Reconstructing the Unwritten: Methodological Advances and Challenges in the Study of Ancient Languages

Alter Frisco*, Charlotte Wilkinson

Philology and Phonetics, Faculty of Linguistics, Bangor University, Bangor, United Kingdom

*Corresponding author: Alter Frisco, alterffc@edu.bg.uk

Abstract

The study of ancient languages, a cornerstone of historical and comparative linguistics, has been profoundly transformed in the 21st century. Moving beyond the traditional confines of philology and the Comparative Method, the field now integrates computational modeling, phylogenetic analysis, and sophisticated socio-linguistic frameworks. This paper provides a critical overview of the methodological evolution in ancient language research. It examines the enduring principles of the Comparative Method and internal reconstruction, highlighting their successes in reconstructing proto-languages like Proto-Indo-European. The paper then delves into the paradigm shift brought about by computational phylogenetics and cladistic models, assessing their power in modeling language divergence and dating proto-language splits, while also addressing critiques regarding their handling of language contact and horizontal transmission. Furthermore, the study explores the burgeoning field of corpus linguistics applied to digitized ancient texts, demonstrating how quantitative analysis can reveal syntactic patterns, semantic shifts, and sociolinguistic variation previously obscured by manual analysis. A significant portion of the discussion is dedicated to the central challenge of fragmentary data and the "unwritten" aspects of ancient languages (e.g., phonetics, register variation, non-standard dialects). The paper argues that a robust, multi-disciplinary approach-combining traditional philological rigor, computational power, and insights from archaeology, genetics, and anthropology-is essential for constructing more nuanced and testable hypotheses about our linguistic past. Three conceptual figures illustrate the comparative phylogenetic model, the feedback loop of multi-disciplinary research, and the spectrum of data completeness in ancient language corpora.

Keywords

Ancient Languages, Historical Linguistics, Computational Phylogenetics, Proto-Language Reconstruction, Language Contact, Indo-European Studies

1. Introduction

The allure of ancient languages is as old as scholarship itself. From the classical grammarians of Greece and Rome to the medieval Arabic scholars who pondered the relationships between Semitic languages, humanity has long sought to understand its linguistic heritage. The formal discipline of historical linguistics, born in the 19th century with the seminal work of scholars like Franz Bopp and Jacob Grimm, established the principle of regular sound change and provided the first systematic framework for relating languages to a common ancestor. This led to spectacular achievements, most notably the detailed reconstruction of Proto-Indo-European (PIE), the hypothesized ancestor of languages from English and Hindi to Greek and Russian.

The primary goal of ancient language research is not merely to decipher scripts or compile dictionaries, but to reconstruct the linguistic systems of the past in their phonological, morphological, syntactic, and semantic dimensions. This endeavor allows us to peer into the cognitive worlds of ancient peoples, to understand their social structures, beliefs, and material culture through the very language they used to describe them. Furthermore, the family trees of languages often serve as proxies for human migration and cultural contact, making linguistic paleontology a valuable, if contentious, tool for archaeologists and geneticists.

However, the 21st century has ushered in both new opportunities and new challenges. The digital revolution has made vast corpora of ancient texts available for computational analysis, while theoretical advances in evolutionary biology have provided models for understanding language divergence and change. Yet, the fundamental problem remains: the data for ancient languages are inherently and often severely fragmentary. We are left with a fossil record composed almost exclusively of written texts, which represent a narrow, formal, and often ideologically charged slice of the full linguistic repertoire [1].

This paper aims to survey the current landscape of ancient language research. It will first revisit the foundational methodologies of the 19th and 20th centuries, acknowledging their strengths and inherent limitations. It will then explore the computational turn, critically assessing the promises and perils of phylogenetic and corpus-based

approaches. A central theme will be the problem of data fragmentation and the strategies researchers employ to mitigate it. Finally, by advocating for a deeply interdisciplinary model, this article will propose a path forward for the field, one that leverages all available tools to reconstruct not just the forms of ancient languages, but the vibrant, spoken realities they once represented.

2. Foundational Methodologies: The Bedrock of Reconstruction

The edifice of ancient language studies rests upon two pillars: the Comparative Method and Internal Reconstruction. These methodologies, developed and refined over two centuries, remain the gold standard for establishing genetic relationships and reconstructing unattested ancestral forms.

2.1 The Comparative Method

The Comparative Method (CM) is a procedure for hypothesizing a common parent language from a set of attested daughter languages. Its core principle is the Neogrammarian doctrine of the *Ausnahmslosigkeit der Lautgesetze*-the exception lessness of sound laws. This posits that sound change is regular and phonetically conditioned; it affects all words in a language that contain the sound in the same phonetic environment, without exception [2].

