



Research Article

© 2025 Sonila Daiu and Blevis Allushi

This is an open access article licensed under the Creative Commons
Attribution-NonCommercial 4.0 International License
(<https://creativecommons.org/licenses/by-nc/4.0/>)

Analyzing English evolution through semantics, lexicology, syntax, context, using Machine Learning and Data Analysis

Dr Sonila Daiu

University Metropolitan Tirana, Albania

Blevis Allushi

University Metropolitan Tirana, Albania

DOI: <https://doi.org/10.2478/ejels-2025-0002>

Abstract

Whether a committed reader or not, it is clear that the English language has evolved throughout the years into many different nuances. What is considered acceptable in terms of written form has been adjusted to match cultural, regional, and on a macroscopic scale, temporal changes. This paper explores these changes through innovative analysis of semantics, lexicology, syntax, and context, (to be referred to as Linguistic Patterns), analyzing data and deriving conclusions. It is apparent that there are noticeable differences in the corpora of a 16th-century author to a contemporary one. The most accessible examples of such changes are spotted in written works, whether poetry, books, or documents, as they provide valuable insight into the analysis of change. The methodology incorporates pre-existing and specifically trained models specialized in data analysis to understand and compile these changes. This study showcases the evolution of the English language interpreted by Machine Learning (ML) models and methods such as Natural Language Understanding (NLU). By feeding data, specifically written works, into such models with the foresight of expecting a wide range of differing results and analyzing the changes through the scope of time, this study showcases the change of Linguistic Patterns. The decision between model preference and proficiency is made by comparing the quality of data outputs, and systematically evaluating different model archetypes, such as Generative Pre-training Transformers (GPT) or Bidirectional Encoder Representations from Transformers (BERT). The evaluation of changes in linguistic patterns is quantifiable through statistical measures, embeddings and syntactic parsing scores. Through these steps, this study derives that the English language has experienced a robust alteration in its core, from the elimination of now-considered archaic lexicology, differences in structural and contextual cues as well as notable evolution in semantics. These findings can be utilized in historical linguistic analysis and education, as well as improving Natural Language Understanding.

Keywords: English language evolution, Linguistic Patterns, Machine Learning models, Data Analysis, semantics, lexicology, syntax, Natural Language Understanding.

1. Introduction

The English language possesses the greatest number of non-native speakers (SIL International, 2024). Through English ideas can be expressed for readers of all nationalities to understand. English can be dubbed the most internationalized language, due to its robust nature and potential for ever-increasing globalization. The unification English has brought to the world is of concern to the socio-political faction of linguists. This study is concerned with the evolution of English, not its spread nor the reasons for it. In this paper the goal is to have a clearer grasp of the historical evolution of English.

1.1. The timeframe of English

The earliest varieties of a form of the English language, named Old English, or “Anglo-Saxon”, evolved from a group of Germanic dialects due to cultural exchanges. This marks the start of English as somewhere around the AD 500s (Short, 2002). The progression of English then extends to Middle English, which finds in genesis in the late AD 1000s, taking place after the Normal Conquest of England, which lead to a significant amount of Old French vocabulary being incorporated (Crystal, The Cambridge Encyclopedia of the English Language, 2003). Early Modern English can be dated to the AD 1400s. To summarize, English follows a timetable as shown below [Table 1].

Table 1 – Timeframe of English Language

Old English	AD 500 – AD 1000
Middle English	AD 1000 – AD 1400
Early Modern English	AD 1400 – AD 1600
Modern English	AD 1600 – present

1.2. Decision of studying Early Modern English to Modern English

Old English is not particularly comparable to Modern English, but rather a historical remnant valuable only to a niche research field. The alphabet has remained largely the same, with few changes. This is due to the fact that it is a direct adaptation of the Latin alphabet (Fortescue, 1714). The modern alphabet is largely the same, and that simplifies the analysis of text due to making the input easier to process, in regard to preprocessing of the data, however this is a detail that will be largely elaborated in the description of the methodology used throughout this study. Currently the main concern is establishing a clear boundary between what is to be utilized and what isn't.

