encyclopedia

MDPI

Entry

The Computational Study of Old English

Javier Martín Arista

Department of Modern Languages, University of La Rioja, 26004 Logroño, Spain; javier.martin@unirioja.es

Definition

This entry presents a comprehensive overview of the computational study of Old English that surveys the evolution from early digital corpora to recent artificial intelligence applications. Six interconnected domains are examined: textual resources (including the *Helsinki Corpus*, the *Dictionary of Old English Corpus*, and the *York-Toronto-Helsinki Parsed Corpus*), lexicographical resources (analysing approaches from Bosworth–Toller to the *Dictionary of Old English*), corpus lemmatisation (covering both prose and poetic texts), treebanks (particularly Universal Dependencies frameworks), and artificial intelligence applications. The paper shows that computational methodologies have transformed Old English studies because they facilitate large-scale analyses of morphology, syntax, and semantics previously impossible through traditional philological methods. Recent innovations are highlighted, including the development of lexical databases like *Nerthusv5*, dependency parsing methods, and the application of transformer models and NLP libraries to historical language processing. In spite of these remarkable advances, problems persist, including limited corpus size, orthographic inconsistency, and methodological difficulties in applying modern computational techniques to historical languages. The conclusion is reached that the future of computational Old English studies lies in the integration of AI capabilities with traditional philological expertise, an approach that enhances traditional scholarship and opens new avenues for understanding Anglo-Saxon language and culture.

Keywords: Old English; historical language processing; natural language processing; artificial intelligence; computational linguistics; corpus lemmatisation; digital lexicography; universal dependencies

Academic Editor: Raffaele Barretta

Received: 25 June 2025
Revised: 31 July 2025
Accepted: 28 August 2025
Published: 4 September 2025

Citation: Martín Arista, J. The Computational Study of Old English. *Encyclopedia* **2025**, *5*, 137. https://doi.org/10.3390/encyclopedia5030137

1. Introduction

This entry provides a comprehensive overview of the linguistic computational approaches to studying Old English. The computational study of Old English applies digital methodologies to analyse the earliest diachronic stage of the English language (c. 600–1150 CE), focusing on digital and computational methodologies rather than the medieval computus tradition. This interdisciplinary field integrates philology, corpus linguistics, and artificial intelligence in order to carry out large-scale data-driven studies in Old English texts. Digitising manuscripts, annotating syntactic structures and training machine learning models allow scholars of Old English to capture the various synchronic and diachronic phenomena at the morphological, syntactic, and semantic levels of analysis. Many of these aspects were previously imperceptible through traditional manual analysis. The foundation of this approach lies in curated resources such as electronic corpora, machine-readable dictionaries and annotated treebanks, which provide structured datasets for computational models.

The remainder of this entry is structured as follows: Section 2 examines textual resources, including digital corpora like *The Helsinki Corpus*, *The Dictionary of Old English*

Corpus, and *The York-Toronto-Helsinki Parsed Corpus of Old English Prose*. Section 3 explores lexicographical resources and discusses how dictionaries from Bosworth–Toller to *The Dictionary of Old English* have been digitised and enhanced with textual data. Section 4 tackles corpus lemmatisation methodologies for both prose and poetry. Section 5 deals with treebanks and focuses on the application of Universal Dependencies frameworks to Old English. Section 6 investigates artificial intelligence developments through the progression from rule-based systems to modern transformer models. Section 7 draws the main conclusions of this research.

2. Textual Resources

2.1. The Helsinki Corpus of English Texts (1980s–1990s)

The computational study of Old English began in the 1980s with the *Helsinki Corpus of English Texts* (HCET), compiled at the University of Helsinki. The HCET comprises texts from the 8th to the 18th centuries, with Old English comprising approximately 413,000 words. It introduced manual annotation of part-of-speech tags and metadata such as genre, authorship and chronology [1] (p. 1). The corpus is structured with four periods for Old English: O1 (pre–850), O2 (850–950), O3 (950–1050), and O4 (1050–1150). Text categorisation parameters include genre, dialect, manuscript date, composition date, author, audience/level of formality, and prototypical text category. The corpus contains law texts, religious treatises, biblical translations, historical writings, charters, documents, and scientific/medical texts [1] (pp. 18–24). The *Helsinki Corpus* remains a cornerstone for diachronic linguistic research. Its structured sampling methodology, which stratifies texts by genre, region, and period, is the basis of studies on phenomena like the decline of inflectional morphology. Despite its limitations, such as rule-based annotation and a focus on prose, the HCET established the standard for historical corpus design and inspired subsequent projects like the *York-Toronto-Helsinki Parsed Corpus of Old English* [1] (p. 29).

2.2. The Dictionary of Old English Corpus (1981–Present)

A breakthrough in the field of Old English studies occurred in 1981 with the *Dictionary of Old English Corpus* (DOEC), led by Antonette diPaolo Healey at the University of Toronto. The DOEC digitised nearly all surviving Old English texts (approximately 3000), totalling approximately 3 million words, including poetry, prose, and glosses. The integration of the DOEC with the *Dictionary of Old English* (DOE) has set a benchmark for digital lexicography because each entry is based on corpus data [2] (p. 3). For instance, the DOE entry for *mōd*, 'mind, spirit', documents its usage in 847 instances in religious, poetic, and legal texts, which illustrates its semantic range and grammatical behaviour. The DOE Web Corpus offers search functionalities with simple search (basic word forms) and advanced search (complex Boolean searches, wildcards). The XML of the corpus architecture facilitates machine-readable queries, which set a precedent for later digital editions. These advances led to a degree of textual accuracy compatible with computational analysis.

2.3. The York-Toronto-Helsinki Parsed Corpus (1990s)

The 1990s saw the appearance of syntactically annotated corpora, most notably the York-Toronto-Helsinki Parsed Corpus of Old English Prose (YCOE). The YCOE annotated 1.5 million words of Old English prose with part-of-speech and phrase-structure trees. This has resulted in studies in the syntactic phenomena of Old English such as verb-second order and clausal subordination [3] (p. 1). The segment in (1) can be described as follows: the YCOE uses a part-of-speech (POS) tagging system that displays part-of-speech and morphological features.

(1)
*He*ˆPNnom *aras*ˆVPS *þa*ˆADV *of*ˆP *þam*ˆD *deaþe*ˆNSD *on*ˆP *life*ˆNSD.