The process involves:

1.**Cognate Identification:** Assembling sets of words (cognates) from related languages that derive from a common ancestral form (e.g., English *father*, German *Vater*, Latin *pater*, Sanskrit *pitár*-).

2.**Sound Correspondence Establishment:** Identifying systematic patterns of sound relationships (e.g., initial /f/ in English and German corresponds to /p/ in Latin and Sanskrit).

3.**Proto-Form Reconstruction:** Postulating a proto-phoneme for each correspondence set (e.g., */p/ for the above set) and reconstructing the ancestral word (e.g., PIE **ph₂tḗr*).

The success of the CM in reconstructing large parts of the PIE lexicon and grammar is undeniable. It has provided the backbone for the classification of most of the world's language families, from Afro-Asiatic to Sino-Tibetan. However, its limitations are well-known. It works best with a substantial number of well-attested daughter languages. It struggles with situations of intense language contact and borrowing, which can create correspondence sets that mimic genuine genetic relationship [3]. Furthermore, the CM can only take us back in time as far as the deepest reconstructable splits in a family; beyond that, relationships become speculative (e.g., the proposed Nostratic macro-family).

2.2 Internal Reconstruction

Internal Reconstruction (IR) is a method applied within a single language to recover earlier stages of its history by analyzing synchronic irregularities. The underlying assumption is that morphological alternations and anomalies are often the fossilized remnants of regular phonological changes.

A classic example is the reconstruction of PIE laryngeals. The anomalous long vowels and ablaut patterns in Ancient Greek and other Indo-European languages were systematically explained by Saussure (1879) through the postulation of lost "coefficients sonantiques," later confirmed by the discovery of Hittite, an Anatolian language that preserved some of these sounds as laryngeals (/h/-like sounds). This was a triumph of IR, demonstrating its power to predict the existence of unattested phonemes.

IR is particularly valuable for languages with limited or no known relatives, such as Sumerian or Basque. However, it cannot establish genetic relationships between languages, and its results are often chronologically ambiguous, as it is difficult to determine the relative age of different internally reconstructed layers [4].

2.3 The Challenge of Language Contact and Areal Linguistics

While the Comparative Method and Internal Reconstruction provide the bedrock for establishing genetic relationships, their models are primarily linear and vertical, presupposing a clear, branching descent from a common ancestor. A significant methodological challenge arises from the pervasive phenomenon of language contact [5]. Languages do not evolve in isolation; they exist in spatio-temporal contexts where speakers interact, trade, conquer, and coexist, leading to borrowing, convergence, and even the emergence of new, mixed languages.

The Neogrammarian doctrine of exceptionless sound change can be confounded by extensive borrowing. For instance, a set of words showing a systematic sound correspondence might not be cognates but rather a layer of ancient loanwords from a substrate or adstrate language. This is particularly evident in cases like the Uralic substrate in Slavic languages, where certain phonological and lexical features are argued to be the result of language shift rather than inheritance. Distinguishing between deep genetic relationship and long-term contact remains one of the most thorny problems in the field [6]. The proposed "Altaic" family, for example, is now widely regarded by many linguists not as a genetic unit but as a *Sprachbund* (linguistic area), where Turkic, Mongolic, and Tungusic languages have converged typologically over centuries of contact, sharing features like vowel harmony and agglutinative morphology without necessarily sharing a common proto-language.

To address this, researchers have turned to **areal linguistics** and **contact linguistics**. These subfields provide frameworks for identifying features typical of language contact, such as:

•**Structural Borrowing:** The transfer of morphological patterns or syntactic constructions, which is less common than lexical borrowing and often indicates intense, long-term contact. The Balkan *Sprachbund*, where languages from different families (Albanian, Bulgarian, Greek, Romanian) share features like a postposed definite article and the loss of the infinitive, is a classic example [7].

•**Substrate Influence:** The effects of an indigenous language on a newly arrived one, often evident in place names and phonological patterns. The Celtic substrate in Proto-Germanic, for instance, has been hypothesized to explain certain lexical and phonological peculiarities.

Integrating these models with the Comparative Method allows for a more realistic, multi-faceted understanding of linguistic prehistory, acknowledging that language families are not always discrete, tree-like entities but can form intertwined networks of influence.

3. The Computational Turn: Phylogenetics and Corpus Analysis

The advent of powerful computers and sophisticated algorithms has revolutionized many scientific fields, and ancient language studies are no exception. Two approaches, in particular, have garnered significant attention and debate [8].

