonnegut. A methodology like ours is unlikely to produce such a perfect shape. Even the stories which were found to most closely align with such a shape in the (2016) Reagan *et al.* paper had deviations from the emotional arc they were fit to. There are a few reasons this may have been the case in our particular study. For one, the texts we're analyzing contain roughly 100,000 words per story. With that many data points, a generated emotional arc is bound to experience many more inflection points and relative extrema than the handful in Vonnegut's idealized shapes.

Another explanation for this may lie in the Homeric epics' unique structure. The Iliad and Odyssey have undeniable structural differences from the stories composed and disseminated in the last few centuries. The stories most of us think of today were likely composed and distributed almost exclusively in written mediums like the novel. These mediums in term have structural norms which are generally followed, among them a tidy beginning, middle, and end. But this is not the way the Homeric epics were composed or disseminated. We know that ancient experienced these stories through oral performance, and we also know that the stories would have been performed in their entirety very infrequently, if ever, because of how long it would take. A twenty-four-hour-long performance of the epics would be a daunting task for

not just the rhapsode performing but for the audience listening. As a result, the epics were usually performed in smaller episodes. These smaller sections of the poems had beginnings, middles, and ends of their own.

The *Patrokleia*, or "Patroclus Episode" in Books 16 and 17 of *the Iliad*, for instance, is a story all its own. Patroclus dons Achilles' armor to trick the Trojans into thinking Achilles has returned to the war, and this scheme allows the Achaean forces to successfully push the Trojans back away from the Achaean camp. Patroclus was advised not to risk going further than that, but in the moment he decides to take advantage of their momentum and push on. He forces the Trojans all the way back to the walls of Troy before he is tragically struck down by a Trojan fighter and then killed by Hector. A brutal fight for the possession of Patroclus' body then ensues between the Achaeans and Trojans. Vonnegut would surely classify it a classic Icarus story, and a look at the corresponding sections of the annotated time series for *the Iliad* undeniably show the precipitous rise and fall of sentiment. Thus, maybe the Homeric epics are best understood as a collection of smaller, discrete story shapes within the greater emotional arc.

3.3.2 The Annotated Emotional Arcs

The annotated arcs in Figures 3.6 and 3.7 allow us to examine the degree to which a window's average sentiment score corresponds to our perception of the emotional valence of the story at that time point. Before we discuss the degree to which the results meet our expectations, though, it is prudent to remind ourselves what exactly we are doing when we perform sentiment analysis with the labMT dictionary. When a score is computed for any given window, our methods do not entail a holistic and

complex analysis whereby the section of the text is read through some artificially intelligent means and graded on the sentiment spectrum accordingly. Rather, this method involves simply averaging the scores of the individual words in that window which are in the labMT dictionary and within our lexical lens. As a result, an overall positive sentence like "With this fatal blow, the villain was defeated." would nevertheless result in a negative sentiment evaluation from labMT due to the very low scores for words like *fatal*, *villain*, and *defeated*.

It is on this basis that some dismiss this method of sentiment analysis altogether. However, semantically negative language in a text undoubtedly has an effect on the overall happiness or emotional valence of the story, and the ability for this method to detect high and low emotional points in a story has been demonstrated. (Reagan et al., 2016) So we point this out not to call into question our findings, but to help us interpret what we have found.

The emotional arc of *the Iliad* paints a complicated picture in this regard. Are any moment's in a war story truly happy or positive? Perhaps it depends on the story. British or American audiences may have little trouble identifying moments of victorious triumph, relief, and gratification in a story told in World War II, where they can easily identify with one side and cast the other as the villains of the story. But *the Iliad* is not that kind of story. We spend time with characters on either side of the war, and experience the hopes, fears, and heartache of Trojans and Achaeans alike. We grieve for Trojan Andromache as she watches her husband's corpse be dragged around the walls of Troy by wrathful Achilles and knows it means her own death is not far off; so too do we grieve for Achilles himself when he learns Patroclus has been killed, tearing at his hair and wailing in grief, utterly inconsolable and wanting to

die. The moments of triumph for one side in *the Iliad* (e.g. Books 10, 11, 14, 22, etc.) always necessarily mean defeat for the other, and the nature of these moments means semantic analysis of them is complex. The Achaeans' driving the Trojans away from their ships in Book 16 might be a victory for their side, but it is one that involves a lot of the most negatively scored words in the labMT dictionary: *death*, *blood*, and *killed*.