The segment in (1) can be understood as follows: He (personal pronoun, nominative), aras (verb, past, singular, 3rd person), þa (adverb), of (preposition), þam (demonstrative), deaþe (noun, singular, dative), on (preposition), and life (noun, singular, dative). For phrase structure parsing, the YCOE uses a labelling system that represents syntactic dependency, hierarchy and linearisation. A parsed sentence appears as shown in (2), where constituent structure is derived from bracketed labelling and indentation.

(2)
```
( (IP-MAT (NP-NOM (PRO he))
          (VBD aras)
          (ADVP-TMP (ADV þa))
          (PP (P of)
              (NP (D þam) (N deaþe)))
          (PP (P on)
              (NP (N life)))
          (. .))
```

Studies using YCOE data demonstrated, among other things, the gradual rigidification of subject-verb-object word order in late Old English. The seminal study by Taylor and Pintzuk [3] (pp. 12–15) showed that OV order declined from 75% in early Old English to 40% by 1100 CE, correlating with the rise in SVO structures.

2.4. Corpus Search Tools

CorpusSearch [4] is a query and search tool specifically designed for parsed corpora, particularly used with the Penn Historical Corpora including the YCOE. Its main functionalities include searching for syntactic structures, node labels (NP, VP, PP), tree relationships, clause types, and lexical items [3] (p. 18). Query types include basic searches (exists, precedes, dominates, hasSister, isDaughter, iPrecedes) and complex searches using AND, OR, NOT, wildcards, and regular expressions. Output options include various display formats and statistical functions. For example, to search for subordinate clauses with the conjunction *þæt*, where the subject is postposed after the finite verb, the node reference and the query string presented in (3) would be used:

(3)
node: IP-SUB*
query: (IP-SUB* dominates C þæt) AND (IP-SUB* dominates NP-NOM) AND (IP-SUB* dominates VBD | VBP | VB | MD) AND (NP-NOM precedes !VBD | VBP | VB | MD)

2.5. Lexical Databases of Old English

Lexical databases contribute to bridging the gap between traditional lexicographical resources and computational linguistics. The Nerthus Project, based at the Universidad de La Rioja, has developed a series culminating in Nerthusv5 (2024). This database functions as an interface between textual and lexicographical sources, thus allowing researchers to relate attested forms to headwords and meaning definitions from three complete dictionaries of Anglo-Saxon, as well as to link textual to lexicographical resources of Old English [5]. Nerthusv5 contains 32,812 headword files, with spelling largely guided by the Clark Hall Dictionary. For the AI segment, it also provides corresponding headwords from the DOE. The database incorporates a simplified analysis of derivational morphology, including morphological status, base of derivation, and category.

The Nerthus database has fostered studies in Old English linguistics. Martín Arista's work [6] has explored lexical databases, derivational maps and three-dimensional represen-

tations of complex polysemy. Other researchers have examined specific aspects. To cite just a few, Ojanguren López [7,8] addressed verbal classes in a corpus of Old English, Mateo Mendaza [9] analysed the Old English exponent for the semantic prime MOVE, Novo Urraca [10] investigated deadjectival paradigms and Vea Escarza [11,12] examined morphological recursivity and adjectival affixation. The integration of these lexical databases with corpus resources has significantly advanced computational approaches to Old English and provided functionalities of complex search across multiple levels of linguistic structure.

2.6. The Parallel Corpus of Old English Prose (ParCorOE)

ParCorOE [13] is a digital resource developed by the Nerthus Project. ParCorOE is an open-access parallel corpus of Old English–contemporary English that is fully annotated and aligned at the levels of the text, fragment and word. The corpus has gone through several versions, with v3.0 released in 2024. In its current state, ParCorOE contains over 300,000 entries of parallel text. Key features include a token identification system and detailed linguistic annotation including POS tagging, sources, and translation. The 2024 release added texts such as Ælfric's *Catholic Homilies I*, *The Anglo-Saxon Chronicle A-E*, *The Benedictine Rule*, Charters and Wills, Laws, and St Augustine's *Soliloquies*. ParCorOE offers several search options, both by individual field and by multiple fields, including lemma, category, inflectional form, gloss, textual source, and translation. For instance, ParCorOE could solve searches like identifying 935 instances of the instrumental case or examining 950 occurrences of *weorðan*, 493 of which are inflected for the past. It also facilitates thematic lexical studies, such as words related to disease, which include terms like *ābrocen* 'diseased', *ādl* 'disease', *ādlig* 'diseased', and many others. ParCorOE has been used by Martín Arista et al. [14] for training automatic dependency parsing models for Old English.

2.7. Comparative Perspective: Old English Computational Resources in Context

The computational resources available for Old English can be meaningfully contextualised through comparison with similar tools developed for other historical languages, particularly Latin and Old Norse. Latin computational linguistics benefits from a significantly larger corpus (approximately 100 million words versus the 3 million in Old English written records) and more standardised orthography. The Latin BERT model [15] achieves higher accuracy rates in tasks like POS tagging and lemmatisation compared to Old English tools, primarily due to this advantage in data volume. Similarly, the Perseus Digital Library provides extensive annotated Latin texts with established evaluation metrics for accuracy. Old Norse computational approaches, meanwhile, share many problems with Old English, including dialectal variation and limited corpus size. The Medieval Nordic Text Archive (Menota) has developed XML-based approaches to representing manuscript variation that parallel efforts in Old English digital editions.

Latin computational resources have emphasised supervised learning from hand-annotated data, benefiting from extensive ecclesiastical text collections and standardised orthography. In contrast, Old Norse computational approaches have developed methods for handling orthographic variation, including normalised-to-diplomatic text mapping systems that maintain both standardised and manuscript-specific representations in parallel. The XML-TEI framework adopted by Menota preserves multiple textual levels (facsimile, diplomatic, normalised) within a single document architecture. Old English computational approaches have necessarily developed hybrid strategies that combine elements from both traditions by combining normalisation techniques for consistency and the preservation of manuscript fidelity through variant tracking systems.

These cross-linguistic comparisons highlight that computational approaches to Old English must address specific questions related to orthographic inconsistency and limited

data that require specialised solutions distinct from those applied to better-resourced historical languages. The development of resources like Nerthusv5 and ParCorOE represents field-specific contributions within the broader context of historical language processing.

