

AI-Driven Intralingual Translation across Historical Varieties

Theoretical Frameworks and Examples from Early Modern English

Fabio Ciambella

Sapienza University of Rome

ORCID: <https://orcid.org/0000-0002-6761-2031>

Email: fabio.ciambella@uniroma1.it

Keywords

Early Modern English

Literary texts

Specialized texts

Intralingual translation

Modernization

AI

Abstract

This article examines how Artificial Intelligence (AI) can transform intralingual translation, with a focus on historical varieties, particularly Early Modern English. It explores the foundations of AI-based translation, addressing both the challenges and opportunities of applying this technology to linguistic change over time. Drawing from models in linguistics, natural language processing, and historical linguistics, the study investigates the modernization of Early Modern English, using literary and specialized texts to demonstrate AI's effectiveness in handling complex vocabulary and syntax. The paper also evaluates AI's impact on fields like historical linguistics and digital humanities, discussing both the benefits and the limitations, such as the risk of anachronism and the need for human oversight. Additionally, it considers how AI-driven translation can contribute to the digitization and accessibility of historical texts, broadening access to linguistic resources and enhancing appreciation for language evolution.

1. Introduction: Early Modern English and Artificial Intelligence

It is well-known that navigating the linguistic currents of Early Modern English poses significant challenges due to the rapid evolution of the language over time. With the advent of Artificial Intelligence (AI), scholars have gained powerful tools to delve deeper into this linguistic realm (see Zheng et al. 2023; Massey et al. 2022; Quah 2006), employing innovative approaches to normalizing spelling,¹ unraveling authorship mysteries, etc.

Surely, one of the best-known examples of AI applied to Early Modern English is the VARD2 software, which will be employed in this article. Developed by Paul Rayson and Dawn Archer

¹ In line with Domingo and Casacuberta (2023), which to my knowledge is one of the few studies devoted to machine learning techniques and historical documents, in this article I distinguish between spelling normalization and language modernization as two distinct processes, one concerning only the Early Modern English spelling and the other working with language as a whole, meaning also modernizing morphosyntactic and lexical structures.

at Lancaster University, VARD2 employs computational algorithms to automatically normalize the spelling of historical texts (see Baron 2008 for further details). Through pattern recognition and linguistic rules, VARD2 maps archaic spellings to their contemporary counterparts, facilitating smoother comprehension and analysis of Early Modern English texts. VARD2 helps users overcome myriad challenges posed by language evolution's fluidity. Each text presents its lexical labyrinth, demanding adaptability and precision from this AI tool. Despite its prowess, VARD2 remains a tool in progress, evolving with each encounter to enhance its linguistic understanding (see, among others, Marquilhas and Hendrickx 2014).

Even ChatGPT can be used to normalize Early Modern English texts, although it may not be as specifically tailored for this task as tools like VARD2. In the first phase of this study, the results provided by VARD2 and ChatGPT concerning spelling normalization will be compared and discussed, while in the second phase, only ChatGPT is employed for language modernization, since VARD2 cannot perform this task. ChatGPT, developed by OpenAI, is a state-of-the-art language generation model trained on a vast amount of text data, including a wide range of historical and contemporary literature. When provided with Early Modern English text as input, ChatGPT leverages its understanding of language patterns, syntax, and semantics to generate output that closely resembles contemporary English. While its primary function is not spelling normalization, ChatGPT can infer modern equivalents for archaic spellings and grammatical structures based on its training data. This happens because of semantic understanding, contextual learning, and language generation. First of all, ChatGPT can comprehend the meaning of the text and infer the intended message despite the archaic language forms. It can identify words and phrases with modern equivalents and generate output that conveys the same message in contemporary language. Then, through its training on a diverse range of texts, ChatGPT has learned contextual relationships between words and phrases. This allows it to accurately interpret the context of Early Modern English passages and generate modernized versions that maintain coherence and relevance. Finally, ChatGPT's natural language generation capabilities enable it to produce fluent and grammatically correct text. It can generate modernized versions of Early Modern English passages by replacing archaic spellings and structures with their contemporary counterparts while preserving the overall meaning and style. While ChatGPT can perform reasonably well in modernizing Early Modern English texts, it may encounter challenges with highly specialized vocabulary or obscure linguistic features unique to that period. Additionally, the quality of modernization may vary depending on the complexity and ambiguity of the original text. Nonetheless, ChatGPT's versatility and adaptability make it a valuable tool for linguistic research and text transformation tasks, including the normalization and modernization of historical texts.

Beyond spelling normalization, authorship attribution is another point of contact between Early Modern English texts and AI. Here, projects like Hugh Craig's Intelligent Archive beckon, promising insights into the authorship of Early Modern English texts. Hugh Craig, founder and director of the Centre for Literary and Linguistic Computing (CLLC)² at the University of Newcastle, Australia, devoted most of his scholarly career to statistical analyses connected with the authorship of Early Modern English texts. Through the Intelligent Archive, Craig harnesses AI algorithms to scrutinize textual patterns, unraveling the threads of authorial identity woven into the fabric of historical texts. The Intelligent Archive serves as a beacon of innovation, illuminating authorship attribution with computational precision. By analyzing linguistic features such as vocabulary, syntax, and style, Craig's AI navigator discerns subtle nuances indicative of individual authorial voices (see Craig and Greatley-Hirsch 2017; Craig and Kinney 2010 for further details).

Machine learning techniques such as Graig's Intelligent Archive are also used to detect unidentified printers within the realm of Early Modern English printed books. For instance, Vogler et al.'s (2023) focus lies on correlating distinctively damaged character-type imprints found in anonymously printed books with known printer attributions, thereby furnishing insights into their provenance. Traditionally, such investigations have relied on manual scrutiny by analytical bibliographers. Vogler et al. introduce a contrastive attention-based metric learning strategy for discerning similar damage patterns across character image pairs. This approach proves adept at detecting even the subtlest discrepancies in glyph shapes while remaining resilient to assorted sources of noise inherent in digitized historical texts. To address the paucity of annotated data, Vogler et al. devise a randomized data synthesis method aimed at replicating bends, fractures, and ink variations characteristic of early printing techniques. The group of scholars' methodology notably enhances the matching of damaged type-imprints across printed works from this era, a conclusion supported by expert evaluation within the field. As often happens in the field of Early Modern English, most of the AI-based devices concern William Shakespeare and his works, and specifically the capacity of AI tools to be creative enough to emulate the playwright's style, thus raising issues of creativity connected with the use of AI.³ Three AI writing software are of particular interest in this brief overview of the best-known AI tools devoted to Shakespeare, that is Deep-speare, ShakespeareGPT and AI Shakespeare.⁴

² See <https://www.newcastle.edu.au/research/centre/cllc> for further details. All websites last visited 12/12/2024.

³ For matters concerning AI and creativity, see, among others, Ivcevic and Grandinetti 2024; Vinchon et al. 2023; van Heerden and Bas 2021; Boden 1998; Dartnall 1994.

⁴ I discovered them while browsing the website *There's an AI for that*, which is dedicated to recognizing AI-based software, platforms, or websites tailored to users' specific needs (<https://theresanaiforthat.com>).

Deep-speare is a project developed by Jue Wang, who worked with a team at MIT, including Piji Li, Zhao Song, and Hsin-Yu Ha (see van Heerden and Bas 2021; Lau et al. 2020; 2018). The project aimed to create a deep-learning model capable of generating Shakespearean-style sonnets. The model was trained on a dataset of Shakespeare's sonnets and utilized a recurrent neural network (RNN) architecture. Deep-speare used a combination of two types of neural networks: an encoder-decoder architecture and a generative adversarial network (GAN). The encoder-decoder model was responsible for generating the initial sequences of text, while the GAN helped refine the output to ensure it closely resembled Shakespearean style and language. The model's training involved exposing it to a large corpus of Shakespeare's work, allowing it to learn the patterns, structures, and vocabulary typical of Shakespearean sonnets. Once trained, Deep-speare could generate new sonnets in a style reminiscent of Shakespeare, incorporating themes, language, and poetic devices characteristic of his writing. While Deep-speare represents an interesting application of deep learning in creative writing, it is essential to note that the quality of the generated sonnets may vary, and they may not always match the complexity and depth of Shakespeare's original work. However, the project demonstrates the potential of AI in creative endeavors, showcasing how machine-learning techniques can be applied to generate human-like artistic content.