1.3. How a language changes

Contact with other languages is a common alteration to endemic languages. The interactions between speakers of different languages can lead to borrowing of terms,

switching, of vocabularies, and the emergence of new expressions. An example of such changes is evident in the Norman Conquest of England, which took place in AD 1066. French largely influenced English and changed Old English to what we refer to as Middle English (Hock & Joseph, 2009). Social differentiation is an important factor to consider when considering how a language progresses. The sociolinguistic factors such as class, profession and geography to this day shape the type and level of verbosity (Gillard, 2018). An example of assimilation and dissimilation is The Great Vowel Shift of AD 1450 – AD 1700 (Liberman, 2003). Syncope and apocope refer to the loss of sounds within a word, such as the omission of certain vowels or syllables. Once again, these simplifications, or rather barbarisms, can bleed into standard written form (McMahon, 2012).

1.4. How and why machine learning can be utilized

Machine Learning is a field of study within artificial intelligence. It is specifically concerned with the development and study of statistical algorithms. These algorithms can learn from data and extrapolate new information and visualize unseen data (IBM, 2021).

2. Methodology

This study adopts an experimental methodology utilizing collected data and analyzing it through transformer architecture models, known to excel in showing details a human analysis is either unable to show or not efficient. Text analysis through machine learning is revolutionary in the extraction of insight, which will greatly increase the precision and accuracy of the outputs. The research is being approached by collecting data on written works, preprocessing and cleaning the data, analyzing through models, getting feedback and repeating necessary steps to achieve better results, and in the end organizing the data into digestible, statistically appropriate and logical manners. The structure of such a process is coined by the term Machine Learning Pipeline (IBM, 2024) (Williams, 2024) (Dinohund, 2023).

2.1. Research design

An experimental approach has been taken to have empirical data based on numbers. The transformers in LLMs are neural networks, and these neural networks attempt to understand and generate text, hence granting us Natural Language Understanding, and Natural Language Generation, which lead us to Natural Language Processing. All LLMs are NLUs, but not all NLUs are LLMs (IBM, 2024). In this study's methodology in order to mitigate errors in processing, attention to clarity of prompts and parallel repetition will be utilized to gather the most unbiased and accurate data possible.

2.2. Data collection methods

Different regions write differently, however since the oldest and most studied geographical, historical, and socio-political documentation can be found in the British Isles, this region will be the one considered by this study regarding data collection

(Yorke, 2006). Since there are differences in writing documents, poetry, and prose, this study accounts for all three. By studying them separately, it is aimed to find independent differences and patterns, referred to as Linguistic Patterns, which have been defined as differences and patterns in semantics, lexicology, syntax and context.

2.2.1 1600s

Chosen document: *Leviathan* (1651) by Thomas Hobbes (Hobbes, 1651).

Chosen poetry: *Paradise Lost* (1667) by John Milton (Milton, 1667).

Chosen prose: *The Anatomy of Melancholy* (1621) by Robert Burton (Burton, 1621).

2.2.2 1700s

Chosen document: *The Wealth of Nations* (1776) by Adam Smith (Smith, 1776).

Chosen poetry: *The Rape of the Lock* (1712) by Alexander Pope (Pope, 1712).

Chosen prose: *A Modest Proposal* (1729) by Jonathan Swift (Swift, 1729).

2.2.3 1800s

Chosen document: *On Liberty* (1859) by John Stuart Mill (Mill, 1859).

Chosen poetry: *Poems* by Alfred Tennyson, Volume I (1842) by Alfred Lord Tennyson (Tennyson, 1842).

Chosen prose: *Wuthering Heights* (1847) by Emily Brontë (Brontë, 1847).

2.2.4 1900s

Chosen document: *Why I write* (1946) by George Orwell (Orwell, 1946).

Chosen poetry: *The Highwayman* (1906) by Alfred Noyes (Noyes, 1906).

Chosen prose: *A Portrait of the Artist as a Young Man* (1916) by James Joyce (Joyce, 1916).

2.2.5 2000s

Chosen document: *The Creative Commons* (2001) by Lawrence Lessig (Lessig, 2004).

Chosen poetry: *A Visit from St. Nicholas* (2024) by Clement Clarke Moore (Moore, 2024).

Chosen prose: *The Cathedral and the Bazaar* (2000) by Eric S. Raymond (Raymond, 2000).