2.8. Critical Assessment of Corpus Representativeness and Bias

The computational study of Old English faces questions about corpus representativeness that affect the validity of research findings. The surviving textual records represent a heavily skewed sample of historical language use. Moreover, there is significant overrepresentation of religious and legal texts (comprising approximately 70% of the DOEC) and underrepresentation of spoken registers and secular literature. This bias has methodological implications for computational analyses. Statistical studies of lexical frequency may reflect textual genre distribution rather than actual language use patterns. Similarly, syntactic analyses based on the YCOE may overrepresent formal prose registers but provide limited insight into colloquial Old English syntax. Geographical and temporal sampling is not without problems either. West Saxon texts dominate the corpus (approximately 60% of surviving material), potentially skewing computational models toward this dialectal variety. The temporal distribution is similarly uneven, with significantly more material from the late Old English period (post-950 CE) than from earlier centuries. These sampling biases call for careful interpretation of computational findings and highlight the need for explicit acknowledgment of corpus limitations in published research. Future computational work should incorporate bias-aware methodologies that account for these distributional skews in both model training and result interpretation.

3. Lexicographical Resources

3.1. Historical Lexicography: Approaches and Evolution

The three major 19th-century Old English dictionaries represent different approaches. Bosworth's Dictionary (1838, revised 1848) contains approximately 26,000 entries and uses normalised West Saxon spellings while acknowledging dialectal variants. Its entry structure is complex, featuring headwords with variants, etymological information, meaning hierarchies, Latin equivalents, illustrative quotations, and compound words integrated into main entries [15] (p. 188). Sweet's *Student's Dictionary of Anglo-Saxon* (1896) was more selective, containing around 12,000 entries and employing strict West Saxon normalisation. It was designed for student use, with streamlined headwords, brief definitions, minimal etymological information, few textual citations, and a focus on common vocabulary [16] (p. 190). Clark Hall's *Concise Anglo-Saxon Dictionary* (1894) constitutes a middle ground but exceeds both in scope, with approximately 35,000 entries. It employs a mixed normalisation system and draws from a broader textual base. The entry structure features headwords with limited variants, concise definitions, some etymological notes, brief citations, and extensive cross-referencing [17] (p. 192). Ellis [18] (p. 194) demonstrates that Bosworth–Toller generally presents forms with <eo> as main entries, relegating <io> variants to cross-references but maintaining <ie> variants within entries. Clark Hall tends to prefer later West Saxon forms, standardising toward <y> spellings over both <ie> and <i>. Sweet's dictionary shows the most systematic approach, given that it consistently normalises to Late West Saxon forms and indicates important dialectal variants.

3.2. Bosworth–Toller Anglo-Saxon Dictionary

The digitisation of the Bosworth–Toller *Anglo-Saxon Dictionary* [19] marked a milestone in Old English lexicography. Hosted by Charles University, the online edition integrates advanced search functionalities that allow users to filter entries by dialect, word class and period. A sample entry for *werod*', troop, host, and multitude shows the main headword

with spelling variants, followed by grammatical information, definitions and semantic domains, lists compounds and related forms, as can be seen in (4).

(4)

werod, *weorod*, *werud*, *wered*. *es*; *n*.
A troop, band, company, host [Latin: turba, caterva]
of people generally: Micel werod 'great multitude' - Luke 6:17
specifically: a band of warriors, army; host of heaven, angelic host; company of saints/blessed
Compounds: *heofon-wered*, *sige-wered*, *þēod-wered*.
Related forms: *wered-lēst*, *werodiān*.

The structure of this entry comprises extensive cross-referencing of attestations, hierarchical organisation of meanings, Latin glosses, citations from both prose and poetry, clear marking of grammatical information, etymology suggestions, and related compound forms. Despite retaining some 19th-century lexicographical biases, this resource remains the standard reference work of Old English lexicology and lexicography because it is more extensive that the other dictionaries published at the turn of the 20th c. and, above all, because it is complete.

3.3. The Dictionary of Old English

The *Dictionary of Old English* (DOE) [20] currently covers letters A-Le and represents a more modern lexicographical approach that integrates indexes, databases and concordances. The sample entry for hand given in (5) illustrates its structure.

(5)

hand (*hond*) noun fem.
FORM: sg nom/acc: *hand*, *hond*; sg gen: *handa*, *honda*; sg dat: *handa*, *honda*; pl nom/acc: *handa*, *honda*; pl gen: *handa*, *honda*; pl dat: *handum*, *hondum*
FREQUENCY: c. 2900 occurrences
MEANING DEFINITION: The human hand as anatomical part referring to physical hand; in reference to handedness; hand as measure/bodily reference; in measurement; as directional indicator.
Extended meanings: power, control, authority; possession, custody; agency, work; pledge, guarantee.
Compounds: *hand-æx*, *hand-bell*, *hand-geweorc*.
Collocations: *on hand gan*, *to handa*.
Latin correspondences: *manus*, *potestas*.

The DOE entry given in (5) shows the complete morphological paradigm, precise frequency data, chronological organisation within semantic fields, extensive documentation of usage, clear marking of syntactic patterns, careful attention to Latin correspondences, exhaustive citation of occurrences, and detailed subentry structure. Key differences with respect to Bosworth–Toller include more systematic organisation, statistical information, complete morphological information, more extensive citations, and greater attention to collocations and structural patterns.

3.4. The Lemmatisation Project

A significant advance of Old English corpus linguistics and lexicography has been the development of automated lemmatisation systems. Lemmatisation, assigning the dictionary headword to each word form, is essential for linking corpus occurrences to dictionary entries and for carrying out statistical analyses across inflected forms. The combination of these lemmatisation resources with dictionaries creates a powerful framework for computational analysis of Old English texts that launches sophisticated queries and finds statistical patterns impossible with raw text alone [21] (p. 187). It also facilitates comparative studies of prose-poetry and helps to disambiguate homographs (identical

forms representing different lemmas). For instance, *bær* could be a form of *beran*, 'to bear', or *bær*, 'bare', and only context can determine which is correct [21] (p. 185). The question of lemmatisation is addressed in more detail in Section 4.