ShakespeareGPT⁵ is a specialized version of the Generative Pre-training Transformer (GPT), aimed to transform contemporary English texts into the Early Modern English language, specifically in the style of Shakespearean English. Leveraging the mechanisms of GPT, this tool seamlessly converts everyday English input into the more elaborate and antiquated style of Shakespeare's literary masterpieces. Although access to its features requires a ChatGPT Plus subscription, the basic, free version of ShakespeareGPT offers a unique opportunity for users to immerse themselves in and deepen their understanding of Shakespearean language. Upon activation, users are greeted with a friendly message inviting them to input their text for translation. Once initiated, users simply enter their text into the tool, which swiftly generates a rendition in Shakespearean English. Whether for educational pursuits or leisurely exploration, ShakespeareGPT provides an engaging platform for exploring the distinct linguistic rhythm of Shakespeare, nurturing a heightened appreciation for his unparalleled contribution to the English language.

Similarly to ShakespeareGPT, but unfortunately no longer available, AI Shakespeare,⁶ developed by DAISYS.ai, was a groundbreaking platform leveraging artificial intelligence to swiftly craft entertaining, fully narrated Shakespearean plays. With the aid of AI technology,

⁵ Available at <https://flowgpt.com/p/shakespearegpt>.

⁶ Former URL: https://ai-shakespeare.ai/?ref=taaft&utm_source=taaft&utm_medium=referral.

the platform supplied users with remarkably human-like voices, imbued with emotion and purpose, thereby ensuring captivating recordings. Users were afforded the flexibility to provide prompts across a spectrum of themes, ranging from vampire dentists to time-traveling goldfishes and houses haunted by OCD ghosts. Subsequently, the platform generated an entirely original Shakespearean play, meticulously crafted through algorithmic prowess, devoid of any human intervention. Nevertheless, it is imperative to acknowledge that while the AI model's output promised creativity, it occasionally deviated from user expectations or historical accuracy. Thus, the end products were primarily intended for entertainment and inspiration, rather than professional counsel or direction. By engaging with AI Shakespeare by DAISYS.ai, users implicitly assumed responsibility for any claims or disputes arising from the platform's output or performance. DAISYS.ai explicitly disclaimed liability for any dissatisfaction, offense, or harm resulting from the utilization of its service. With all its limits, AI Shakespeare by DAISYS.ai stood as a testament to the transformative potential of AI within the creative arts sphere. It showcased how technology can revolutionize conventional entertainment mediums, offering users the means to craft unique theatrical experiences sans the need for professional actors or writers. This innovative platform exemplified how AI can breathe new life into traditional forms of entertainment, fostering immersive audio experiences enriched by dynamic, lifelike voices.

2. Methods and materials

The research conducted in this essay is mainly divided into two phases. The first phase aims to compare and assess the effectiveness of two intralingual translation tools, VARD2 and ChatGPT,⁷ in normalizing the spelling of both literary and non-literary texts belonging to Early Modern English. I use a small DIY corpus of literary and non-literary texts that I am familiar with from previous research in various fields. These texts are representative of the period's genres and span a significant portion of Early Modern English. Specifically, I will consider three literary texts (a sonnet by Richard Barnfield, a monologue from a comedy by William Shakespeare, and one from Fletcher's tragedy *Bonduca*), and three non-literary texts (an ophthalmic treatise by Walter Bailey, a passage from a treatise about warfare by Thomas and Dudley Digges, and a paragraph from a treatise about Restoration dietary habits by Thomas Tryon). As previously hinted at, the number of tokens and span of time of the texts considered is as representative and balanced as possible, as shown in Table 1:

⁷ The version of ChatGPT used for this research is ChatGPT-4. For the sake of readability, it will not be explicitly mentioned each time it appears in the text of this essay.

Literary texts	N. of tokens	Year
Richard Barnfield – Final sonnet, from <i>The Affectionate Shepherd</i>	112	1594
William Shakespeare – from <i>The Taming of the Shrew</i> , Folio text, vv. 2694-2737	365	1623
John Fletcher – from <i>Bonduca</i> , 1.1.1-17	138	1647
Total n. of tokens	615	

Non-literary texts	N. of tokens	
Walter Bailey – <i>A Treatise on Eyesight</i> , from MS V.a.140 at the Folger Shakespeare Library	316	1586
Thomas and Dudley Digges – from <i>Four Paradoxes or Political Discourses</i> , A2<r>, ll. 72-85	116	1604
Thomas Tryon – <i>What Planet and Constellation Governs each Instrument of Music</i> , from <i>The Long Way to Health and Long Life</i> , chapter XXI, pp. 487-488	217	1683
Total n. of tokens	649	

Tab. 1: Description of the corpus

The selection of texts for the DIY corpus was deliberate, aiming to encompass both literary and non-literary genres and discourses belonging to the late 16th and 17th centuries. The three literary texts selected belong to different genres and cover more than 50 years. Barnfield’s sonnet is a love poem that closes the poet’s well-known *The Affectionate Shepherd*, while Katherine’s final monologue is taken from Shakespeare’s comedy, *The Taming of the Shrew*. Finally, Bonduca’s monologue is taken from the very first lines of Fletcher’s eponymous tragedy. Regarding non-literary texts, spanning almost 100 years, Bailey’s *Treatise on Eyesight* is included in a manuscript held at the Folger Shakespeare Library, from the collection of Early Modern English recipe books. The extract from Thomas and Dudley Digges’ *Four Paradoxes* discusses officers and different ranks in the English army, while the passage from Thomas Tryon’s *The Way to Health and Long Life* concerns musical instruments, their characteristics, and their astronomical connections. Despite being small in terms of tokens, this corpus aims to be balanced and representative of the Early Modern English language from the end of the 1500s to the second half of the 1600s.

As already hinted at in the introduction, VARD2 serves as a primary intralingual translation tool designed explicitly for normalizing historical texts. It operates on rule-based algorithms, identifying and correcting spelling variations to align with contemporary norms. In contrast, ChatGPT offers a more flexible approach to intralingual translation. Although not specifically designed for spelling normalization, ChatGPT’s natural language processing capabilities enable it to generate text that closely resembles human language, making it a suitable candidate for this study.

In the second phase, I just used ChatGPT not only to normalize the spelling of the texts but also to create an out-and-out intralingual translation of the texts where morphosyntactic and

lexical traits are modernized. To this aim, after ChatGPT generated the full modernization of the corpus selected, I interviewed the platform itself (see the Appendix) to understand what criteria ChatGPT uses when modernizing Early Modern English texts. From the interview, it emerges that the criteria adopted concern five aspects/principles:

1. Lexis and semantics: ChatGPT's task involves updating the language and vocabulary to align with contemporary usage, while retaining the original meaning and intention of the text.
2. Morphology and syntax: Necessary adjustments to the grammar and syntax are made to adhere to present-day norms, while carefully preserving the structure and fluency of the original text.
3. Clarity and readability: The goal is to ensure that the updated text is lucid and easily comprehensible to modern readers, steering clear of outdated expressions or constructions that might impede understanding.
4. Cultural context: Any cultural references or contexts that may require clarification or modernization for present-day audiences are considered.
5. Respect for the original: Throughout the process of modernization, the platform upholds the integrity of the original text and the author's style, ensuring that the essence and tone of the initial work remain intact (other criteria or specifications can be found in the Appendix).

The procedures for this research involve at least three steps (pre-editing, machine intralingual translation, and post-editing) that mirror the new tasks of the human translator in the era of AI-based translations. Professional translators, when not involved directly in the translation process which is entrusted to machines, are responsible for the pre- and post-editing phases, as affirmed in the literature (see, among others, Mattioda et al. 2023; Guerberof Arenas 2020; Kliffer 2008). Hence, in the pre-editing phase, I preprocess the selected texts to ensure consistency and compatibility with the intralingual translation tools. This may include formatting adjustments and minor corrections⁸ to enhance the accuracy of the translation process. Once prepared, the texts undergo spelling normalization using both VARD2 and ChatGPT, and language modernization using ChatGPT only. Each tool is tasked with normalizing the spelling of the original texts while maintaining coherence, style, and semantic fidelity. Evaluation of the intralingual translation process is based on several criteria. First of all, the accuracy of the normalized texts in aligning with contemporary spelling conventions is assessed. Additionally, in the second phase, the coherence of the modernized texts is evaluated,

⁸ In the examples provided below, however, some minor spelling typos in the texts taken from EEBO have been intentionally left unchanged. This is to underscore the importance of pre-editing operations; otherwise, the software, particularly VARD2, may not process the text correctly.

considering readability and overall linguistic flow. Style preservation is also a crucial aspect of the evaluation process, examining the ability of ChatGPT to retain the stylistic nuances and linguistic features of the original works. Finally, I also consider semantic fidelity, ensuring that the modernized texts retain the original meaning and intent of the source material.