2.3. Tools and techniques

The two main tools to be utilized are ChatGPT and Gemini. The main idea is to utilize pre-existing LLMs and train them through prompting on understanding semantics, lexicology, syntax, and context, also referred to as Linguistic Patterns.

2.4. Procedure

The chosen experimental approach begins with defining the categories of Linguistic Patterns. Semantics, lexicology, syntax and context. The mother science of them is Linguistics. The previous process gives statistical, and articulable Linguistic Patterns for each analysis it completes. The methodology for data analysis is a practical one,

of experimentation and documentation, and of course, iteration.

3. Results

The compilation process of the aforementioned works in their respective categories has yielded valuable insight worthy of showcase and discussion, through the computation of the data provided by the algorithms and models mentioned, to get the results which are to be showcased.

3.1. The analysis and raw data

The output provided by the models, averaged to each other's data to find the most accurate representation. First and foremost a table is provided with the initial number of words in texts analyzed per period. The table can be seen below [Table 2].

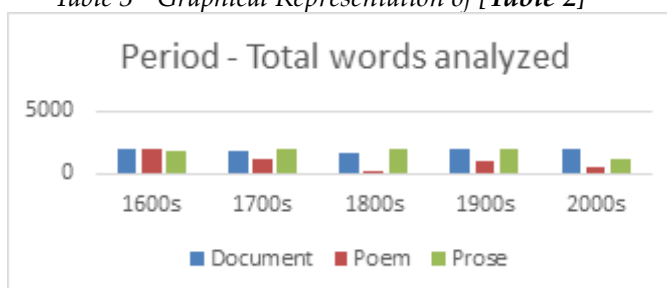
Table 2 - Period to Total words analyzed

Period	Document words analyzed	Poem words analyzed	Prose words analyzed
1600s	1977	1938	1907
1700s	1853	1130	1974
1800s	1634	222	1913
1900s	1993	1004	2058
2000s	1999	540	1222

3.2. Analysis of documents

As can be seen above, the number of words analyzed is consistent, and the only inconsistencies are in the poem category regarding the AD 1800s and the AD 2000s. The numerical values appear valid and worthy of a visual representation. [Table 3]

Table 3 - Graphical Representation of [Table 2]



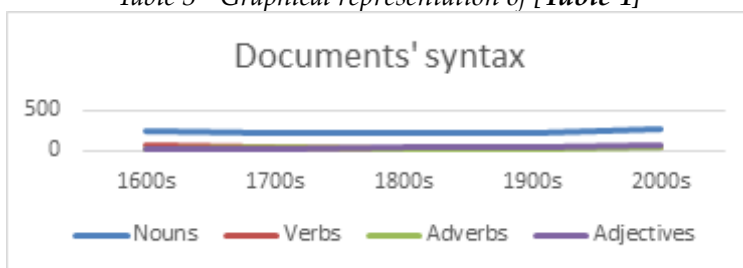
Shown below are tables of the document analysis regarding a semantically meaningful showcase of the usage of nouns, verbs, adverbs and adjectives [Table 4].

Table 4 - Documents' syntax

Period	Nouns	Verbs	Adverbs	Adjectives
1600s	251	80	27	17
1700s	219	32	35	25
1800s	221	29	27	43
1900s	208	46	28	52
2000s	283	40	31	60

A noticeable consistency can be seen in the number of nouns and adverbs used, however when looking at verbs it appears there was a peak usage of verbs in the AD 1600s followed by a dip that remained consistent. [Table 5].

Table 5 - Graphical representation of [Table 4]



An analysis of the most used words in documents is also provided by the experiment and is shown in the table below. The axiomatic expression that “the” is the most used word in the English language has been proven (although not for the first time). The most interesting insight is that “the” has remained the most used word throughout all of the specified timeframe, closely followed by “i”. [Table 6].