4. Corpus Lemmatisation

4.1. Three Dissertations on the Lemmatisation of the Old English Verb

Metola Rodríguez's [22] dissertation addresses the lemmatisation of strong verbs in the DOEC. The author develops a four-step methodology combining computational queries with manual verification. The approach creates specifically designed query strings targeting stems, inflections and prefixes. For example, when lemmatising a Class II strong verb like *beodan*, 'to command', the system identifies forms like *bead* (preterite) and *budon* (plural preterite), distinguishing them from homographic nouns and adjectives. The research achieves approximately 80% accuracy of lemma identification before manual revision.

García Fernández's [23] dissertation addresses the lemmatisation of preterite-present, contracted, anomalous, and strong class VII verbs. The author develops a semi-automated methodology that combines computational searches with manual verification. For preterite-present, contracted, and anomalous verbs, the method leverages morphological relatedness to find derivatives based on simplex forms. The research produces a comprehensive inventory of lemmas and inflectional forms, particularly valuable for the letters I-Y which were not previously available from lexicographical sources.

Tío Sáenz's [24] dissertation addresses the lemmatisation of weak verbs across all three subclasses. The author develops a four-step methodology combining automatic searches with manual validation. The approach creates patterns for non-canonical lemmatisation based on inflectional endings and stem vowel alternations. For example, when processing the Class II weak verb *gearwian*, 'to prepare', the system identifies not only canonical forms like *gearwode* but also non-canonical variants such as *gærwod* and *gegarwed*. The dissertation fills an important gap, as weak verbs represent the largest and most productive verbal class in Old English.

4.2. The Lemmatisation of Old English Prose

The Variation in Old English (VariOE) project, developed at the University of Łódź, has created a fully lemmatised version of the YCOE. This project addresses a critical pending question of Old English corpus linguistics by transforming a syntactically parsed corpus into one that is also lemmatised. The project has created two online tools which enable investigating phraseology and the relations between syntax and lexicon in Old English prose: a morphological dictionary for the YCOE corpus and a dictionary of Old English collocations [25]. The project employs a multi-stage methodology combining computational techniques with manual philological expertise. The initial stage applies rule-based algorithms that match inflected forms to dictionary headwords based on morphological patterns and part-of-speech information. What distinguishes this methodology is its treatment of ambiguous forms. The system implements a hierarchical disambiguation process considering morphosyntactic information from YCOE tags, contextual evidence from surrounding words, statistical probability based on attested patterns, and manual review of challenging cases. The lemmatisation achieves an accuracy rate of over 95% across the corpus. The lemmatised data preserves the original YCOE annotation while adding lemma information, which allows researchers to perform searches based on lexical items across all inflectional variants, specific morphosyntactic features, combinations of lexical and syntactic criteria, and dialectal and chronological variations. The project maintains a comprehensive lexicon of approximately 30,000 lemmas. The complete lemmatisation has also yielded a fully searchable lemmatised version of the YCOE corpus compatible with the

original parsed format, a detailed inventory of Old English lemmas with their inflectional variants, documentation of systematic patterns in dialectal and orthographic variation, and tools for identifying and analysing lexical collocations and patterns.

4.3. The Lemmatisation of Old English Poetry

Hamdoun Bghiyel's [26] dissertation engages in the lemmatisation of Old English poetry. It undertakes the lemmatisation of the *York-Toronto-Helsinki Parsed Corpus of Old English Poetry* (YCOEP) with the Interface of Old English Dictionaries (IOED; Martín Arista [26]) integrated in *Nerthusv5*. Hamdoun Bghiyel's research systematically groups attested spellings from the YCOEP under unified headwords using the IOED relational database. This database contains key Old English lexicographical resources, including the lemmatisation provided by other corpora and dictionaries. The methodology combines morphology, lexicography, and corpus analysis in four main steps: compiling the inventory of attested forms from the YCOEP, correlating these forms with IOED sources to assign lemmas, aligning the assigned lemmas with the final ParCorEv2 lemma list [14], and validating the lemmatised inventory against reference sources. The IOED methodology successfully assigned lemmas automatically to 73% of the unique attested forms, though accuracy varied across lexical categories. Functional categories like prepositions and conjunctions had accuracy rates exceeding 95%, while about 27% of the inventory required manual lemma assignment. The methodology assigned ParCorEv2 lemmas to 99% of the attested forms, with only 1% requiring new lemmas. The main contribution of this dissertation is the creation of a comprehensive lemmatised inventory for Old English poetry, including around 7000 lemmas assigned to 69,456 attested forms. The study has also produced 72 new lemmas, primarily for compound words not yet registered by the DOE, including 28 noun lemmas, 19 adjectival lemmas, 11 verbal lemmas, and 13 lemmas for adverbs and other grammatical categories.

4.4. Evaluation Metrics for Lemmatisation Systems

The effectiveness of Old English lemmatisation systems can be assessed through established evaluation metrics rather than anecdotal evidence. The main performance measures are accuracy (the percentage of correctly lemmatised tokens) and disambiguation accuracy (the correct identification of lemmas for homographic forms). Published evaluations demonstrate varying performance across different approaches. The Nerthus Project's lemmatisation methodology achieves 73% initial automatic accuracy with specific performance variations across grammatical categories [26]. Functional categories like prepositions and conjunctions show accuracy rates exceeding 95%, while complex forms like compounds raise some issues. The VariOE project reports overall accuracy of approximately 95% for Old English prose lemmatisation [25]. These evaluations typically employ manually annotated gold standard test sets comprising selected samples that represent the linguistic diversity of Old English texts. The quantified performance metrics demonstrate both the significant progress in computational approaches to Old English and the persistent difficulties that require hybrid automated-manual methodologies, particularly for poetic and dialectally diverse texts.

5. Treebanks of Old English

5.1. Universal Dependencies for Old English

While the YCOE takes a phrase-structure approach to syntax, more recent efforts have focused on dependency grammar, particularly within the Universal Dependencies (UD) framework [27]. The UD project aims to develop cross-linguistically standardised treebank annotation for many languages, including historical ones like Old English [28]. Unlike

constituent-based approaches, dependency parsing represents syntactic structure through directed relationships between words, without intermediate phrasal nodes. Each word is attached to its head through a typed dependency relation. This approach has advantages for analysing Old English, including better handling of its relatively free word order and allowing cross-linguistic comparison with related Germanic languages. Martín Arista [29] pioneered work on adapting the UD framework for Old English, converting a subset of the YCOE treebank to dependency structures. Similarly, Villa and Giarda [30] have worked on developing dependency-based representations of Old English morphosyntax, particularly focusing on verbal constructions and argument structure. Recent work by Martín Arista et al. [14] has extended the UD approach to Old English by implementing artificial intelligence techniques. These authors have developed a method for the automatic annotation of Old English texts with the UD framework. Their study found that pre-training the tok2vec stage yielded better results than the default pipeline configuration of the NLP library spaCy, in such a way that larger training corpora significantly improved performance [14] (p. 6).