Despite this, there is logic to the annotated emotional arc of this text. The lowest average sentiment points of the arc largely correspond to negative moments in the story, particularly if we adopt an Achaean perspective—the particularly desperate moments for the Achaean forces in Books 11 and 15, for instance, or the collapse of the truce in Book 4, or Patroclus' death in Book 16. Conversely, moments that are comparatively more positive correspond to sentiment peaks: Agamemnon's apology (by proxy) to Achilles in Book 9, Hera's seduction of Zeus and the Achaean's subsequent advances in the war in Book 14, and the funeral games held for Patroclus and Achilles and Agamemnon's reconciliation in Book 23. There are some moments where the sentiment score is less intuitive, namely the scene in Book 18 where Achilles learns of Patroclus death, which is a relative maxima in the emotional arc, though the computation of the sentiment of this moment is likely complicated by the scene where Hephaestus crafts a glorious set of armor for Achilles and where the sight of him drives the Trojans away from the Achaean camp as they flee in terror.

The annotated *Odyssey* emotional arc (Figure 3.7) similarly makes sense and doesn't. The moments we might consider to be maxima on the graph (the reunion of Telemachus and Odysseus, or that of Penelope and Odysseus) are actually quite close to the middle of the arc, though the latter reunion does correspond to a relative

maxima after a particularly brutal set of scenes in Book 22. Instead, it is moments of generous hospitality that seem to rise to the top, such as when Odysseus is aided by the Princess Nausicaa after a shipwreck off the coast of Phaeacia, or when he is welcomed in disguise by the swinelord Eumaeus, or when Telemachus is sent off with generous provisions by Menelaus and Helen.

The two most extreme minima of the arc, though, correspond to what could be plausibly considered *the Odyssey's* darkest moments: the first is a scene in which Odysseus performs a sacrifice and then speaks to the largely very unhappy spirits of the dead (Book 11), and the second is the sequence of scenes where Odysseus and Telemachus slaughter the invading suitors and then kill the slave girls who slept with them and are thus considered guilty by association (Book 22).

3.3.3 Limitations in Explaining Divergence and Extrema

While we have attempted to explain the peaks and valleys of the average emotional arcs of *the Iliad* and *the Odyssey*, understanding why individual translations diverge from the average arc is at any given moment is much more difficult, and requires further analysis that is outside the scope of this inquiry. Our prior knowledge of what happens in the *Odyssey's* Book 23, for instance, might help us explain why it corresponds to a relative maxima on the average arc of the text, but it doesn't tell us why that section of the Book is so much more negative in valence in Pope's translation than in Lattimore's or Lombardo's.

Making sense of these differences is possible, but would require the use of both

close-reading methods and a different computational tools than the one at use in this study. For instance, a close-reading approach to this question *the Odyssey* might hypothesize that the particularly brutal language used towards the slave women at the end of Book 22 in Fagles' and Lombardo's translations is responsible for the dip their arcs experience at that point in the story. In order to know for certain, though, we'd want to use an approach like Word Shifts (Dodds et al., 2011), which would allow us to directly compare two translator's versions of the same section of text and identify which words are the most significant contributors to the divergence in sentiment between the two. Word Shifts could also be used to investigate the peaks and valleys of any given translation's emotional arc by comparing two scenes within the same translation of different emotional valence. Because of the benefit these further analyses would provide to the interpretation of these analyses, our future plans for this work involve the use of Word Shift technology to address both these questions.

Chapter 4

Conclusion: The Invisible Waters of the Wine-Dark Sea

4.1 Summary

While digital approaches to the humanities have become increasingly accepted, computational work is still relegated to the fringes of the field. In recent years, strides have been made in the relatively nascent science of stories, and tools have been developed which have particular potential for classical studies. The purpose of this project was to experiment with these tools to see what, if anything, they could tell us about Homer in translation. We found that both allotaxonometry and hedonometry show promise for use in studies of classical texts in translation.