While my study offers valuable insights into the intralingual translation process, it is essential to acknowledge its scope and limitations from the very beginning. The small size of the DIY corpus may limit the generalizability of some findings. Additionally, the intralingual translation tools employed in this study may encounter challenges in accurately modernizing certain linguistic features or stylistic elements unique to Early Modern English. Such challenges will be the object of discussion in the post-editing step.

Therefore, the first phase of this study utilizes a combination of VARD2 and ChatGPT to assess the effectiveness of intralingual translation in spelling normalization from Early Modern English to contemporary English, while in the second phase, the effectiveness of ChatGPT in language modernization will be assessed. By analyzing both literary and non-literary texts, the aim is to provide a comprehensive evaluation of the intralingual translation process and its implications for historical linguistics and computational linguistics.

3. Examples from literary and technical texts

3.1 VARD2 vs ChatGPT: Challenges concerning Early Modern English spelling normalization

In this section, I will introduce examples from the selected corpus to highlight both similarities and differences in spelling normalization performed by VARD2 and ChatGPT. The first two case studies presented are Barnfield's final (Shakespearean) sonnet from *The Affectionate Shepherd* (see Table 2), and Bailey's *Treatise on Eyesight*, from MS V.a.140 at the Folger Shakespeare Library. Both texts offer interesting considerations about spelling normalization practices. It is important to note that VARD2 was not pre-configured for normalization; default settings for automatic normalization were utilized in terms of confidence score (50%) and f-score weight (1).⁹

⁹ To define confidence score, I will quote from Baron and Rayson (2008), "any word not found in the tool's modern lexicon is marked as a potential variant, a list of candidate modern equivalents ranked by 'confidence' is produced for each potential variant and is presented to the user for consideration. [...] [T]he user can also provide a threshold 'confidence,' which is the minimum score the candidate must reach for it to be used. By using this feature with a relatively high threshold, the user can automatically replace most common variant forms, thereby saving a substantial amount of time processing the text." VARD2 is configured with a default threshold score of 50%. This means it automatically normalizes variants that have been previously normalized to a corresponding term in contemporary English in more than half of the occurrences stored in its translation memory. F-weight score, instead, is defined in the VARD2 user guide available online at <https://ucrel.lancs.ac.uk/ward/userguide/>: "To calculate overall confidence scores for methods and replacements, an f-score is calculated by combining the precision and recall scores. Usually, equal weight is given to precision and recall (F-Score weight 1), but a user may give priority to either precision or recall by altering this weight using the slider in the toolbar. Moving the slider either towards precision or recall will bias all F-Scores in VARD accordingly. Weights under 1 will bias towards

This decision was made to ensure parity between VARD2 and ChatGPT: the corpus was input into VARD2 with default settings, exactly as it was copied and pasted into ChatGPT with no specific criteria requested, but simply the prompt “Please normalize the spelling of the following texts.”

Original text	
<p>Loe here behold these tributarie Teares, Paid to thy faire, but cruell tyrant Eyes; Loe here the blossome of my youthfull yeares, Nipt with the fresh of thy Wraths winter, dyes, Here on Loues Altar I doo offer vp This burning hart for my Soules sacrifice; Here I receaue this deadly-poysned Cu, Of Circe charm'd; wherein deepe Magickelyes. Then Teares (if you be happie Teares indeed), And Hart (if thou be lodged in his brest), And Cup (if thou canst helpe despaire with speed); Teares, Hart, and Cup conioyne to make me blest: Teares moue, Hart win, Cup cause, ruth, loue, desire, In word, in deed, by moane by zeale, by fire.</p>	
VARD2	ChatGPT
<p>Lo here behold these tributary Tears, Paid to <u>thy</u> fair, but cruel tyrant Eyes; Lo here the blossom of my youthful years, Nipped with the <u>fresh</u> of <u>thy</u> <u>Wraths</u> winter, <u>dyes</u>, Here on <u>Loves</u> Altar I doo offer up This burning <u>hart</u> for my <u>Souls</u> sacrifice; Here I receive this deadly-poysned <u>Cue</u>, Of <u>Circa</u> charmed; wherein deep <u>Musicales</u>. Then Tears (if you be happy Tears indeed), And <u>Hart</u> (if thou be lodged in his breast), And Cup (if <u>thou</u> can help despair with speed); Tears, <u>Hart</u>, and Cup <u>conjoin</u> to make me <u>blessed</u>: Tears move, <u>Hart</u> win, Cup cause, ruth, love, desire, In word, in deed, by moan by zeal, by fire.</p>	<p>Lo, here behold these tributary Tears, Paid to <u>your</u> fair, but cruel tyrant Eyes; Lo, here the blossom of my youthful years, Nipped with the <u>frost</u> of <u>your</u> <u>Wrath's</u> winter, <u>dies</u>. Here on <u>Love's</u> Altar I do offer up This burning <u>heart</u> for my <u>Soul's</u> sacrifice; Here I receive this deadly-poisoned <u>Cup</u>, Of <u>Circe</u> charmed; wherein deep <u>Magic lies</u>. Then Tears (if you be happy Tears indeed), And <u>Heart</u> (if you be lodged in his breast), And Cup (if <u>you</u> can help despair with speed); Tears, <u>Heart</u>, and Cup <u>combine</u> to make me <u>blest</u>: Tears move, <u>Heart</u> win, Cup cause, ruth, love, desire, In word, in deed, by moan by zeal, by fire.</p>

Tab. 2: Spelling normalization of a sonnet by Barnfield with both VARD2 and ChatGPT

precision, weights over 1 will bias towards recall.” The default f-score setting is 1, balanced between precision (in Statistics: the ratio between the number of correct predictions of an event (class) to the total number of times the model predicts it) and recall (measurement of the model’s sensitivity. It is the ratio between the correct predictions for a class and the total number of cases where it occurs).

Original text	
<p>A breif Treatise towchinge the preservacion of the eyesight Consistinge partely in good order of dyet & partely in vse of medycynes Occasioned thorough certayne speches had with somme of myne honorable good frendes to wryte myne opinion of means to preserue the sighte in good integrity I haue performed the same in this litle pamphlett wherein I haue dyrected my penn rather to leave rules for those which haue not in them selves suffycient knowledge then to satisfy the lerned who I know canne devise nuche better means and remydyes: And my intent beinge onelye to sett downe somme order as may be fytest for all sortes of men to contynue in perfecion there sight I thought not good to inserte anye discourse of the dyseases of the eyes which will require a longe treatis & cannot well be executed but by men which haue skyl in the arte of phesicke: therefore it semed mete vnto me at this tyme to omytt that parte and to leave the reader for the knowledge and Cure of suche diseases to the professors thereof by whose helpe they may receyve remedye of all infirmytyes and affectes, which shall happen to the eyes: purposinge in this litle trestis onelye to declare how the sight might be contynued in his integritye, by suche Common order as maye of all men in their trade of lyef be resonably followed leavinge all other harder matters to a farther dyrection of the lerned And now followinge the laudable custome begunne in anncyent tyme & contynued in theis our dayes of presentinge our frendes with new yeres gyftes for lacke of other thinges of greater pryce I offer vnto you this litle pamphlet as a signification of my good will wishinge that the same maye be a token of many good & prosperous yeres which god the gyver of all goodnes sende vnto you.</p>	
VARD2	ChatGPT
<p>A brief Treatise <u>touching</u> the <u>preservacion</u> of the eyesight Consisting partly in good order of diet & partly in use of medicines Occasioned thorough certain speeches had with <u>sum</u> of mine <u>honourable</u> good friends to write mine opinion of means to preserve the sight in good integrity I have performed the same in this little pamphlet wherein I have directed my <u>penn</u> rather to leave rules for those which have not in <u>them selves</u> sufficient knowledge then to satisfy the learned who I know <u>canne</u> devise much better means and <u>remydyes</u>: And my intent being only to set down <u>sum</u> order as may be fittest for all sorts of men to continue in <u>perfecion</u> there sight I thought not good to insert any discourse of the diseases of the eyes which will require a <u>lunge treatis</u> & cannot well be executed but by men which have skill in the art of physic: therefore it seemed <u>mete</u> unto me at this time to <u>omytt</u> that part and to leave the reader for the knowledge and Cure of such diseases to the professors thereof by whose help they may receive remedy of all infirmities and affects, which shall happen to the eyes: purposing in this little <u>trestis</u> only to declare how the sight might be continued in his integrity, by such Common order as may of all men in their trade of <u>lief</u> be reasonably followed leaving all other harder matters to a farther direction of the learned And now following the laudable custom begun in ancient time & continued in their our days of presenting our friends with new <u>years</u> gifts for lack of other things of greater price I offer unto you this little pamphlet as a signification of my <u>good will</u> wishing that the same may be a token of many good & prosperous years which <u>god</u> the giver of all goodness send unto you.</p>	<p>A brief Treatise <u>touching</u> the <u>preservation</u> of the eyesight Consisting partly in good order of diet & partly in use of medicines Occasioned through certain speeches had with <u>some</u> of my <u>honorable</u> good friends to write my opinion of means to preserve the sight in good integrity I have performed the same in this little pamphlet wherein I have directed my <u>pen</u> rather to leave rules for those who have not in <u>themselves</u> sufficient knowledge than to satisfy the learned who I know <u>can</u> devise much better means and <u>remedies</u>: And my intent being only to set down <u>some</u> order as may be fittest for all sorts of men to continue in <u>perfection</u> their sight I thought not good to insert any discourse of the diseases of the eyes which will require a <u>long treatise</u> & cannot well be executed but by men who have skill in the art of physic: therefore it seemed <u>meet</u> unto me at this time to <u>omit</u> that part and to leave the reader for the knowledge and Cure of such diseases to the professors thereof by whose help they may receive remedy of all infirmities and affects, which shall happen to the eyes: purposing in this little <u>treatise</u> only to declare how the sight might be continued in its integrity, by such Common order as may of all men in their trade of <u>life</u> be reasonably followed leaving all other harder matters to a farther direction of the learned And now following the laudable custom begun in ancient time & continued in these our days of presenting our friends with new <u>year's</u> gifts for lack of other things of greater price I offer unto you this little pamphlet as a signification of my <u>goodwill</u> wishing that the same may be a token of many good & prosperous years which <u>God</u> the giver of all goodness send unto you.</p>