Table 6 – Documents' most used words & number of rare words

Period	Most Common Words	Least Common Words (Count)
1600s	the (146), of (87), and (73), in (56), is (41), a (38), that (38), which (28), to (28), as (28)	464
1700s	the (158), of (145), in (68), a (47), and (43), to (41), is (35), are (20), it (17), but (17)	363
1800s	the (135), of (97), to (57), and (45), a (33), in (31), it (24), as (23), was (23), or (18)	501
1900s	the (97), i (79), of (65), and (63), to (55), a (52), in (47), was (23), my (22), that (20)	672
2000s	the (134), of (79), that (58), in (54), a (54), to (44), and (43), is (36), it (28), was (27)	508

3.3. Analysis of poems

Following the same logic as done with documents, expressing similar tables and values, the data appears as such [Table 7].

Table 7 - Poems' syntax

Period	Nouns	Verbs		Adverbs	Adjectives
1600s	217	39		9	37
1700s	173	24		7	23
1800s	27	3		0	3
1900s	83	31		3	1
2000s	61	10		5	4

It is apparent that due to the lesser number of words available for the AD 1800s and AD 2000s there is not much misrepresentation, and a fair consistency along with the other periods of time, proportionally speaking. The number of verbs, adverbs and adjectives is low. While the use of nouns is dominant, however consistently decreasing. Going onto the most common words and number of rare words used, we can refer to the next table [Table 8].

Table 8 - Poems' most used words & number of rare words

Period	Most Common Words	Least Common Words (Count)
1600s	and (84), the (81), of (57), to (44), his (40), in (36), that (33), with (32), from (23), or (20)	744
1700s	the (67), and (49), of (26), to (25), in (21), a (19), her (16), their (15), with (12), what (11)	575
1800s	the (11), of (10), her (8), and (7), your (5), to (4), you (3), that (3), a (3), as (3)	145
1900s	the (98), a (31), and (31), he (24), her (24), in (23), of (21), his (21), to (18), they (14)	259
2000s	the (32), a (29), and (28), his (21), of (16), to (16), he (13), in (11), i (10), like (8)	218

3.4 Analysis of prose

Similarly done to documents and poetry we can continue the same process of showing results for prose [Table 9].

Table 9 - Proses' syntax

Period	Nouns	Verbs	Adverbs	Adjectives
1600s	273	17	12	31
1700s	203	39	34	38

1800s	196	71	45	35
1900s	223	37	7	13
2000s	61	10	5	4

There is a clear consistency in nouns, and a seemingly pyramidal presentation of verbs and adverbs, while a consistency was apparent in adjectives there is has been a dip in use throughout the past two hundred years. Word usage is presented below [Table 10].

Table 10 - Proseses' most used words & number of rare words

Period	Most Common Words	Least Common Words (Count)
1600s	and (101), of (78), the (75), in (41), to (41), that (28), his (23), all (22), a (21), is (20)	753
1700s	the (98), of (67), and (64), to (60), a (58), in (44), i (32), be (29), for (27), that (27)	615
1800s	the (97), of (73), and (64), a (63), i (56), to (47), my (28), in (25), his (19), with (18)	872
1900s	the (153), and (100), his (54), a (53), he (50), was (47), to (44), had (40), in (34), of (31)	491
2000s	the (50), to (47), i (41), a (32), and (24), was (22), of (19), for (18), it (18), that (15)	442

There was more diversity in the lexicology of AD 1600s – AD 1900s, with a noticeable dip throughout AD 1900s – AD 2000s.

3.5 Linguistic Patterns

The numerical analysis shows clear insight, but the slightly more introspective analysis of Linguistic Patterns is expected to show a more detailed overview into the similarities and differences across time periods.

3.5.1 Semantic Patterns

In the AD 1600s there is generally philosophical and abstract writing, even in official documents. Terms like “sense”, “fancy”, and imagination are used heavily. Words carried layered meaning, rooted in the metaphysical or philosophy. Documents assume educated audiences familiar with philosophy.

The AD 1700s progressed to be more analytical and technical. The focus is on economic and practical ideas with terms like “labor,” and “division” being used frequently. The language is more empirical. This reflects the Enlightenment period and the emphasis on reason and observation.

A political and ideological structure can be observed in the AD 1800s. The focus is on liberty and authority, there are terms such as “liberty” and “democracy” which are central to repetition. The semantic field is broad and reflects evolving political and societal structures.

AD 1900 manifest text that appears more personal and reflective, with a focus on individual experiences. The use of “I” is especially common. There is a blend of personal and societal narratives.