After putting the texts through a robust, multi-round standardization process, which itself revealed interesting differences between translations, we ran allotaxonometric comparisons for all possible combinations of the translations of *the Iliad* and *the Odyssey* in our sample. We classified our findings into nine areas, some of which

pertained to the general divergence between texts (e.g. computed divergence scores, the shape of data on allotaxonometric histograms), and others of which related to specific choices made by individual translators in our sample (the evidence of Alexander Pope's rhyme scheme, for instance, or matching word pairs which show how specific words in the text were translated differently). By integrating these findings with existing classical scholarship and traditional, close-reading methods of textual analysis, we were able to identify ways in which our findings aligned with and built on work in the field of Homer in translation.

Using hedonometry and the labMT dictionary, we identified and analyzed the emotional arcs of *the Iliad* and *the Odyssey*. By annotating the emotional arcs of these texts, we were able to investigate how well this method of sentiment analysis matched up with our preconceptions about the emotional valence of the texts. We also compared the emotional arcs of all of the translations of each text against one another to identify areas where there was the most sentimental disagreement between translators and identified areas for future study.

In sum, these findings indicate that there is a place for not only digital but computational methods in the classical field, particularly as pertains to the critical yet often overlooked work of translation.

4.2 Limitations and Scope

> "Diction, word choice, is important but not everything. Rhythms are at least as important, and crafting verse lines and sequences based on natural speech rhythms occupies more of my attention than any other aspect of

translation." (Lombardo, 2010, p. 228)

While word-based methods of computational analysis are undeniably of value to this field of inquiry, some of the most defining characteristics of Homeric epic cannot be captured using this approach. Additionally, though words are treated as individual units in the bulk of this analysis, no word is an island, and the connections between words influence their meaning in a way neither form of analysis used here can truly account for.

One thing that might stick out to readers of Homer in translation is that we do not at any point in this study prescribe any particular translation of the Homeric epics or even propose to evaluate the quality of the epics in our sample. While some might view this as a limitation of this project, it is in fact a question of scope.

A great deal of scholarship exists which evaluates the effectiveness of various translations of Homer. Theoretically, this would involve first defining what makes a translation of the epics successful, and subsequently evaluating translations based on this criteria. Unfortunately, there is agreement on the specifics of neither of these among classicists or non-Greek-reading audiences of Homer in translation.

Consider these two comments on Richmond Lattimore's 1951 translation of *the Iliad*, which is credited with a close adherence to the literal meaning of Homer's words. Lattimore receives praise for this from some quarters:

> "(Lattimore's Iliad) captured the majestic repose of Homer's verse with great fidelity. Lattimore is scrupulously faithful to the original ... He hews closely to Homer's text, keeping interpretation at a minimum. And except for an occasional archaism, his language is straightforward and unadorned. Lattimore's fluid, flexible six-beat line is especially apt for

> the quiet moments and lyric touches of the poem. He gets admirably close to Homer's voice," (Farrell, 2012)

and just as much censure from others:

> "(Lattimore) is quaint, prosaic, garrulous, low, and unidiomatic as well. What has happened to us in the twentieth century that translations such as this have appeared, sometimes to critical acclaim, and are used by students in serious academic courses? We have become so anxious to be true to Homer, so reluctant to introduce anything which is not in the original, that literalness has become a virtue." (Geddes, 1988, p. 11)

Literalness in a translation can be seen as a proxy for *fidelity to the original text*, a standard to which all translators are expected to aspire. But Geddes argues that Lombardo is literal in the wrong way, or that his pursuit of literalness in diction leads him astray in capturing more important Homeric qualities: "sincerity and honesty are not enough. Fine language (and poetry is language at its finest) needs artifice ... And if the original is poetry, then inarticulate plain-speaking is not more but less true to the original." (p. 12)

This is only one small example of the way translations of Homeric epic are compared and assessed, and mentioned only to illustrate why this was outside the scope of our work. Our aim in this study was not to make judgments on the quality or fidelity of the translations in our sample, but simply to identify differences between them using computational methods and in turn to understand and explain these differences. This is partly due to the unsuitability of the tools at our disposal to answer questions like those posed by Geddes, but also because we believe that the question,

"Which of these translations is best?" is fundamentally less interesting than "How are these translations different, why is it so, and why does it matter?"