5.2. From YCOE to UD: Rule-Based Parsing

Villa and Giarda [31] have undertaken a rule-based approach to identify the root of dependency trees in Old English sentences. This is the first step toward a rule-based automatic conversion of the entire YCOE treebank into the UD format. Their method applies hierarchical rules to identify the root based on the original YCOE constituency annotation. The rules aim at identifying the root based on part-of-speech tags and syntactic structures. This approach achieved a precision of 89.49% in root identification for a test set of 390 sentences, significantly outperforming previous methods that reached only 78.46% accuracy [31] (p. 26). This rule-based approach has proven particularly effective for handling specific syntactic constructions in Old English, including verbal constructions with the verb 'to be', passive constructions, and coordinated verb phrases. Error analysis revealed that most challenges occurred with Latin sentences embedded within Old English texts, nominal sentences lacking a verb, and copular constructions with negated verbs [31] (p. 27).

5.3. Comparative Performance of Parsing Approaches

The various approaches to Old English dependency parsing can be evaluated on the grounds of their performance on benchmark datasets. These comparative evaluations provide context for understanding the relative advantages of different computational approaches to Old English syntactic analysis and help explain why no single methodology has emerged as definitively superior in the field. Three primary methodologies have emerged in recent years: (i) rule-based conversion from constituency structures; (ii) statistical parsers trained on annotated data; and (iii) neural approaches using transformer architectures. The rule-based approach developed by Villa and Giarda [31] achieves 89.49% accuracy in root identification on their test set of 390 sentences, with lower performance on complete dependency structures. Neural approaches using spaCy's implementation [14] demonstrate strong overall performance but require substantial computational resources. Each approach has specific strengths: rule-based methods effectively capture linguistic generalisations but may lack robustness to variation; statistical parsers balance performance and interpretability; and neural methods handle complex syntactic patterns but depend on larger training sets than are typically available for historical languages.

6. AI Developments in Old English Studies

6.1. Overview: Historical Development of AI Applications

The evolution of AI applications for Old English studies reflects broader trends in artificial intelligence. Early systems from the 2000s relied primarily on rule-based approaches and simple statistical models. For example, early lemmatisers used predefined paradigms and exception lists to handle Old English morphology [32] (p. 395). By the early 2010s, machine learning approaches began to gain traction as researchers applied techniques like maximum entropy models and conditional random fields to tasks such as POS tagging and syntactic parsing. These approaches achieved higher accuracy than rule-based systems but required extensive feature engineering [32] (p. 396). The mid-2010s saw the rise in neural network approaches, initially with recurrent neural networks (RNNs) and then increasingly with attention-based models. These systems proved particularly effective for handling the complexities of Old English syntax and morphology, as they could learn hierarchical patterns without explicit rules [33] (p. 42). The most recent developments since 2018 have relied on transformer-based models and transfer learning approaches. Researchers have achieved impressive results despite the relatively small size of Old English datasets [15] (p. 8). The recent work by Martín Arista et al. [14] that applies spaCy and transformer models to Old English dependency parsing represents the current state of the art. This work demonstrates that even with limited training data, modern NLP techniques can achieve satisfactory results for historical language processing.

6.2. Methodological Challenges in AI for Historical Languages

Applying modern AI techniques to historical languages raises several methodological challenges beyond data limitations [34]. First, there is the issue of evaluation: without native speakers, assessing the quality of model outputs relies on specialist scholarly judgement, which may itself be subject to debate [35] (p. 1410). Second, there is the risk of anachronism: modern computational models, trained primarily on contemporary language, may inadvertently impose modern linguistic patterns or categories on historical data. This requires careful calibration and historical linguistic expertise in both model design and interpretation [35] (p. 1412). Villa and Giarda [30] (p. 35) address some of these challenges and note that parsers frequently fail to adequately parse distinctive features of Old English syntax, such as the freedom of word order, case syncretism, the co-existence of prepositional and postpositional government, as well as the various types of marking of relative clauses. Similarly, Martín Arista et al. [14] (p. 8) identify several linguistic phenomena that cause specific problems for automatic dependency parsing, including negative contractions, flat multiword expressions, appositive constructions and various clause-level structures. Third, there is the challenge of interdisciplinary communication: effective collaboration between AI researchers and historical linguists calls for terminological and methodological links between fields [36] (p. 100).

The application of transformer models to Old English requires specific architectural adaptations to meet the characteristics of the language. The primary transformer architectures that have been tested include the following: (i) BERT-based models with vocabulary adaptations for Old English-specific graphemes (æ, Æ, ð, Ð, þ, Þ); (ii) RoBERTa variants with modified tokenisation strategies; and (iii) lighter implementations like MobileBERT that require fewer computational resources. Ref. [14] evaluated these architectures on standard NLP tasks including POS tagging, lemmatisation, and dependency parsing. Their findings indicate that pre-training on larger corpora followed by fine-tuning on Old English outperforms training exclusively on the limited Old English corpus. The adaptation strategies for these models include character-level tokenisation to handle orthographic variation and modified vocabulary representations to accommodate historical language

features. Performance metrics indicate promising results but remain constrained by the limited corpus size available for training.

Fine-tuning strategies include gradient accumulation to compensate for the small batch sizes imposed by the limited corpus, and specialised learning rate schedules that prioritise retention of pre-trained knowledge and accommodate Old English linguistic features. In this respect, character-level tokenisation proves essential for handling orthographic inconsistency, with experiments showing that byte-pair encoding (BPE) tokenisers struggle with the variable spelling conventions of Old English. For instance, when processing forms like *scip/scyp*, 'ship', or *micel/mycel*, 'great', subword tokenisation strategies often fragment these variants inconsistently, which hampers model performance. Experimental results demonstrate that models employing character-level tokenisation achieve 5–8% higher accuracy in lemmatisation tasks compared to standard BPE approaches. Cross-validation experiments demonstrate that pre-training on modern English followed by fine-tuning outperforms models trained exclusively on Old English by approximately 12% across standard evaluation metrics [14].