3.1 Computational Phylogenetics

Inspired by methods in evolutionary biology, computational phylogenetics applies cladistic models to linguistic data to infer family trees and estimate divergence times. Researchers code linguistic features-typically lexical cognates from standardized Swadesh lists or morphological characters-into a matrix. Software packages like BEAST (Bayesian Evolutionary Analysis Sampling Trees) then use probabilistic models to generate phylogenetic trees and calculate confidence intervals for divergence dates [9].

•**Strengths:** This approach offers a quantitative, reproducible, and testable framework for modeling language divergence. It can handle large, complex datasets and provide explicit estimates for proto-language split dates, a long-standing challenge in glottochronology. For instance, the much-cited study by Bouckaert et al. (2012) modeled the expansion of the Indo-European languages and supported an Anatolian origin hypothesis.

•**Challenges and Critiques:** The method is heavily dependent on the quality and coding of the input data. Critics argue that treating languages as perfectly branching trees ignores the pervasive reality of dialect continua and borrowing (wave theory). The models for lexical replacement are often simplistic, and the dating relies on controversial "clock" assumptions [10]. The results can be sensitive to the choice of languages and characters included in the analysis, leading to ongoing debates, such as the Anatolian vs. Steppe hypothesis for PIE origins.

Figure 1 explain this flowchart illustrates the standard pipeline for computational phylogenetic analysis in historical linguistics, from data preparation to the interpretation of results, highlighting its structured, hypothesis-testing nature.

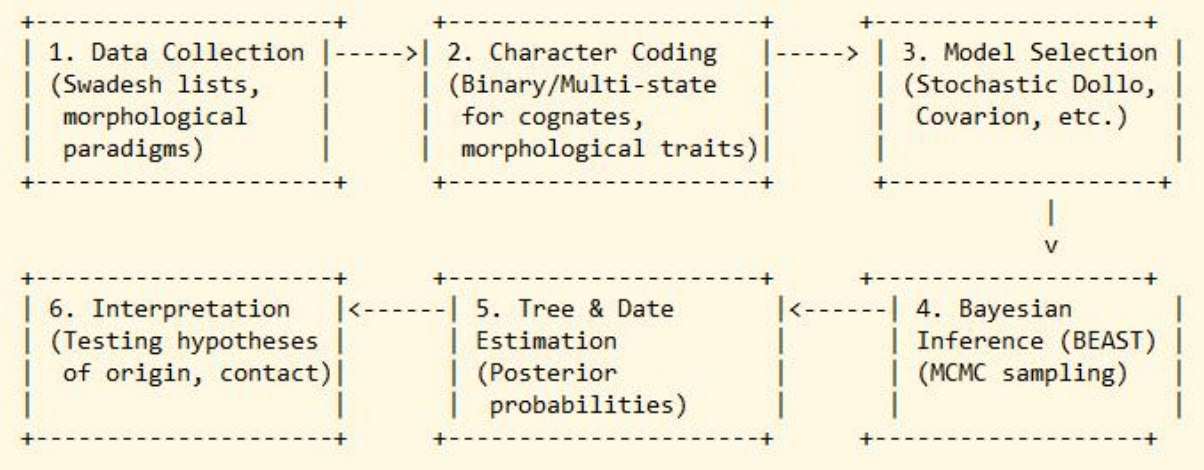

Figure 1. A simplified model of computational phylogenetics in linguistics.

3.2 Digital Corpus Linguistics

The mass digitization of ancient texts-from the epigraphic corpora of the Roman Empire to the manuscript collections of Medieval Europe-has enabled the application of corpus linguistics. This involves the quantitative analysis of large, machine-readable text collections to identify patterns of frequency, collocation, and syntactic construction [11].

•**Applications:**

○**Syntactic Change:** Tracking the decline of the Latin accusative and infinitive construction and its replacement by subordinate clauses introduced by *quod* or *quia*.

○**Semantic Shift:** Mapping the changing meanings of words over time through vector space models applied to co-occurrence data.

○**Socio-linguistics:** Identifying register variation by comparing texts from different genres (e.g., legal documents vs. private letters vs. epic poetry).

•**Limitations:** The results are only as good as the corpus. Biases in the original textual production and preservation are baked into the digital corpus. Texts are often lemmatized and tagged by automated systems that can introduce errors, especially for fragmentary or non-standard texts.