Tab. 3: Spelling normalization of Bailey’s treatise with both VARD2 and ChatGPT

The tables above show both strengths and weaknesses of the two normalizations. First of all, it is evident that, from a shallow reading of the two versions, ChatGPT generally performs better than VARD2 in both literary and non-literary texts. Just to mention a few examples, ChatGPT always recognizes the inflection -s as an old form of the Saxon genitive, thus transforming it into the clitic modern form 's, while VARD2 does not seem to recognize it, at least with default settings:

Original texts: Loues Altar new yeres gyftes
 VARD2: Loves Altar new years gifts
 ChatGPT: Love's Altar new year's gifts

Moreover, VARD2 seems to struggle with proper names, especially those of non-English origin. This weakness of the software is evident in Table 2 (Circa/Circe) but is also confirmed by Table 4. Most of the issues with VARD2 stem from its default threshold confidence score, set at 50%. This means that only variants normalized with a specific term for over half of the total occurrences in the software's memory are automatically normalized. Part of the problem can be solved by lowering the threshold score and shifting the f-score weight towards precision, but even in this case, some issues persist (see Tables 4 and 5):

Original text	
THE hardy Romanes? O ye gods of Britain, the rust of Arms, the blushing shame of souldiers; are these the men that conquer by inheritance? the Fortune-makers? these the Julians, that with the Sun measure the end of Nature, making the world but one Rome and one Caesar? Shame, how they flee! Caesars soft soul dwells in 'em; their mothers got 'em sleeping, pleasure nurst 'em, their bodies sweat with sweet oils, loves allurements, not lustie Arms. Dare they send these to seek us, these Romane Girls? Is Britain grown so wanton? Twice we have beat 'em, Nennius, scatter'd 'em, and through their big-bon'd Germans, on whose Pikes the honour of their actions sit in triumph, made Themes for songs to shame 'em, and a woman, a woman beat 'em, Nennius; a weak woman, a woman beat these Romanes.	
VARD2 (threshold confidence score: 50%; f-score weight: 1; default settings)	VARD2 (threshold confidence score: 0%; f-score weight: 1/5)
THE hardy Romans? O ye gods of Britain, the rust of Arms, the blushing shame of soldiers; are these the men that conquer by inheritance? the Fortune-makers? these the <u>Julians</u> , that with the Sun measure the end of Nature, making the world but one Rome and one Caesar? Shame, how they flee! <u>Caesars</u> soft soul dwells in 'em; their mothers got ' <u>em</u> sleeping, pleasure nursed ' <u>em</u> , their bodies sweat with sweet oils, loves allurements, not lusty Arms. Dare they send these to seek us, these Roman Girls? Is Britain grown so wanton? Twice we have beat ' <u>em</u> , <u>Nennius</u> , scattered ' <u>em</u> , and through their <u>big-bon'd</u> Germans, on whose Pikes the honour of their actions sit in triumph, made Themes for songs to shame ' <u>em</u> , and a woman, a woman <u>beat</u> ' <u>em</u> , <u>Nennius</u> ; a weak woman, a woman <u>beat</u> these Romans.	THE hardy Romans? O ye gods of Britain, the rust of Arms, the blushing shame of soldiers; are these the men that conquer by inheritance? the Fortune-makers? these the <u>Juliennes</u> , that with the Sun measure the end of Nature, making the world but one Rome and one Caesar? Shame, how they flee! <u>Caesuras</u> soft soul dwells in <u>em</u> ; their mothers got <u>em</u> sleeping, pleasure nursed <u>em</u> , their bodies sweat with sweet oils, loves allurements, not lusty Arms. Dare they send these to seek us, these Roman Girls? Is Britain grown so wanton? Twice we have beat <u>em</u> , <u>Ninnies</u> , scattered <u>em</u> , and through their <u>big-boned</u> Germans, on whose Pikes the honour of their actions sit in triumph, made Themes for songs to shame <u>em</u> , and a woman, a woman <u>beaten</u> , <u>Ninnies</u> ; a weak woman, a woman <u>beat</u> these Romans.

Tab. 4: Spelling normalization from Fletcher's *Bonduca* with VARD2 (different threshold levels)

Original text	
<p>Drums are under the dominion of Saturn and Mars, in the Signs Capricorn and Aries, which their dull, heavy, melancholly, rattling, jarring Sounds do manifest; for as they arise from the dark wrathful Centre in Nature; so they carry with them the power of the same Properties, and therefore do encourage Wrath and Violence, the pleasant Element of Air, which is the Life of all Harmony, is here encompassed or penned up so close, that it cannot have its free Egress and Regress, which suffocates the thin spirituous Vapours thereof, and that causeth Sounds, heavy dull and frightful. The Air being the pleasant Life in all things, and the original of all Sounds, Voices and Harmony; and in what Things, Instrument or Creature soever the Properties of Saturn and Mars are strong, and the sweet Influences of this Element obstructed, from that Thing or Creature can proceed no delightful Sounds or M^elody, but dull, heavy, dolorous Jarrings, Noises or Sounds; for there is no true Life, Light, Harmony, or pleasant Sound can arise from any, in which the Properties of Nature are unequal in their qualifying or operation; for then the wrath of Nature becomes fierce and raging, and destroys the Moderator or friendly quality, and so Nature can no longer continue in equal weight and measure.</p>	
VARD2 (threshold confidence score: 50%; f-score weigh: 1; default settings)	VARD2 (threshold confidence score: 0%; f-score weight; 5)
<p>Drums are under the dominion of Saturn and Mars, in the Signs Capricorn and Aries, which their dull, heavy, melancholy, rattling, jarring Sounds do manifest; for as they arise from the dark wrathful Centre in Nature; so they carry with them the power of the same Properties, and therefore do encourage Wrath and Violence, the pleasant Element of Air, which is the Life of all Harmony, is here encompassed or penned up so close, that it cannot have its free Egress and Regress, which suffocates the thin spirituous Vapours thereof, and that causes Sounds, heavy dull and frightful. The Air being the pleasant Life in all things, and the original of all Sounds, Voices and Harmony; and in what Things, Instrument or Creature <u>soever</u> the Properties of Saturn and Mars are strong, and the sweet Influences of this Element obstructed, from that Thing or Creature can proceed no delightful Sounds or <u>M^elody</u>, but dull, heavy, dolorous <u>Jarrings</u>, Noises or Sounds; for there is no true Life, Light, Harmony, or pleasant Sound can arise from any, in which the Properties of Nature are unequal in their qualifying or operation; for then the wrath of Nature becomes fierce and raging, and destroys the Moderator or friendly quality, and so Nature can no longer continue in equal weight and measure.</p>	<p>Drums are under the dominion of Saturn and Mars, in the Signs Capricorn and Aries, which their dull, heavy, melancholy, rattling, jarring Sounds do manifest; for as they arise from the dark wrathful Centre in Nature; so they carry with them the power of the same Properties, and therefore do encourage Wrath and Violence, the pleasant Element of Air, which is the Life of all Harmony, is here encompassed or penned up so close, that it cannot have its free Egress and Regress, which suffocates the thin spirituous Vapours thereof, and that causes Sounds, heavy dull and frightful. The Air being the pleasant Life in all things, and the original of all Sounds, Voices and Harmony; and in what Things, Instrument or Creature <u>sever</u> the Properties of Saturn and Mars are strong, and the sweet Influences of this Element obstructed, from that Thing or Creature can proceed no delightful Sounds or <u>M^elody</u>, but dull, heavy, dolorous <u>Jarring</u>, Noises or Sounds; for there is no true Life, Light, Harmony, or pleasant Sound can arise from any, in which the Properties of Nature are unequal in their qualifying or operation; for then the wrath of Nature becomes fierce and raging, and destroys the Moderator or friendly quality, and so Nature can no longer continue in equal weight and measure.</p>