To the methodological issues discussed above, the lack of standardised evaluation protocols for historical language processing should be added. Unlike modern NLP tasks with established benchmarks (such as GLUE or SuperGLUE), Old English computational work relies on ad hoc evaluation sets that vary across studies. This makes cross-study comparisons difficult and limits reproducibility. Furthermore, the absence of inter-annotator agreement scores for many Old English resources raises questions about annotation quality. For instance, while the YCOE provides syntactic annotation, no published inter-annotator agreement statistics exist for its dependency conversion, which makes it challenging to assess the reliability of training data for machine learning models. Future work should prioritise establishing standardised evaluation protocols and reporting inter-annotator agreement scores to enhance the scientific rigour of computational Old English studies.

6.3. Corpus Linguistics with Ad Hoc Tools

This foundational layer involves custom computational tools designed for specific philological tasks, often developed by scholars to address immediate research questions. These tools prioritise precision over scalability and are typically implemented in scripting languages like Python, Perl or R [37] (p. 543). Early projects like the Microfiche Concordance to Old English employed purpose-built algorithms to index word forms and their contexts [2] (p. 15). Custom scripts also calculate word-frequency distributions across genres or periods and identify statistically significant word pairs. While these tools offer high interpretability, as scholars directly control input parameters and output formats, they are labour-intensive to maintain and lack interoperability with broader NLP pipelines [37] (p. 552).

6.4. From Ad Hoc Tools to AI

The integration of AI into Old English studies has accelerated since the early 2020s. Transformer models like BERT, fine-tuned on Old English corpora, have achieved state-of-the-art performance in tasks such as part-of-speech tagging and named entity recognition. These advances, however, depend on high-quality annotated corpora, which highlights the relevance of resources like the YCOE and DOEC [38] (p. 129). Recent work by Martín Arista et al. [14] has successfully applied modern NLP techniques to Old English, testing three training corpora and three configurations for automatic Universal Dependencies annotation. Their study found that larger training corpora significantly improved performance, and that pre-training the tok2vec stage yielded the best results [14] (p. 5). The three methodologies differ significantly in development time, resource requirements, interpretability and

accuracy. Ad hoc tools are ideal for small-scale studies but lack scalability. NLP libraries balance automation with customisation but are limited to certain tasks. Transformer models achieve the highest accuracy (80–92%) but demand cloud-based infrastructure and suffer from low interpretability [38] (p. 131).

6.5. Natural Language Processing Libraries

The second layer resorts to general-purpose NLP libraries like spaCy, Stanza or Natural Language Tool Kit (NLTK), which offer pre-trained models and standardised annotation frameworks. These tools automate tasks such as tokenisation, lemmatisation and syntactic parsing and allow for customisation for Old English's linguistic features [38] (p. 134). Recent work by Martín Arista et al. [14] has demonstrated the effectiveness of using spaCy for Old English analysis. They implemented a pipeline architecture with several processing stages: tokenisation, tok2vec/transformer, morphological analysis, lemmatisation, and dependency parsing. Their experiments showed that pre-training the tok2vec stage using a large unlabelled corpus significantly improved performance across all tasks, with precision rates between 75% and 90% depending on the evaluation parameter [14] (p. 6). Although these customised pipelines achieve 85% accuracy in lemmatisation, they present lower accuracy, for instance, with poetic compounds like *hronrād*, 'whale-road, sea', which underscores the need for hybrid rule-based/AI approaches [21] (p. 190).

6.6. AI-Driven Solutions with Transformers and LLMs

The most recent layer employs transformer architectures and large language models (LLMs), which learn contextual representations from vast text corpora. These models excel at tasks requiring semantic and pragmatic understanding [38] (p. 136). The type of contextual representation based on word embeddings represent words as dense vectors in a high-dimensional space, where words with similar meanings or grammatical functions cluster together. This allows for sophisticated analyses of semantic relationships and lexical patterns [39] (p. 2). These models can also disambiguate polysemous terms like tīd, which can mean both 'time' and 'hour/festival' depending on context [15] (p. 7). Models like OEC-BERT, pre-trained on the DOEC, generate contextual embeddings for disambiguating polysemous terms. Faulkner [40] has demonstrated that computational approaches can reveal previously unrecognised patterns in scribal practices and orthographic standardisation in eleventh-century manuscripts, which provides a foundation for applying transformer-based models to specific historical contexts. Martín Arista et al. [14] explore using MobileBERT for Old English analysis, noting that standard pre-trained models for modern English were unsuitable due to Old English-specific graphemes (æ, Æ, ð, Ð, þ, Þ). These authors implement a new tokeniser and language model using the MobileBERT architecture. Their experiments show that while this approach performs well, it does not outperform the pre-trained tok2vec model, which suggests that transformer-based models may require significantly more training data to reach their full potential for historical language processing [14] (p. 7). This is a consequence of data scarcity, as the 3-million-word corpus of Old English limits transformer training, and computational costs, given that training a 110M-parameter transformer requires ~100 GPU hours [15] (p. 5).

6.7. Evidence-Based Integration of AI and Philology

The integration of AI capabilities with traditional philological expertise in Old English studies takes place in several forms that bring about measurable benefits. These approaches include: (i) computer-assisted philology, where AI tools largely replace manual analysis; (ii) philologically informed AI, where linguistic expertise guides model development; and (iii) hybrid workflows combining automated and manual processes. The publications reviewed in this entry demonstrate the effectiveness of these integrated approaches. For

example, computer-assisted collation systems have significantly reduced the time required for manuscript comparison and maintained high accuracy rates compared to purely manual methods. In lexical semantic research, computational approaches have identified previously unrecognised meaning relationships in Old English word fields that complement traditional semantic analyses. In syntactic analysis, combined automatic–manual approaches have shown substantial efficiency gains that conform to scholarly standards of quality. The lemmatisation projects discussed in Section 4 demonstrate that hybrid approaches combining computational methods with manual verification can achieve more comprehensive coverage than either approach alone. Similarly, the development of resources like ParCorOE illustrates how computational alignment methods can facilitate traditional philological analysis and maintain scholarly standards. These integrated approaches typically involve automated processing followed by expert validation, which allows scholars to apply computational pattern recognition while preserving philological accuracy. The methodology generally follows a workflow of computational preprocessing, scholarly evaluation, and iterative refinement. While systematic comparative studies remain limited, the projects reviewed in this entry suggest that such integration can enhance both the efficiency and scope of Old English research without compromising scholarly rigour. These outcomes ultimately substantiate the claim that integration of AI and philology represents a productive direction for Old English studies. The effectiveness of this integration depends on balanced methodologies that combine the strengths of computational pattern recognition and contextual philological knowledge while acknowledging the contributions of each approach.