In the current era, AD 2000s, text is informative and digital, and discusses technology, creativity and intellectual property. Technical terminology paralleled to the age of computers dominates.

3.5.2 Lexical Patterns

AD 1600s express through archaic words such as “vertue” and “divers” are common. Compound words such as “counter-pressure”, and the capitalization of nouns seem characteristic in older English.

Throughout the AD 1700s the vocabulary is made simpler and daily terms such as “division”, “labor” are used instead of their counterparts. This reflects an increase in interest in economic discourse.

The 1800s have a mix of abstract and concrete terms. The vocabulary is that of political discussion and social debates. Terms such as “liberty”, “oppression”, and “authority” are common.

Conversation is narrative driven in the AD 1900s, explaining ideas more vividly and imaginatively. Words are modern, however formal. Examples include “egoism”, and “aesthetic”.

Contemporary vocabulary of the AD 2000s is simpler, and much more precise. Technical terms are more viable and used, with terminology such as “copyright”, and “digital consumer” finding more use.

3.5.3 Syntactic Patterns

Syntax of AD 1600s is one of long, complex sentences composed of multiple clauses. Inversions in sentences are common, an example is “Him the Almighty Power hurld headlong flaming”. Sentence structures are elaborate and formal, especially resembling philosophical and theological patterns.

AD 1700s show more balance in structure, often with symmetrical clauses. The sentences are constructed logically and there is more use of rhetorical questions, such as, “Say what strange motive, Goddess!”.

The syntax in the AD 1800s blends narrative flow and detailed descriptions. Sentences are varied in length, with some long and descriptive and others shorter and more impactful. The style is formal and structured, yet more conversational than the previous. There is utilization of detailed imagery in descriptions, for example descriptions such as “stormy weather” or “gaunt thorns” are much more common than previously.

AD 1900s syntax is fragmented and rawer. Modernist trends can be observed, especially in the shorter sentences and occasional, much rarer, longer sentences.

Repetition is more common, and there is extensive use of dialogue and monologue. Contemporary AD 2000s syntax is clear. Short to medium-length sentences are preferred. Technical jargon is more common, and structure finds itself standardized. Syntax is tailored to be as practical and informative as possible.

3.5.4 Contextual Cues

Context of the AD 1600s is deeply theological and philosophical. Texts assume familiarity with Christianity and Greco-Roman literature. References to biblical events are common, and a scholarly tone is often prevalent. Context is provided through morality and metaphysical questions.

The AD 1700s emphasize the Enlightenment, especially reason and critique. Socio-economic issues are addressed, and context is provided through analytical tones with notes of humor and irony, appearing satirical and notably logical.

The AD 1800s consist of political revolutions and multiple societal transformations. Introspection is the focus, and contextual cues are dependent on authority, social responsibility and liberty. The Gothics and Romantics influenced descriptive nature and emotivity.

The context of the AD 1900s is individualistic and introspective – a reflection of modernism. Psychological depth and psychoanalysis evolve through personal experiences. There are contemporary cultural references apparent, which serve to provide context to the abundance of themes regarding identity and alienation.

AD 2000s are shaped by the digital age. Texts assume familiarity with technical concepts and especially reference online culture. Themes of practicality and innovation are often the desire of contextual cues as far as attention is concerned. There is pragmatism, and efficiency and functionality are necessary to address problems.

4. Discussion

Arguably most of the discussion has been made through the [Results], however it is necessary to compile information and discuss implications and relevance.

4.1. Linguistic Patterns across periods

The findings reveal evolution in the Linguistic Patterns throughout the specified periods. From the predominance of theological and philosophical themes to the change of the AD 2000s which emphasize clarity and practicality and efficient communication. The switch from complex and archaic syntax and lexicology to a much simpler and digestible methodology of writing is a testament to the practical nature of humans, and more importantly, adaptability to contemporary problems and concerns. To each period there are societal priorities which are reflected in the Linguistic Patterns.

5. Recommendations

A suggestion to further researchers is to compare and focus the findings of this paper in terms of geographical and historical analysis. Although history is partially accounted for in explaining changes in Linguistic Patterns throughout this study, along with geographical conditions it is not in the scope of this methodology, and frankly, this paper.