Tab. 5: Spelling normalization from Tryon’s *The Way to Health and Long Life* with VARD2 (different threshold levels)

Even by lowering the threshold confidence score for automatic normalization, some issues persist, as shown in Tables 4 and 5. This is due to the software’s over-normalization of some variants (e.g., *Julians* as *Juliennes*, *Caesars* as *Caesuras*, *soever* as *sever*, *Jarrings* as *Jarring*). In some cases, VARD2 does not manage to normalize the variant, since it is not present in its database, thus manual normalization must be carried out (e.g., *M^elody*, just because of the acute

accent of the letter *e*). On the contrary, ChatGPT, being a neural system which considers also the lexical and semantic network between different words, manages to fulfill better spelling normalizations. Nevertheless, some problems arise in the examples from the corpus. First of all, ChatGPT tends to normalize the Early Modern English second-person singular pronoun *thou* with the modern form *you*, thus eliminating the very important distinction between *thou* and *you* in 16th- and 17th-century English. Moreover, the fact that ChatGPT also considers lexical and semantic networks which statistically occur together, it sometimes over-normalizes words that do not need to be normalized. In the example from Barnfield's poem in Table 2, *fresh* is normalized with *frost* due to the lexical neighborhood of the word, which includes lexical items like *nipped* and *winter* that are generally more semantically connected with *frost* than *fresh*. However, such connections, common in technical texts, cannot be assumed in literary texts with their frequent use of figures of speech, such as oxymorons and synaesthesias. Finally, when asked to normalize the spelling of Early Modern English texts, ChatGPT shows the tendency to use American, unless otherwise instructed (e.g., in Table 3, the adjective *honourable* is normalized as *honorable*).

Last, but not least, ChatGPT also appears to overcome issues related to typos and mistakes found in the digitized source text, which VARD2 cannot do. For instance, typos such as *cu* instead of *cup* and *magicklyes* instead of *magicke lyes* are normalized as *cue* and *musicales* by VARD2, demonstrating its inability to grasp the semantic network between words. In contrast, ChatGPT successfully addresses such issues and normalizes them as *cup* and *magic lies*. Similarly, in Table 4, VARD2 normalizes *beat'em* as *beaten* because the two words are not written separately in the digitized version of the text taken from EEBO TCP. This demonstrates the importance of pre-editing operations on the source text, as hinted at in the previous section.

3.2 ChatGPT and the modernization of Early Modern English texts

The modernization of morphosyntactic and lexicosemantic features of Early Modern English texts constitutes a form of intralingual translation. However, modernizing 16th- and 17th-century texts raises age-old questions that extend beyond the scope of this essay. Here, my focus is on understanding the type of text modernization ChatGPT can perform when provided with either literary or non-literary texts from the Early Modern English period, the criteria it follows, and whether it produces a text that is more understandable for contemporary readership. Studies in the field of historical linguistics have demonstrated that the primary challenges of Early Modern English texts arise from antiquated grammar, particularly syntax, and vocabulary, as noted by Crystal (2012, 10-15). This includes “overcomplicated sentence structure to say simple things,” as succinctly summarized by one student, along with lexical and semantic complexities such as archaisms, infrequently used items in modern English, colloquial

language, malapropisms, culturally contemporary references, and false friends (Murphy et al. 2020, 303; 312).

When prompted with the inquiry, “Can you modernize this text?”, it appears that the modernization performed by ChatGPT leads to overly drastic alterations in syntactic structures, lexical items, and semantic nuances and connotations – not to mention prosodic features, alliterations, etc. in texts in verses – as evidenced by the findings presented in Tables 6 and 7. These tables illustrate the extent of the changes made during the modernization process, highlighting a notable discrepancy between the original text and its modernized rendition. Such radical transformations may raise concerns regarding the adherence and coherence of the modernized output compared to the intended modernization goal.

Original text	ChatGPT
<p>Fie, fie, vnknit that thretaring vnkinde brow, And dart not sornefull glances from those eies, To wound thy Lord, thy King, thy Gouverneur. It blots thy beautie, as frosts doe bite the Meads, Confounds thy fame, as whirle winds shake faire budds, And in no sence is meete or amiable. A woman mou'd, is like a fountaine troubled, Muddie, ill seeming, thicke, bereft of beautie, And while it is so, none so dry or thirstie Will daigne to sip, or touch one drop of it. Thy husband is thy Lord, thy life, thy keeper, Thy head, thy soueraigne: One that cares for thee, And for thy maintenance. Commits his body To painfull labour, both by sea and land: To watch the night in stormes, the day in cold, Whil'st thou ly'st warme at home, secure and safe, And craues no other tribute at thy hands, But loue, faire lookes, and true obedience; Too little payment for so great a debt. Such dutie as the subiect owes the Prince, Euen such a woman oweth to her husband: And when she is froward, peeuish, sullen, sowre, And not obedient to his honest will, What is she but a foule contending Rebell, And gracelesse Traitor to her louing Lord? I am asham'd that women are so simple, To offer warre, where they should kneele for peace: Or seeke for rule, supremacie, and sway, When they are bound to serue, loue, and obey. Why are our bodies soft, and weake, and smooth, Vnapt to toyle and trouble in the world, But that our soft conditions, and our harts, Should well agree with our externall parts? Come, come, you froward and vnable wormes, My minde hath bin as bigge as one of yours, My heart as great, my reason haplie more, To bandie word for word, and frowne for frowne; But now I see our Launces are but strawes: Our strength as weake, our weaknesse past compare, That seeming to be most, which we indeed least are. Then vale your stomackes, for it is no boote, And place your hands below your husbands foote: In token of which dutie, if he please, My hand is readie, may it do him ease.</p>	<p>Come on, now, relax that threatening frown, And don't shoot those scornful looks at me, To hurt your partner, your leader, your guide. It ruins your beauty, like frost on flowers, Ruins your reputation, like storms on buds, And it's not at all acceptable or attractive. A woman in a rage is like a troubled fountain, Muddy, unattractive, thick, devoid of beauty, And while she's like that, nobody, no matter how thirsty, Will want to drink from her. Your husband is your partner, your life, your support, Your leader, your king: Someone who cares for you, And for your well-being. He puts his body Through hard work, both on sea and land: To watch through storms at night, and endure cold days, While you're warm at home, safe and sound, And asks for nothing more than your love, your kindness, and your obedience; Too small a payment for such a big debt. Just as a subject owes duty to the ruler, So does a woman owe to her husband And when she's stubborn, irritable, sulky, bitter, And doesn't follow his honest wishes, What is she but a rebellious troublemaker, And disloyal traitor to her loving husband? I'm ashamed that women are so foolish, To seek conflict, when they should seek peace: Or strive for power, control, and dominance, When they're meant to serve, love, and obey. Why are our bodies soft, weak, and smooth, Not suited for toil and trouble in the world, If not so that our gentle natures and hearts, Should match our outward appearance? Come now, you stubborn and incapable creatures, My mind has been as strong as yours, My heart as great, my intellect perhaps greater, To argue back and forth, and scowl in return; But now I see our arguments are futile: Our strength is weak, our weaknesses incomparable, What seems strongest is actually weakest. So lower your pride, because it's pointless, And show your submission to your husband As a sign of this duty, if he wishes, My hand is here, ready to bring him comfort.</p>

Tab. 6: ChatGPT's modernization of the selected monologue from Shakespeare's *The Taming of the Shrew*