7. Conclusions

The computational study of Old English has evolved from manual corpus annotation to sophisticated AI-driven analysis, fundamentally transforming scholarly engagement with early medieval texts. Foundational resources like the Helsinki Corpus, the DOEC, and the YCOE continue to underpin research and will guide large-scale investigations into syntactic change, lexical semantics, and dialectal variation for a long time. Recent advances, exemplified by the integration of NLP libraries and transformer models as shown in the work of Martín Arista et al. [13], highlight the potential of the field to address longstanding questions in historical linguistics. The emergence of dependency-based approaches, particularly within the Universal Dependencies framework [27–29,31,32], offers new perspectives on Old English syntax that complement traditional constituent-based analyses. However, persistent issues of data scarcity and orthographic variation stress the need for interdisciplinary collaboration. The limited textual corpus of Old English, with approximately 3 million words, poses significant problems for AI-driven analyses, as does the orthographic inconsistency across texts and dialects.

The development of lexical databases like Nerthusv5, which relate corpus data to lexicographical resources, has significantly advanced our ability to conduct comprehensive linguistic analyses of Old English. These databases have facilitated research across multiple linguistic domains, including morphology, syntax, semantics, and lexicography. The lemmatisation project associated with this database has contributed to the understanding of the morphological and spelling variation in Old English.

The computational methods discussed in this entry may help revise certain traditional understandings of Old English. Large-scale corpus analyses have revealed patterns of linguistic variation that may lead to reconsidering the conventional dialectal classifications presented in standard reference works. Computational studies of derivational morphology have identified productive patterns previously unrecognised in traditional grammatical descriptions. The emergence of sophisticated lexical databases may result

in more comprehensive semantic analyses that reveal subtle meaning relationships unavailable in traditional philological approaches. These developments demonstrate that computational methods are not just tools for more efficient analysis but can reshape central aspects of Old English scholarship.

Several specific research directions emerge from this survey that could significantly advance computational Old English studies. First, multimodal approaches combining textual analysis with paleographic and codicological information could provide richer computational models that account for manuscript context. Second, the application of few-shot learning techniques specifically adapted for historical languages could help address data scarcity issues. And third, the field would benefit from collaborative benchmark development, including standardised evaluation datasets, inter-annotator agreement protocols, and shared task competitions similar to those that have accelerated progress in modern NLP. Such infrastructure development would result in more rigorous comparison of computational approaches and faster identification of effective methodologies for historical language processing.

As machine learning methodologies advance, they promise to unlock new dimensions of the linguistic and cultural legacy of Old English that will undoubtedly stress its relevance for the field of digital humanities. The future lies in the integration of computational power with philological expertise, combining the pattern-recognition capabilities of AI with the contextual understanding of historical linguistics. If this balance is kept, computational approaches can enhance rather than replace traditional scholarship and open new avenues for understanding the language and culture of Anglo-Saxon England.

Funding: This research was funded by MCIN/AEI/10.13039/501100011033, grant number PID2023-149762NB-100.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Kytö, M. *Manual to the Diachronic Part of the Helsinki Corpus of English Texts: Coding Conventions and Lists of Source Texts*; University of Helsinki: Helsinki, Finland, 1996; pp. 1, 18–24, 29.
2. Healey, A.D.; Venezky, R.L. *A Microfiche Concordance to Old English*; University of Toronto: Toronto, ON, Canada, 1985; pp. 3, 15.
3. Taylor, A.; Pintzuk, S. *The York-Toronto-Helsinki Parsed Corpus of Old English Prose*; Department of Language and Linguistic Science, University of York: York, UK, 2012; pp. 1, 12–15, 18.
4. Randall, B. *CorpusSearch*; University of Pennsylvania: Philadelphia, PA, USA, 2003.
5. Martín Arista, J.; Domínguez Barragán, S.; Fidalgo Allo, L.; García Fernández, L.; Hamdoun Bghiyel, Y.; Lacalle Palacios, M.; Mateo Mendaza, R.; Novo Urraca, C.; Ojanguren López, A.E.; Ruíz Narbona, E.; et al. *Nerthusv5: Interface of Textual, Lexicographical and Secondary Sources of Old English*; Nerthus Project, Universidad de La Rioja: Logroño, Spain, 2024.
6. Martín Arista, J. The Semantic Poles of Old English: Toward the 3D Representation of Complex Polysemy. *Digit. Scholarsh. Humanit.* **2018**, *33*, 96–111. [CrossRef]
7. Ojanguren López, A.E. Old English perspectives on the complement shift: Toward the desententialisation of self-manipulative verbs. *J. Hist. Linguist.* **2025**, *15*, 44–77. [CrossRef]
8. Ojanguren López, A.E. Are There Serial Verb Constructions in Old English? A New Perspective on the Changes in Verbal Complementation. *Philol. Canar.* **2024**, *30*, 393–423. [CrossRef]
9. Mateo Mendaza, R. The Old English Exponent for the Semantic Prime MOVE. *Aust. J. Linguist.* **2016**, *36*, 542–559. [CrossRef]
10. Novo Urraca, C. Old English Deadjectival Paradigms: Productivity and Recursivity. *NOWELE-North-West. Eur. Lang. Evol.* **2015**, *68*, 61–80. [CrossRef]