4.3 Final Thoughts

The great writer David Foster Wallace opened his commencement address to the 2005 graduating class of Kenyon College with this joke:

> There are these two young fish swimming along and they happen to meet an older fish swimming the other way, who nods at them and says "Morning, boys. How's the water?" And the two young fish swim on for a bit, and then eventually one of them looks over at the other and goes "What the hell is water?" (Wallace, 2021)

The water of Wallace's joke is a metaphor for all the things we take for granted in society and in life. So many important things in our world go unnoticed on the day-to-day that it as if the scale and importance of a thing is inversely proportional to how much attention is paid to it. The Homeric epics are an example of this, to be sure; how many people who claim no knowledge of Homer nevertheless know the gist of the Trojan Horse story or what it means for something to be an "Achilles' heel?" If the Homer's influence is invisible, though, the practice of translation of Homer is doubly so.

Translators are often encouraged to make themselves invisible. Lawrence Venuti opens his book, "The Translator's Invisibility" (2008) with this quote from translator Norman Shapiro: "I see translation as the attempt to produce a text so transparent that it does not seem to be translated. A good translation is like a pane of glass. You

only notice that it's there when there are little imperfections—scratches, bubbles." (p. 1)

Venuti's thesis is that this largely British and American paradigm of translation, whereby translations must be invisible to be successful, must be reevaluated. He says,

> "A translated text ... is judged acceptable by most publishers, reviewers and readers when it reads fluently, when the absence of any linguistic or stylistic peculiarities makes it seem transparent, giving the appearance that it reflects the foreign writer's personality or intention or the essential meaning of the foreign text—the appearance, in other words, that the translation is not in fact a translation, but the 'original."' (Venuti, 2008, p. 1)

But sometimes, as in the case of the Homeric epics, composed for and disseminated for centuries in an oral culture millennia distant from our own, the foreign text *is* linguistically or stylistically peculiar. Thus, Venuti argues that the demands for invisibility—the demands that translators produce a maximally assimilated, "domesticated" version of the foreign text—paradoxically produce a version of the text through which the original is less visible.

Despite this, the illusion that a translation can and should be a pane of glass through which the original can be fully taken in persists, and it has lead to the work of the translator so being misunderstood and marginalized that a translator's work is sometimes erased altogether. A 1984 report on the state of translation in the publishing industry painted a dire picture:

> "Many newspapers, such as *The Los Angeles Times*, do not even list the translators in headnotes to reviews, reviewers often fail to mention that

> a book is a translation (while quoting from the text as though it were written in English), and publishers almost uniformly exclude translators from book covers and advertisements." (Christ, 1984)

One would hope that things have improved in the last 40 years, but as of March 4th, 2026, a search for an English-language translation of *the Odyssey* on Barnes & Nobel's website indicates that the translator's name appears on the cover of only two of the top ten results.

Two translations of the epics which always have the translator's name on the cover are Emily Wilson's. Wilson has cited Venuti when referencing her own philosophy of translation, and has even named "Be (Somewhat) Visible" as one of her guiding principles. Part of this visibility is her openness in discussing that which often gone unsaid in translation of the epics. In her epilogue for "Homer's Daughters," (2019) she lays out her principles for translation as a feminist classicist and reflects on the misogyny that has long proliferated in translations of Homeric epic. She ends with this:

> "The lesson to draw here is that a translator's ideological approach makes a real difference, and that the world of classics desperately needs more diligent, thoughtful, literary translators who are willing to be more critical of the power imbalances both of the classical world and of our own societies. ... Canonical texts such as *the Odyssey* are often read by impressionable young people—and, in fact, older people and scholars—not simply as historical documents, but as lessons in how society really works, and how human life has always been. If we present non-specialists with *Odysseys* that normalize the linguistic and physical abuse of women (for instance by

> glorifying the hanging of the 'slut' slaves), we are failing to tell the truth about the original text, and we may also be doing real ethical damage. Feminist classicists need to translate as well as communicate, with each other but also with a wider world. The dead white men, including Homer, are no longer the exclusive property of living white men." (pp. 296–297)

Wilson's message feels particularly urgent, now. In recent years, the appropriation of the classics by far-right movements has reached new heights. While not a new phenomenon—the book *Greeks, Romans, and Germans* details the Nazi's obsession with the classical societies (Chapoutot & Nybakken, 2016), for instance—our online world seems to be a particularly suitable breeding ground for these misappropriations, especially in misogynistic (Zuckerberg, 2018) and white nationalist (Dozier, 2026) circles.