Original text	ChatGPT
<p>Firft, (in the choife of his officers) to get, or accept fuch Free-booters and Theeues, as (onely to haue the name and priuiledge of a fouldier, to escape the paine due by Martiall lawe to fuch vnfouldierlike perfons) will ferue without pay, or with halfe pay. Then, euerie of thefe his officers Lieutenaunt, Enfigne, Serieants &c. (being men of that Crewe) will draw in as many alfo as they can of the fame Moulde, to liue on pickorie without pay, and therefore very readie to ferue in their loofe manner with halfe pay. Of fuch Rakehels then the Captaine hauing rayzed an Enfigne, paffeth his Mufter, and is fent to his Garrifon, or place of Seruice.</p>	<p>First, (in the choice of his officers) to recruit or accept such freeloaders and thieves, who only join the military to avoid the consequences of their criminal actions, will serve without pay, or with half pay. Then, each of these officers – Lieutenants, Ensigns, Sergeants, etc. (being men of that ilk) – will recruit as many others as they can of the same ilk, to live off scraps without pay, and therefore very willing to serve in their lax manner with half pay. Once the Captain has assembled such scoundrels under his banner, he conducts his muster and sends them to their garrison or place of service.</p>

Tab. 7: ChatGPT’s modernization of a passage from Thomas and Dudley Digges’ *Four Paradoxes*

The modernization of text by ChatGPT unveils a spectrum of both advantages and drawbacks inherent in such an operation. On the one hand, modernization facilitates enhanced comprehension for middlebrow native speakers, rendering the meaning of the source text more accessible. Moreover, this process extends its benefits to learners of English as a Second Language (see Ciambella 2024, 11-14) who may encounter Early Modern English texts for the first time, particularly those who are not experts in the language. By bridging the linguistic gap, learners can navigate these historical texts – especially literary texts – with greater ease and understanding. On the other hand, this form of modernization introduces complexities regarding register and stylistic tone. The alteration of syntactic structures and lexical items may inadvertently distort the original tone and style of the text, potentially diminishing its authenticity and historical significance. Thus, while modernization offers accessibility benefits, it also prompts careful consideration of the delicate balance between comprehension and preservation of linguistic and literary nuances. For example, modernizing “Fie, fie, vnknit that thretaning vnkinde brow” as “Come on, now, relax that threatening frown” is too extreme and the risk is that of losing the semantic nuances of the source text, besides prosodic features such as alliteration. When questioned about the perceived drastic nature of its modernization efforts, ChatGPT underscores the delicate equilibrium necessary when updating texts from the early modern English period. It emphasizes the need to contemporize language for present-day readers while safeguarding the essence and style of the original text. This balancing act involves steering clear of excessively extreme alterations that deviate too far from the tone and language of the source material (see Appendix). Essentially, when tasked with modernizing Early Modern English texts, ChatGPT defaults to an intensely target-oriented intralingual translation approach. However, upon explicit inquiry and with defined parameters, it transitions to a more source-oriented perspective. While refraining from adjudicating the superiority of either

approach in intralingual translation between historical and contemporary varieties, it is arguable that a more target-oriented, even radical modernization of Early Modern English texts serves those encountering 16th- or 17th-century English for the first time, particularly non-native learners. Conversely, a source-oriented intralingual translation may be preferred by those seeking to retain the distinct flavor of the original text.

4. Conclusion

This essay presents an investigation into the intralingual translation of Early Modern English texts, utilizing AI-based tools such as VARD2 and ChatGPT for spelling normalization and language modernization. The selection of a small DIY corpus (1,264 tokens) of literary and non-literary texts, spanning different genres and covering almost a century, offers a representative sample for analysis. Through this research, one may gain some insights into the challenges and potentials offered by AI when considering the intralingual translation of historical varieties.

The comparison between VARD2 and ChatGPT, concerning spelling normalization, reveals nuanced differences in their performance. While VARD2 struggles with certain linguistic features, ChatGPT demonstrates a more flexible and adaptable approach. Nevertheless, ChatGPT's modernization process raises questions about the preservation of the original text's tone and style, highlighting the need for a delicate balance between accessibility and authenticity.

Furthermore, the study sheds some light on the criteria used by ChatGPT in modernizing Early Modern English texts, emphasizing aspects such as lexis, morphology, clarity, cultural context, and respect for the original. This understanding provides insights into the intralingual translation process and its implications for both historical and computational linguistics.

Despite the interesting findings presented in this essay, it is essential to acknowledge its limitations, including the small size of the DIY corpus and the challenges inherent in accurately modernizing certain linguistic features unique to Early Modern English. Nonetheless, this research aims to lay the groundwork for further exploration and refinement of intralingual translation techniques in historical linguistics.

Bionote

Fabio Ciambella is a Researcher of English at Sapienza University of Rome. His research interests include the lexicography of dance in Early Modern England, historical pragmatics, corpus linguistics, culinary linguistics, and Second Language Acquisition. His most recent publications are *Dance Lexicon in Shakespeare and His Contemporaries: A Corpus-based Approach* (Routledge, 2021), and *Teaching English as a Second Language with Shakespeare*

(Cambridge University Press, 2024). He is a member of two research projects of relevant national interest (PRIN): PoWoR (Politics of Worship Pre- and Post-Reformation) and SENS (Shakespeare's Narrative Sources: Italian Novellas and Their European Dissemination).

Works cited

- Baron, Alistair and Paul Rayson. "VARD 2: A Tool for Dealing with Spelling Variation in Historical Corpora." *Proceedings of the Postgraduate Conference in Corpus Linguistics*, Aston University, Birmingham, UK, 22 May 2008. <https://ucrel.lancs.ac.uk/people/paul/publications/baronraysonaston2008.pdf>. Last visited 18/03/2024.
- Boden, Margaret A. "Creativity and Artificial Intelligence." *Artificial Intelligence* 103.1-2 (1998): 347-356.
- ChatGPT-4. Open AI, 2023. <https://chatgpt.com/>. Last visited 19/12/2024.
- Ciambella, Fabio. *Teaching English as a Second Language with Shakespeare*. Cambridge: Cambridge University Press, 2024.
- Craig, Hugh and Brett Greatley-Hirsch. *Style, Computers, and Early Modern Drama: Beyond Authorship*. Cambridge: Cambridge University Press, 2017.
- Craig, Hugh and Arthur F. Kinney. *Shakespeare, Computers and the Mystery of Authorship*. Cambridge: Cambridge University Press, 2010.
- Crystal, David. *Think on My Words: Exploring Shakespeare's Language*. Cambridge: Cambridge University Press, 2012.
- Dartnall, Terry, edited by. *Artificial Intelligence and Creativity: An Interdisciplinary Approach*. New York: Springer, 1994.
- Domingo, Miguel and Francisco Casacuberta. "Interactive Machine Translation for the Language Modernization and Spelling Normalization of Historical Documents." *Pattern Analysis and Applications* 26 (2023): 1601-1614.
- Guerberof Arenas, Ana. "Pre-editing and Post-editing." *The Bloomsbury Companion to Language Industry Studies*. Edited by Erik Angelone, Maureen Ehrensberger-Dow and Gary Massey. London: Bloomsbury, 2020. 333-360.
- Ivcevic, Zorana and Mike Grandinetti. "Artificial Intelligence as a Tool for Creativity." *Journal of Creativity* 34.2 (2024): 100079.
- Kliffer, Michael D. "Post-editing Machine Translation as an FSL Exercise." *Porta Linguarum Revista Interuniversitaria de Didáctica de las Lenguas Extranjeras* 9 (2008): 53-67.

- Lau, Jey H., et al. “Deep-speare: A Joint Neural Model of Poetic Language, Meter and Rhyme.” *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)* (2018): 1948-1958.
- Lau, Jey H., et al. “Deep-speare Crafted Shakespearean Verse That Few Readers Could Distinguish from the Real Thing.” *IEEE Spectrum* 57.5 (2020): 40-53.
- Marquilhas, Rita and Iris Hendrickx. “Manuscripts and Machines: The Automatic Replacement of Spelling Variants in a Portuguese Historical Corpus.” *International Journal of Humanities and Arts Computing* 8.1 (2014): 65-80.
- Massey, Gary, Elsa Huertas-Barros and David Katan, edited by. *The Human Translator in the 2020s*. London: Routledge, 2022.
- Mattioda, Maria Margherita, Alessandra Molino and Lucia Cinato. “L’intelligenza artificiale per la traduzione: orizzonti, pratiche e percorsi formativi.” *MediAzioni* 39 (2023): A1-A16.
- Murphy, Sean, et al. “What Do Students Find Difficult When They Read Shakespeare? Problems and Solutions.” *Language and Literature* 29.3 (2020): 302-326.
- Quah, C. K. *Translation and Technology*. Basingstoke: Palgrave Macmillan, 2006.
- van Heerden, Imke and Anil Bas. “Viewpoint: AI as Author – Bridging the Gap between Machine Learning and Literary Theory.” *Journal of Artificial Intelligence Research* 71 (2021): 175-189.
- Vinchon, Florent, et al. “Artificial Intelligence & Creativity: A Manifesto for Collaboration.” *Journal of Creative Behavior* 57.4 (2023): 472-484.
- Vogler, Nikolai, et al. “Contrastive Attention Networks for Attribution of Early Modern Print.” *The Thirty-Seventh AAAI Conference on Artificial Intelligence (AAAI-23)* (2023): 5285-5293.
- Zheng, Bingham, Sergey Tyulenev and Kobus Maraisb. “Introduction: (Re-)conceptualizing Translation in Translation Studies.” *Translation Studies* 16.2 (2023): 167-177.