6. Limitations

While artificial intelligence can serve as a powerful tool in the right hands, it is not perfect. Models rely heavily on their training, and since archaic dialects are not particularly abundant regarding model training, despite the best efforts that can be made to counteract the misunderstanding of linguistic diversity.

The analysis of hundreds of thousands of words and millions of characters is computationally expensive and extremely time consuming, even when powered by the best models available. The main bottleneck of AI, as has been stated and cited many times over, is the computational power required.

7. Conclusion

This study has been a comprehensive journey through the analysis of Linguistic Patterns throughout five centuries, more specifically AD 1600s – AD 2000s. The analysis has involved documents, poetry and prose. Examination consists of semantics, lexicology, syntax, and context. Through the examination of the trends, patterns, similarities and differences have emerged. The evolution of the English Language was analyzed in a linguistic context through utilizing Machine Learning algorithms, specifically LLMs, and the technique of NLU. The analysis through machine learning models has enabled this study to compute large amounts of data and find correlations, statistical data and details that are critical to the understanding of how the English language has progressed. Innovative and important technology has been utilized throughout this study, showing the practicality and convenience of developing artificial intelligence and pushing boundaries.

In this study the process of data collection and preprocessing has been elaborated upon, alongside the ethical implications of copyright law.

This paper revealed gradual simplification of syntax and vocabulary. Showing a diminishing lexical diversity throughout time.

This paper proved the axiomatic expression of the most common word in the English language ("the"), while also documenting common use of other words such as "and", "of". Through such a finding it is established that the English language does have a set schematic as far as structure has been concerned for the past five centuries.

In this study the rise in use of the personal pronoun "I" was evidenced in the AD 1900s, marking the shift toward individualistic ideologies and themes. The adaptability of English along with the mirroring of intellectual and cultural priorities

has been illustrated throughout this paper. It has been shown that there was a gradual shift from theological and philosophical themes to more practical themes. This study labelled the use of nouns, verbs, adverbs and adjectives throughout the specified periods. A clear description of each era's specific semantic patterns, lexical patterns, syntactic patterns and contextual cues has been provided, amounting to the culminant definition of Linguistic Patterns.

Finally, this paper discussed the importance of studying Linguistic Patterns and provided suggestions to researchers continuing this paper's work or developing new papers.

8. Generative AI Statement

"As the authors of this work, we used ChatGPT and Gemini, two AI-supported technologies to compile bulk data and extrapolate statistical elements. After using the aforementioned tools, an evaluation and review were conducted on the final product. As the authors of this paper, we take absolute responsibility for all content in our published work."

References

- Brontë, E. (1847). *Wuthering Heights*. Retrieved 12 25, 2024, from Project Gutenberg: <https://www.gutenberg.org/cache/epub/768/pg768-images.html>
- Burton, R. (1621). *THE ANATOMY OF MELANCHOLY*. Retrieved 12 25, 2024, from Project Gutenberg: <https://www.gutenberg.org/cache/epub/10800/pg10800-images.html>
- Crystal, D. (2003). *The Cambridge Encyclopedia of the English Language*. Cambridge University Press.
- Dinohund. (2023, October 18). *Machine Learning Pipeline in Production*. Retrieved from wikimedia: https://commons.wikimedia.org/wiki/File:Machine_Learning_Pipeline_in_Production.png#/media/File:Machine_Learning_Pipeline_in_Production.png
- dvdogoy. (2021, June 6). *dl-visuals*. Retrieved from github: <https://github.com/dvgodoy/dl-visuals/?tab=readme-ov-file>
- Fortescue, J. (1714). *The Difference between an Absolute and Limited Monarchy*. London: John Fortescue Aland.
- Gillard, D. (2018). *Education in the UK: A history*. Retrieved from Education in the UK: <https://education-uk.org/history/chapter01.html>
- Hobbes, T. (1651). *LEVIATHAN*. Retrieved 12 25, 2024, from Project Gutenberg.
- Hock, H. H., & Joseph, B. D. (2009). *Language History, Language Change, and Language Relationship: An Introduction to Historical and Comparative Linguistics*. Walter de Gruyter.
- IBM. (2021, September 22). *What is machine learning?* Retrieved from IBM: <https://www.ibm.com/think/topics/machine-learning>
- IBM. (2024, 12 19). *What is a machine learning pipeline?* Retrieved from IBM: <https://www.ibm.com/think/topics/machine-learning-pipeline#:~:text=A%20machine%20learning%20pipeline%20is,and%20deploying%20machine%20learning%20models>
- IBM. (2024, December 19). *What is a neural network?* Retrieved from IBM: <https://www.ibm.com/think/topics/neural-networks>
- Joyce, J. (1916). *A Portrait of the Artist as a Young Man*. Retrieved 12 25, 2024, from Project Gutenberg: <https://www.gutenberg.org/cache/epub/4217/pg4217-images.html>