Homer's epics have long been used to reinforce reactionary, conservative ideas of the world by people who believe the classical societies are ones where their values were universally upheld, and to which they have special claim. To a certain type of person, with a certain confused perception of antiquity, Ancient Greece and Rome represent a world where people like them (white, male) were in charge and flourishing, and their enemies (women, people of color) were subjugated and put in their place. It is because of this that women and people of color taking part in Homeric reception, whether by translating or studying the epics, writing adaptations, or playing Homeric characters in film and theater, are met with such remarkable vitriol from the far-right—the presence of women and people of color in classical spaces is seen as an encroachment on their fantasy world.

But, to reiterate Wilson's message, Homer is not the property of white nationalists

and misogynists on Twitter, and if the epics can be said to belong to any non-Greek person, they surely belong to all of us. Each of us can learn to recognize the echoes of Homer which constantly resound all around us. We will hear them better for understanding how their resonance has been shaped by translators, the rhapsodes of our modern times.

Bibliography

Alshaabi, T., Arnold, M. V., Minot, J. R., Adams, J. L., Dewhurst, D. R., Reagan, A. J., Muhamad, R., Danforth, C. M., & Dodds, P. S. (2021). How the world's collective attention is being paid to a pandemic: COVID-19 related n-gram time series for 24 languages on Twitter. *PLOS ONE*, *16*(1). https://doi.org/10.1371/journal.pone.0244476

Arnold, M. (1861). *On translating Homer: Three lectures given at Oxford*. London: Longman, Green, Longman, Roberts. https://www.gutenberg.org/files/65381/65381-h/65381-h.htm

Borges, J., & Levine, S. (1992). Some versions of Homer. *PMLA*, *107*(5), 1134–1138. https://doi.org/https://doi.org/10.2307/462868

Bradley, M. M., & Lang, P. J. (1999). *Affective norms for english words (ANEW): Instruction manual and affective ratings* (tech. rep.). Technical report C-1, The Center for REsearch in Psychophysiology, University of Florida.

Chapoutot, J., & Nybakken, R. R. (2016). *Greeks, romans, germans: How the nazis usurped europe's classical past*. University of California Press. Retrieved March 1, 2026, from http://www.jstor.org/stable/10.1525/j.ctt1f5g5m8

Christ, R. (1984). Translation watch. *PEN American Center Newsletter*, *53*.

Dodds, P. S., Minot, J. R., Arnold, M. V., Alshaabi, T., Adams, J. L., Dewhurst, D. R., Gray, T. J., Frank, M. R., Reagan, A. J., & Danforth, C. M. (2023). Allotaxonometry and rank-turbulence divergence: A universal instrument for comparing complex systems. *EPJ Data Science*, *12*(1). https://doi.org/https://doi.org/10.1140/epjds/s13688-023-00400-x

Dodds, P. S., Clark, E. M., Desu, S., Frank, M. R., Reagan, A. J., Williams, J. R., Mitchell, L., Harris, K. D., Kloumann, I. M., Bagrow, J. P., et al. (2015). Human language reveals a universal positivity bias. *Proceedings of the national academy of sciences*, *112*(8), 2389–2394.

Dodds, P. S., & Danforth, C. M. (2010). Measuring the happiness of large-scale written expression: Songs, blogs, and presidents. *Journal of happiness studies*, *11*(4), 441–456.

Dodds, P. S., Harris, K. D., Kloumann, I. M., Bliss, C. A., & Danforth, C. M. (2011). Temporal patterns of happiness and information in a global social network: Hedonometrics and twitter. *PloS one*, *6*(12), e26752.

Doherty, L. E. (1985). On teaching homer from translations. *The Classical Journal*, *81*(2), 161–166. Retrieved March 1, 2026, from http://www.jstor.org/stable/3296746

Dozier, C. (2026). *The white pedestal: How white nationalists use ancient greece and rome to justify hate*. Yale University Press.

Farley, S. (2016). Verions of Homer: Translation, fan fiction, and other transformative rewriting. *Transformative Works and Cultures*, *21*. https://doi.org/https://doi.org/10.3983/twc.2016.0673

Farrell, J. (2012). The English Iliad. *Los Angeles Review of Books*. https://lareviewofbooks.org/article/the-english-iliad/

Finkelberg, M. (2017). Homer at the Panathenaia: Some possible scenarios. In C.Tsagalis & A. Markantonatos (Eds.), *The winnowing oar—new perspectives in Homeric studies* (pp. 29–40). De Gruyter.