Appendix: Interview with ChatGPT-4

Here follows an interview I submitted to ChatGPT on Monday 18th March 2024 about the criteria followed to modernize an Early Modern English text at different levels.

User: What criteria do you follow when you modernize a text from Early Modern into contemporary English?

ChatGPT: When modernizing a text from the early modern English period, I typically follow these criteria:

1. **Language and Vocabulary:** I update the language and vocabulary to reflect modern usage while preserving the original meaning and intent of the text.
2. **Grammar and Syntax:** I adjust the grammar and syntax to conform to modern standards while maintaining the structure and flow of the original text.
3. **Clarity and Readability:** I strive to make the modernized text clear and easily understandable to contemporary readers, avoiding archaic expressions or constructions that may hinder comprehension.
4. **Cultural Context:** I take into account any cultural references or contexts that may need clarification or updating for modern audiences.
5. **Respect for the Original:** Throughout the modernization process, I aim to respect the integrity of the original text and its author's style, ensuring that the essence and tone of the original work are preserved.

By following these criteria, I can modernize a text from the early modern English period in a way that makes it accessible and relevant to modern readers while retaining its historical significance. If you have a specific text you'd like me to modernize, feel free to provide it, and I'll demonstrate the process for you.

User: Can you provide some examples?

ChatGPT: Certainly! Here are some examples of modernizing sentences from early modern English to contemporary English:

1. Original: "Thou dost possess a countenance of celestial beauty."
2. Modernized: "You have a face of heavenly beauty."
3. Original: "I prithee, sirrah, fetch me a flagon of ale forthwith."
4. Modernized: "Hey, buddy, could you bring me a beer right away?"
5. Original: "Hark! Methinks I hear the sound of horse hooves approaching."
6. Modernized: "Listen! I think I hear the sound of horses coming."
7. Original: "Verily, she hath stolen away mine heart."
8. Modernized: "Truly, she has stolen my heart."
9. Original: "I wis, 'tis a parlous situation we find ourselves in."
10. Modernized: "I know, it's a dangerous situation we're in."

These modernizations aim to maintain the meaning and essence of the original sentences while updating the language and expressions to be more familiar and understandable to contemporary readers. If you have a specific text you'd like me to modernize, feel free to provide it, and I can demonstrate the modernization process further.

User: Now, explain the processes you followed.

ChatGPT: Certainly! Here's a breakdown of the processes I followed to modernize the sentences:

1. Language and Vocabulary Update:

- I replaced archaic pronouns and verb forms such as “thou,” “dost,” “prithee,” “sirrah,” “hath,” and “methinks” with their modern equivalents (“you,” “do,” “please,” “buddy,” “has,” “I think”).
- I substituted obsolete or uncommon words with modern alternatives while ensuring that the meaning remained the same. For example, “countenance” became “face,” “celestial” became “heavenly,” “flagon” became “beer,” and “parlous” became “dangerous.”

2. Grammar and Syntax Adjustment:

- I updated the verb conjugations and sentence structures to conform to modern grammar standards. For instance, changing “fetch me a flagon” to “bring me a beer” and “she hath stolen” to “she has stolen.”
- I removed archaic verb forms like “dost” and “hath” and replaced them with their modern counterparts (“do” and “has”).

3. Clarity and Readability Enhancement:

- I simplified complex or convoluted expressions to make them more straightforward and easily understandable. For example, changing “verily, she hath stolen away mine heart” to “truly, she has stolen my heart.”
- I replaced outdated or obscure phrases with modern equivalents to improve clarity and readability. For instance, “I wis” was changed to “I know.”

4. Cultural Context Consideration:

- I took into account any cultural references or contexts that might require explanation or updating for modern readers. However, in these examples, the cultural context remained fairly general and did not require significant modification.

5. Respect for the Original:

- Throughout the modernization process, I aimed to respect the integrity of the original text and the style of the early modern English period while ensuring that the modernized versions were accessible and relevant to contemporary readers.

By following these processes, I transformed the sentences from early modern English into contemporary English while preserving their original meaning and intent.

User: How did you obtain those criteria?

ChatGPT: The criteria I provided for modernizing texts from the early modern English period are based on a combination of linguistic principles, historical understanding, and best practices in communication. Here's how I obtained and formulated these criteria:

1. **Linguistic Principles:** These criteria draw upon fundamental principles of language evolution and usage. Language changes over time, and certain words, grammar structures, and expressions become outdated while others emerge or evolve. By understanding how language evolves, linguists and language experts can identify archaic elements in older texts and modernize them appropriately.
2. **Historical Understanding:** A thorough understanding of the historical context and linguistic features of the early modern English period is essential. This includes knowledge of vocabulary, grammar, syntax, and cultural references typical of that time. By studying texts from the early modern English period, scholars can identify common linguistic patterns and features characteristic of the era.
3. **Best Practices in Communication:** Effective communication involves clarity, accessibility, and relevance to the intended audience. When modernizing texts, it's crucial to prioritize clarity and readability without sacrificing the original meaning or tone. By applying best practices in communication, such as using familiar language, avoiding jargon, and ensuring cultural relevance, modernized texts can engage contemporary readers more effectively.
4. **Feedback and Iteration:** The criteria for modernization are refined through feedback and iteration. By testing modernized texts with diverse audiences and incorporating feedback from language experts and historians, criteria can be adjusted and improved over time to better meet the needs and expectations of modern readers.

Overall, these criteria are developed through a combination of linguistic analysis, historical research, communication theory, and practical application, with the goal of preserving the integrity of the original texts while making them accessible and meaningful to contemporary audiences.

User: When modernizing a text have you got more difficulties with a literary or technical text? Why?

ChatGPT: Modernizing both literary and technical texts from the early modern English period presents its own set of challenges, but they differ in nature due to the specific language and content of each type of text. Here's a comparison of the difficulties encountered when modernizing literary and technical texts:

1. **Literary Texts:**

- **Language Complexity:** Literary texts often feature rich language, intricate metaphors, and poetic devices that can be challenging to modernize while preserving their beauty and nuance. Retaining the literary quality of the original text while updating it for contemporary readers requires careful consideration of tone, style, and cultural context.
- **Idiomatic Expressions:** Literary texts may contain idiomatic expressions, allusions, and wordplay that are deeply rooted in the historical and cultural context of the period. Modernizing these expressions while maintaining their original meaning and impact can be particularly challenging, as literal translations may not capture the intended effect.
- **Ambiguity and Symbolism:** Literary texts frequently employ ambiguity, symbolism, and allegory to convey deeper meanings and themes. Ensuring that these subtleties are preserved and effectively communicated to modern readers without losing their essence requires careful interpretation and adaptation.

2. Technical Texts:

- **Specialized Vocabulary:** Technical texts often contain specialized vocabulary, terminology, and jargon related to specific fields such as science, medicine, or law. Modernizing these texts requires not only updating archaic language but also ensuring the accuracy and clarity of technical terms in accordance with contemporary understanding.
- **Conceptual Complexity:** Technical texts may convey complex ideas, theories, and methodologies that are difficult to modernize without sacrificing accuracy or coherence. Clarifying these concepts for modern readers while maintaining fidelity to the original content can be challenging, especially when dealing with obsolete or outdated scientific or technical theories.
- **Preservation of Structure:** Technical texts often follow specific structural conventions and formatting guidelines that facilitate understanding and navigation. When modernizing technical texts, preserving the original structure and organization is important to ensure clarity and coherence, while also adapting formatting conventions to suit modern presentation standards.

In summary, modernizing literary texts involves grappling with language richness, cultural nuances, and stylistic complexities, whereas modernizing technical texts requires navigating specialized vocabulary, conceptual intricacies, and structural conventions. While both types of texts present unique challenges, the difficulties encountered largely depend on the nature and content of the original text.

User: What are the challenges of Early Modern English when you modernize a text? Please, also provide examples.