11. Vea Escarza, R. Structural and Functional Aspects of Morphological Recursivity. *NOWELE-North-West. Eur. Lang. Evol.* **2012**, *64*, 155–179. [CrossRef]
12. Vea Escarza, R. Old English Adjectival Affixation: Structure and Function. *Stud. Angl. Posnaniensia* **2013**, *48*, 5–25. [CrossRef]
13. Martín Arista, J.; Domínguez Barragán, S.; Fidalgo Allo, L.; García Fernández, L.; Hamdoun Bghiyel, Y.; Lacalle Palacios, M.; Mateo Mendaza, R.; Novo Urraca, C.; Ojanguren López, A.E.; Ruíz Narbona, E.; et al. *ParCorOEv3: An Open Access Annotated Parallel Corpus Old English-English*; Nerthus Project, Universidad de La Rioja: Logroño, Spain, 2023.
14. Martín Arista, J.; Ojanguren López, A.E.; Domínguez Barragán, S. Universal Dependencies Annotation of Old English with spaCy and MobileBERT. Evaluation and Perspectives. *Proces. Leng. Nat.* **2024**, *73*, 253–262.
15. Bosworth, J.; Toller, T.N. *An Anglo-Saxon Dictionary*; Oxford University Press: Oxford, UK, 1973.
16. Sweet, H. *The Student's Dictionary of Anglo-Saxon*; Cambridge University Press: Cambridge, UK, 1976.
17. Clark Hall, J.R. *A Concise Anglo-Saxon Dictionary*, 4th ed.; University of Toronto Press: Toronto, ON, Canada, 1986.
18. Ellis, J. The Interpretative Spelling in Bosworth's Anglo-Saxon Dictionary, Clark Hall's Concise Anglo-Saxon Dictionary, and Sweet's Student's Dictionary of Anglo-Saxon. *Anglo-Saxon Engl.* **1985**, *14*, 187–207.
19. Toller, T.N.; Sean, C.; Tichy, O. (Eds.) *An Anglo-Saxon Dictionary Online*; Faculty of Arts, Charles University: Prague, Czech Republic, 2014; Available online: https://bosworthtoller.com (accessed on 27 August 2025).
20. Healey, A.D. *Dictionary of Old English: A to I Online*; Dictionary of Old English Project; University of Toronto: Toronto, ON, Canada, 2018.
21. Vatri, A.; McGillivray, B. Lemmatization for Ancient Greek: An Experimental Assessment of the State of the Art. *J. Greek Linguist.* **2020**, *20*, 179–196. [CrossRef]
22. Metola Rodríguez, J. Lemmatisation of Old English Strong Verbs on a Relational Database. Ph.D. Dissertation, Universidad de La Rioja, Logroño, Spain, 2015.
23. García Fernández, L. The Lemmatisation of Old English Preterite-Present, Contracted, Anomalous and Class VII Strong Verbs. Ph.D. Dissertation, Universidad de La Rioja, Logroño, Spain, 2018.
24. Tío Sáenz, M. The Lemmatisation of Old English Weak Verbs on a Relational Database. Ph.D. Dissertation, Universidad de La Rioja, Logroño, Spain, 2019.
25. Cichosz, A.; Pęzik, P.; Grabski, M.; Adamczyk, M.; Rybińska, P.; Ostrowska, A. *The VARIOE Online Morphological Dictionary for YCOE*; University of Łódź: Łódź, Poland, 2021.
26. Hamdoun Bghiyel, Y. The Lemmatisation of The York-Toronto-Helsinki Parsed Corpus of Old English Poetry with a Relational Database of Old English Dictionaries. Ph.D. Dissertation, Universidad de La Rioja, Logroño, Spain, 2025.
27. de Marneffe, M.C.; Manning, C.; Nivre, J.; Zeman, D. Universal Dependencies. *Comput. Linguist.* **2021**, *47*, 255–308. [CrossRef]
28. Martín Arista, J. Toward the Morpho-Syntactic Annotation of an Old English Corpus with Universal Dependencies. *Rev. Lingüística Leng. Apl.* **2022**, *17*, 85–97. [CrossRef]
29. Martín Arista, J. Toward a Universal Dependencies Treebank of Old English: Representing the Morphological Relatedness of Un-Derivatives. *Languages* **2024**, *9*, 76. [CrossRef]
30. Villa, L.B.; Giarda, M. Using Modern Languages to Parse Ancient Ones: A Test on Old English. In Proceedings of the 5th Workshop on Research in Computational Linguistic Typology and Multilingual NLP (SIGTYP 2023), Toronto, ON, Canada, 2–6 May 2023; Association for Computational Linguistics: Stroudsburg, PA, USA, 2023; pp. 30–41.
31. Villa, L.B.; Giarda, M. From YCOE to UD: Rule-Based Root Identification in Old English. In Proceedings of the Workshop on Language Technologies for Historical and Ancient Languages (LT4HALA 2024@LREC-COLING-2024), Turin, Italy, 25 May 2024; ELRA Language Resources Association: Paris, France, 2024; pp. 22–29.
32. Celano, G.G.A.; Crane, G.; Majidi, S. Part of Speech Tagging for Ancient Greek. *Open Linguist.* **2016**, *2*, 393–399. [CrossRef]
33. Hellwig, O. Morphological Disambiguation of Classical Sanskrit. In Proceedings of the International Workshop on Systems and Frameworks for Computational Morphology, Stuttgart, Germany, 17–18 September 2015; Springer: Cham, Switzerland, 2015; pp. 41–59.
34. Bamman, D.; Burns, P.J. Latin BERT: A Contextual Language Model for Classical Philology. *arXiv* **2020**, arXiv:2009.10053. [CrossRef]
35. Panagopoulos, M.; Papaodysseus, C.; Rousopoulos, P.; Dafi, D.; Tracy, S. Automatic Writer Identification of Ancient Greek Inscriptions. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *31*, 1404–1414. [CrossRef] [PubMed]
36. Tracy, S.V.; Papaodysseus, C. The Study of Hands on Greek Inscriptions: The Need for a Digital Approach. *Am. J. Archaeol.* **2009**, *113*, 99–102. [CrossRef]
37. Manning, C.; Schutze, H. *Foundations of Statistical Natural Language Processing*; MIT Press: Cambridge, MA, USA, 1999; pp. 543, 552.
38. Singh, P.; Rutten, G.; Lefever, E. A Pilot Study for BERT Language Modeling and Morphological Analysis for Ancient and Medieval Greek. In Proceedings of the SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature, Virtual, 11 November 2021; Association for Computational Linguistics: Stroudsburg, PA, USA, 2021; pp. 128–137.

39. Sprugnoli, R.; Passarotti, M.; Moretti, G. Vir is to Moderatus as Mulier is to Intemperans-Lemma Embeddings for Latin. In Proceedings of the CLiC-it 2019, Bari, Italy, 13–15 November 2019.
40. Faulkner, M. Computational Approaches to Script and Spelling in Eleventh-Century English. *Digit. Scholarsh. Humanit.* **2021**, *36*, 830–848.