Geddes, A. G. (1988). Homer in translation. *Greece and Rome*, *35*(1), 1–13. Retrieved February 23, 2026, from http://www.jstor.org/stable/643274

Gothard, K., Dewhurst, D. R., Minot, J. R., Adams, J. L., Danforth, C. M., & Dodds, P. S. (2021). The incel lexicon: Deciphering the emergent cryptolect of a global misogynistic community. [arxiv:2105.12006 [cs]]. https://doi.org/10.48550/arXiv.2105.12006

Greenwood, E. (2020). Postcolonial perceptions of Homeric epic. In C. Pache, C. Dué, S. Lupack, & R. Lamberton (Eds.), *The Cambridge Guide to Homer* (pp. 532–535). Cambridge University Press. https://doi.org/10.1017/9781139225649

Hilden, N. (2023). Emily Wilson wants her "Iliad" to make you cry. *Esquire*. https://www.esquire.com/entertainment/books/a45300595/emily-wilson-iliad-interview/

Homer. (1900). *The Odyssey*. (S. Butler, Trans.). London: A. C. Fifield. (Original work published ca. 800 B.C.E.)

Homer. (1909a). *The Iliad*. (A. Pope, Trans.). Cassell. (First published in 1715. Original work published ca. 800 B.C.E.)

Homer. (1909b). *The Iliad*. (S. Butler, Trans.). Longmans, Green, Co.. (Original work published ca. 800 B.C.E.)

Homer. (1951). *The Iliad*. (R. Lattimore, Trans.). University of Chicago Press. (Original work published ca. 800 B.C.E.)

Homer. (1965). *The Odyssey*. (R. Lattimore, Trans.). Harper; Row. (Original work published ca. 800 B.C.E.)

Homer. (1990). *The Iliad*. (R. Fagles, Trans.). Penguin Books. (Original work published ca. 800 B.C.E.)

Homer. (1997). *The Iliad*. (S. Lombardo, Trans.). Hackett Publishing Company. (Original work published ca. 800 B.C.E.)
Homer. (1999). *The Odyssey*. (R. Fagles, Trans.). Penguin Books. (Original work published ca. 800 B.C.E.)
Homer. (2000). *The Odyssey*. (S. Lombardo, Trans.). Hackett Publishing Company. (Original work published ca. 800 B.C.E.)
Homer. (2002). *The Odyssey*. (A. Pope, Trans.). The Project Gutenberg. (First published in 1725. Original work published ca. 800 B.C.E.)
Homer. (2018). *The Odyssey*. (E. R. Wilson, Trans.). W. W. Norton; Company. (Original work published ca. 800 B.C.E.)
Homer. (2023). *The Iliad*. (E. R. Wilson, Trans.). W. W. Norton; Company. (Original work published ca. 800 B.C.E.)
Jenkins, T. E. (2020a). From the cinema to beyond: Homer in comics, television, apps, and new media. In C. Pache, C. Dué, S. Lupack, & R. Lamberton (Eds.), *The Cambridge Guide to Homer* (pp. 541–543). Cambridge University Press. https://doi.org/10.1017/9781139225649
Jenkins, T. E. (2020b). Homer in social thought. In C. Pache, C. Dué, S. Lupack, & R. Lamberton (Eds.), *The Cambridge Guide to Homer* (pp. 526–527). Cambridge University Press. https://doi.org/10.1017/9781139225649
Jenkins, T. E. (2020c). Introduction. In C. Pache, C. Dué, S. Lupack, & R. Lamberton (Eds.), *The Cambridge Guide to Homer* (pp. 517–518). Cambridge University Press. https://doi.org/10.1017/9781139225649
Lombardo, S. (2010). Translating the Iliad for a wider public. *Classical World*, *103*(2), 227–231. https://doi.org/https://doi.org/10.1353/clw.0.0177
Lucic, A., & Blake, C. (2011). Comparing the similarities and differences between two translations. *ADHO 2011-Stanford*. Retrieved November 3, 2025, from https://dh-abstracts.library.cmu.edu/works/1309
Oikonomaki, A. (2018). Achaioi, argeioi, danaoi: Revisiting the system of denomination of the Greeks in the Homeric epics. *The Center for Hellenic Studies at Harvard: Research Bulletin*. http://nrs.harvard.edu/urn-3:hlnc.essay:OikonomakiA.Achaioi_Argeioi_Danaoi.2018
Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. University of Illinois Press.
Pennebaker, J. W., Francis, M. E., Booth, R. J., et al. (2001). Linguistic inquiry and word count: Liwc 2001. *Erlbaum, Mahway*.
Reagan, A. J., Mitchell, L., Kiley, D., Danforth, C. M., & Dodds, P. S. (2016). The emotional arcs of stories are dominated by six basic shapes. *EPJ Data Science*, *5*(1). https://doi.org/https://doi.org/10.1140/epjds/s13688-016-0093-1
Reagan, A. J., Danforth, C. M., Tivnan, B., Williams, J. R., & Dodds, P. S. (2017). Sentiment analysis methods for understanding large-scale texts: A case for