- Lessig, L. (2004). *The Creative Commons*. Retrieved 12 25, 2024, from Montana Law Review: <https://scholarworks.umt.edu/mlr/vol65/iss1/1/>
- Lieberman, M. (2003). *Linguistics 001 Lecture 22 Language Change*. Retrieved from University of Pennsylvania: https://www.ling.upenn.edu/courses/Fall_2003/ling001/language_change.html
- McMahon, A. M. (2012). *Understanding Language Change*. Cambridge University Press.
- Mill, J. S. (1859). *On Liberty*. Retrieved 12 25, 2024, from Project Gutenberg: <https://www.gutenberg.org/cache/epub/34901/pg34901-images.html>
- Milton, J. (1667). *PARADISE LOST*. Retrieved 12 25, 2024, from Project Gutenberg: <https://www.gutenberg.org/cache/epub/20/pg20-images.html>
- Moore, C. c. (2024). *A Visit from St. Nicholas*. Retrieved 12 25, 2024, from Poetry Foundation: <https://www.poetryfoundation.org/poems/43171/a-visit-from-st-nicholas>
- Noyes, A. (1906). *The Highwayman*. Retrieved 12 25, 2024, from Poetry Foundation: <https://www.poetryfoundation.org/poems/43187/the-highwayman>
- Orwell, G. (1946). *Why I write*. Retrieved 12 25, 2024, from The Orwell Foundation: <https://www.orwellfoundation.com/the-orwell-foundation/orwell/essays-and-other-works/why-i-write/>
- Pope, A. (1712). *The Rape of the Lock*. Retrieved 12 25, 2024, from Project Gutenberg: <https://www.gutenberg.org/cache/epub/9800/pg9800-images.html>
- Raymond, E. S. (2000). *The Cathedral and the Bazaar*. Retrieved 12 25, 2024, from The Cathedral and the Bazaar: <http://www.catb.org/esr/writings/homesteading/cathedral-bazaar/>
- Short, I. (2002). Language and Literature". A Companion to the Anglo-Norman World. In I. Short, *Language and Literature". A Companion to the Anglo-Norman World* (pp. 191-214). Boydell and Brewer Limited.
- SIL International. (2024). *What is the most spoken language?* Retrieved from Ethnologue: <https://www.ethnologue.com/insights/most-spoken-language/>
- Smith, A. (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations*. Retrieved 12 25, 2024, from Project Gutenberg: <https://www.gutenberg.org/cache/epub/3300/pg3300-images.html>
- Swift, D. J. (1729). *A Modest Proposal*. Retrieved 12 25, 2024, from Project Gutenberg: <https://www.gutenberg.org/cache/epub/1080/pg1080-images.html>
- Tennyson, B. A. (1842). *The Early Poems of Alfred, Lord Tennyson*. (J. C. Collins, Ed.) Retrieved 12 25, 2024, from Project Gutenberg: <https://www.gutenberg.org/cache/epub/8601/pg8601-images.html>
- Williams, B. (2024, 7 12). *How to Use AI for Text Analysis in Research*. Retrieved from Insight7: <https://insight7.io/how-to-use-ai-for-text-analysis-in-research/>
- Yorke, B. (2006). The Conversion of Britain: Religion, Politics and Society in Britain c.600–800. In B. Yorke, *The Conversion of Britain: Religion, Politics and Society in Britain c.600–800* (p. 54). Pearson Education.