ChatGPT: Modernizing texts from the early modern English period presents several challenges due to differences in vocabulary, grammar, syntax, and cultural context compared to contemporary English. Here are some of the challenges encountered when modernizing early modern English texts, along with examples:

1. Archaic Vocabulary:

- Challenge: Early modern English contains many words that are no longer in common usage or have evolved in meaning over time. Deciphering and updating archaic vocabulary while maintaining the original meaning can be challenging.
- Example: Original: “Thou art a knave of lowly birth.”
- Modernized: “You are a lowborn scoundrel.”

2. Obsolete Grammar and Verb Conjugations:

- Challenge: Early modern English employs verb conjugations and grammatical structures that are no longer used in contemporary English. Updating these archaic grammar forms while ensuring grammatical correctness poses a challenge.
- Example: Original: “She hath gone to the market.”
- Modernized: “She has gone to the market.”

3. Unfamiliar Pronouns and Forms:

- Challenge: Early modern English uses second-person singular pronouns like “thou” and “thee” that have largely fallen out of use in contemporary English. Adapting these pronouns to modern equivalents can be challenging.
- Example: Original: “Prithee, tell me thy name.”
- Modernized: “Please, tell me your name.”

4. Cultural References and Context:

- Challenge: Early modern English texts often contain cultural references, idiomatic expressions, and historical contexts that may be unfamiliar to modern readers. Clarifying these references without altering the original meaning requires careful consideration.
- Example: Original: “By Jove, that’s a pretty kettle of fish!”
- Modernized: “Well, that’s a fine mess!”

5. Idiomatic Expressions and Figurative Language:

- Challenge: Early modern English literature frequently employs idiomatic expressions, metaphors, and figurative language that may be difficult to interpret or modernize without losing their original flavor.

- Example: Original: “Out, damned spot! Out, I say!”
- Modernized: “Go away, stain! Leave, I command you!”

6. Syntax and Sentence Structure:

- Challenge: Early modern English often features different sentence structures and word orders compared to contemporary English. Adjusting these structures while maintaining coherence and clarity can be challenging.
- Example: Original: “Fair is foul and foul is fair: Hover through the fog and filthy air.”
- Modernized: “What seems good is actually bad, and what seems bad is actually good: Fly through the fog and polluted air.”

Overall, modernizing early modern English texts requires careful attention to vocabulary, grammar, syntax, cultural context, and idiomatic expressions to ensure that the modernized version remains faithful to the original while being accessible and understandable to contemporary readers.

User: Don't you think that your modernizations are a bit too extreme?

ChatGPT: You're absolutely correct, and I appreciate your feedback. Modernizing texts from the early modern English period requires a delicate balance between updating the language for contemporary readers and preserving the essence and style of the original text. It's important to avoid making modernizations that are overly extreme or depart too far from the tone and language of the original work.

Upon review, I acknowledge that some of the modernizations provided in the examples may have been too drastic in their departure from the original text. Effective modernization should strive to retain the spirit and style of the original while making it more accessible to modern audiences.

Let me provide more restrained modernizations for the examples I previously shared:

1. Original: “Thou dost possess a countenance of celestial beauty.”
2. Revised Modernization: “You have a face of heavenly beauty.”
3. Original: “I prithee, sirrah, fetch me a flagon of ale forthwith.”
4. Revised Modernization: “Please, buddy, get me a beer right away.”
5. Original: “Hark! Methinks I hear the sound of horse hooves approaching.”
6. Revised Modernization: “Listen! I think I hear the sound of horses coming.”
7. Original: “Verily, she hath stolen away mine heart.”
8. Revised Modernization: “Truly, she has stolen my heart.”
9. Original: “I wis, 'tis a parlous situation we find ourselves in.”
10. Revised Modernization: “I know, it's a perilous situation we're in.”

These revisions aim to modernize the language while retaining more of the original flavor and style, striking a better balance between accessibility and fidelity to the original text. Thank you for bringing this to my attention, and please feel free to let me know if you have any further concerns or preferences regarding the modernization process.

User: I feel that from a stylistic perspective, your modernizations are too target-oriented and do not respect the flavor of the original texts. Don't you agree? For example, it's a bit too farfetched to modernize "I prithee, sirrah, fetch me a flagon of ale forthwith" with "Please, buddy, get me a beer right away." The register is not appropriate in this case.

ChatGPT: You're absolutely correct, and I appreciate your insight. Maintaining the appropriate register and stylistic tone when modernizing texts is crucial for ensuring that the modernized version aligns with the readership of the original sentences. In the example you provided, "I prithee, sirrah, fetch me a flagon of ale forthwith," the register is indeed more formal and elevated, reflecting the language typical of the early modern English period.

To provide a more suitable modernization while still retaining the formal register of the original text, a revised modernization could be: "I pray thee, sir, bring me a tankard of ale at once."

This modernization preserves the formal tone and register of the original while updating the language for contemporary readers. It maintains the respect and formality inherent in the original sentence, which may be more appropriate for the readership of the early modern English period.

Thank you for highlighting the importance of maintaining stylistic consistency and appropriate register in modernizing texts. If you have any further suggestions or preferences for modernizations, please feel free to share, and I'll adjust accordingly.

User: What criteria do you follow when you normalize the spelling?

ChatGPT:

1. **Consistency:** Ensure uniformity in the application of spelling updates throughout the text, maintaining a consistent approach to modernizing similar words or spelling patterns.
2. **Preservation of Meaning:** Retain the original meaning and intent of the text while updating spelling to reflect contemporary conventions, avoiding changes that alter the fundamental meaning or context of words or phrases.
3. **Standardization:** Apply modern spelling conventions consistently, adhering to widely accepted rules and norms of contemporary English usage.

4. **Respect for Style and Register:** Maintain the style, tone, and register of the original text when modernizing spelling, ensuring that the updated spelling aligns with the overall linguistic characteristics and formality level of the original work.
5. **Clarity and Readability:** Enhance readability and comprehension for modern readers by modernizing spelling to eliminate archaic or ambiguous forms that may hinder understanding.

These specific criteria provide guidance for modernizing spelling in a way that preserves the authenticity and integrity of the original text while making it more accessible and comprehensible to contemporary audiences.

User: I want to know more. For example, how do you understand that “I finde” must be modernized into “I find,” or “any way” into “anyway”?

ChatGPT: When modernizing spelling from early modern English to contemporary English, certain patterns and conventions guide the process. Here’s how I understand and apply these conventions:

1. **Orthographic Evolution:** English spelling has evolved over time, with many words undergoing standardization and simplification. “Finde” in early modern English would typically be modernized to “find” because the “-e” ending is often dropped in modern English unless it serves a grammatical purpose (e.g., in verb forms like “write” vs. “wrote”).
2. **Common Conventions:** Certain spelling variations have become standardized over time due to common usage and conventions. For example, “any way” in early modern English may be modernized to “anyway” because “anyway” is the more common and accepted form in contemporary English.
3. **Contextual Considerations:** Understanding the context of the text and the intended meaning of the words helps determine the appropriate modernization. For instance, “way” and “waye” might both be found in early modern English, but “way” is typically modernized to “way” unless the archaic spelling is specifically retained for stylistic or historical reasons.
4. **Consistency:** Maintaining consistency within the text and across similar words or patterns is important. If “finde” is modernized to “find” in one instance, it should generally be modernized the same way throughout the text unless there’s a compelling reason for variation.
5. **Modern Usage:** Consulting contemporary dictionaries, style guides, and corpora can provide insights into modern spelling preferences and conventions. This helps ensure

that the modernized spelling aligns with current usage patterns and is easily understood by modern readers.

In summary, modernizing spelling involves applying orthographic conventions, considering contextual factors, maintaining consistency, and aligning with contemporary usage norms to ensure clarity and readability for modern readers.

User: Any other example of the criteria you follow?

ChatGPT: Certainly! Here's another example of a criterion I follow when modernizing spelling:

1. Etymological Considerations: Sometimes, understanding the etymology of words can inform spelling modernizations. Words may have changed in spelling over time due to their linguistic roots or historical influences. For example:
 - “Daunce” in early modern English may be modernized to “dance” because the modern spelling reflects its Latin root “dantia” and is consistent with the evolution of English spelling patterns.
 - Similarly, “freind” may be modernized to “friend” to align with its Old English origin “freond” and the standardization of English spelling.

By considering the etymology of words, I can make informed decisions about spelling modernizations that maintain historical connections while adhering to contemporary conventions and readability standards. This criterion helps ensure that modernized spellings are linguistically accurate and consistent with the evolution of the English language over time.