using continuum-scored words and word shift graphs. *EPJ Data Science*, *6*(1), 28.

Russo, D. (2024). The representation of female characters in Butler's translation of the "Odyssey": A corpus-based approach. *AOQU (Achilles Orlando Quixote Ulysses)*, *5*(2), 235–258. https://doi.org/10.54103/2724-3346/27696

Ryabko, B., & Savina, N. Comparative method for quantitative assessment of the quality of translations. In: 2024 IEEE International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON). 2024. https://doi.org/10.1109/SIBIRCON63777.2024.10758508

Rybicki, J. (2006). Burrowing into translation: Character idiolects in Henryk Sienkiewicz's Trilogy and its two english translations. *Literary and Linguistic Computing*, *21*(1), 91–103. https://doi.org/10.1093/llc/fqh051

Stupinski, A. M., Alshaabi, T., Arnold, M. V., Adams, J. L., Minot, J. R., Price, M., Dodds, P. S., & Danforth, C. M. (2022). Quantifying changes in the language used around mental health on Twitter over 10 years: Observational study. *JMIR Mental Health*, *9*(3). https://doi.org/10.2196/33685

The Hanna Holborn Gray Special Collections Research Center. (2014). Homer in print: The transmission and reception of Homer's works. Chicago, IL. https://www.lib.uchicago.edu/collex/exhibits/homer-print-transmission-and-reception-homers-works/

Tsagalis, C. (2020a). Pisistratus. In C. Pache, C. Dué, S. Lupack, & R. Lamberton (Eds.), *The Cambridge Guide to Homer* (pp. 532–535). Cambridge University Press. https://doi.org/10.1017/9781139225649

Tsagalis, C. (2020b). Panathenaia. In C. Pache, C. Dué, S. Lupack, & R. Lamberton (Eds.), *The Cambridge Guide to Homer* (pp. 187–189). Cambridge University Press. https://doi.org/10.1017/9781139225649

Venuti, L. (2008). *The translator's invisibility: A history of translation* (2nd ed.). Routledge.

Vonnegut, K. (1995). Shapes of stories. https://www.youtube.com/watch?v=oP3c1h8v2ZQ

Wallace, D. (2021). This is Water by David Foster Wallace [Speech transcript]. *Farnam Street*. https://fs.blog/david-foster-wallace-this-is-water/

Whissell, C. (2004). ''the sound must seem an echo to the sense"': Pope's use of sound to convey meaning in his translation of homer's iliad. *Perceptual and motor skills*, *98*(3), 859–864.

Wikipedia. (2026). English translations of Homer — Wikipedia, the free encyclopedia. https://en.wikipedia.org/wiki/English_translations_of_Homer

Wilson, E. (2017). A translator's reckoning with the women of the Odyssey. *The New Yorker*. https://www.newyorker.com/books/page-turner/a-translators-reckoning-with-the-women-of-the-odyssey

Wilson, E. (2019). Epilogue: Translating Homer as a woman. In F.Cox & E. Theodorakopoulos (Eds.), *Homer's daughters: Womenâs responses to Homer in the twentieth century and beyond* (pp. 279–297). Oxford University Press.

Wilson, E. (2023). Exit Hector, again and again: How different translators reveal the 'Iliad' anew. *The New York Times*.

Zuckerberg, D. (2018). *Not all dead white men: Classics and misogyny in the digital age*. Harvard University Press. Retrieved March 1, 2026, from http://www.jstor.org/stable/j.ctv24w63tr