3.3 Computational Semantics and Pragmatic Reconstruction

The digital revolution extends beyond phylogenetics and corpus frequency counts. The burgeoning field of computational semantics applies vector space models and word embeddings to diachronic corpora, enabling the quantitative tracking of semantic change [12]. By analyzing the statistical distribution of words and their co-occurrence patterns across texts from different periods, researchers can map subtle shifts in meaning that might be opaque to manual reading. For example, the trajectory of the Latin word *virtus* (from "manliness, courage" to "moral virtue") can be visualized and quantified, revealing the precise historical period when its collocates (words it frequently appears with) began to shift from a martial to an ethical domain. This approach moves semantic analysis from anecdotal observation to a testable, data-driven hypothesis.

Furthermore, the digitization of diverse text genres opens the door to the nascent field of **historical pragmatics**. This involves reconstructing the unwritten rules of language use in context-the patterns of politeness, discourse markers, and speech acts that characterized ancient communication. While the precise phonetic contour of intonation is lost, its pragmatic function can sometimes be inferred. For example, analysis of particle usage in Classical Greek drama (e.g., the Attic Greek particle *μήν*) can reveal its role in marking affirmation or emotional intensity in dialogue, providing a glimpse into the interactive, spoken nature of the language beyond its written form [13].

The limitations, however, are significant. These computational models are trained on modern languages and may not capture the unique semantic fields or pragmatic structures of ancient tongues. The results are also entirely dependent on the quality and breadth of the underlying corpus; a model trained only on formal legal texts will produce a skewed semantic map of a word compared to one trained on a corpus including poetry, letters, and technical manuals.

4. The Central Challenge: The Tyranny of the Fragmentary Record

Perhaps the most profound issue in ancient language research is the incomplete nature of the data. This "fragmentary record problem" manifests in several ways, each imposing significant constraints on our interpretations.

4.1 The Written Bias

With rare exceptions (e.g., Vesuvius victims), we have no direct access to the spoken language of the past. What survives is the written register, which is typically more conservative, formal, and standardized than everyday speech. The colloquialisms, slang, and phonological variations that characterized spoken interaction are largely lost. For example, our knowledge of Classical Latin is based on the works of elite authors like Cicero and Virgil, while the spoken Vulgar Latin that evolved into the Romance languages must be painstakingly reconstructed from inscriptions, graffiti, and later texts [14].

4.2 Geographic and Social Gaps

Our textual records are often concentrated in political and religious centers. The dialects of rural areas, non-elite groups, and marginalized communities are severely underrepresented. For instance, our understanding of Ancient Egyptian is dominated by the formal hieroglyphs of monuments and the hieratic of administrative texts, while the later Demotic script reveals a different linguistic reality. The recent decipherment of Maya glyphs has shown that our previous understanding of Mayan languages was skewed by a limited set of monumental inscriptions, with new genres revealing unexpected syntactic complexity.

4.3 The Problem of "Unwritten" Features

Many linguistic features leave no trace in writing. Prosody, intonation, and exact phonetic quality are notoriously difficult to reconstruct. While comparative evidence and poetic meter can provide clues, the precise sound of a Homeric epic or a Vedic hymn remains a matter of scholarly conjecture.

4.4 The Challenge of "Dark Data" and Epigraphic Gaps

Beyond the biases of the written record lies the problem of "dark data"-texts that once existed but have been lost to time, and texts that physically exist but remain undeciphered, unedited, or inaccessible. The fragmentary record is not merely

a passive gap but an active filter, shaped by historical accidents, climate, and cultural practices of preservation and destruction.

The epigraphic record is a prime example. Inscriptions on stone or metal are often our most direct contemporary sources, but their survival is haphazard. A single new discovery can overturn long-held theories. The 1994 discovery of the Tel Dan Stele, containing the phrase "House of David," provided the first extra-biblical archaeological evidence of the Kingdom of Judah, demonstrating how a single fragment can reshape historical linguistics and history itself. Conversely, the vast majority of wooden tablets, papyri, and textiles from antiquity have decayed, leaving us with a profoundly skewed sample. The languages of pre-colonial Sub-Saharan Africa, for instance, are almost entirely lost to the archaeological record due to the perishability of their primary writing materials [15].

Moreover, many deciphered texts remain "dark" in a practical sense-they are published in obscure epigraphic corpora, lack modern translations, or are not yet digitized. The process of standardizing and linking these disparate datasets (a field known as digital philology) is a monumental task that is a prerequisite for large-scale computational analysis. Projects like the *Digital Corpus of Literary Papyri* and the *EAGLE* (Electronic Archive of Greek and Latin Epigraphy) network are crucial steps in illuminating this "dark data," but the challenge remains global in scale.

Figure 2 mention this spectrum classifies ancient language corpora based on the volume and diversity of surviving texts, illustrating how the scope of feasible linguistic reconstruction is directly constrained by the nature of the available data.

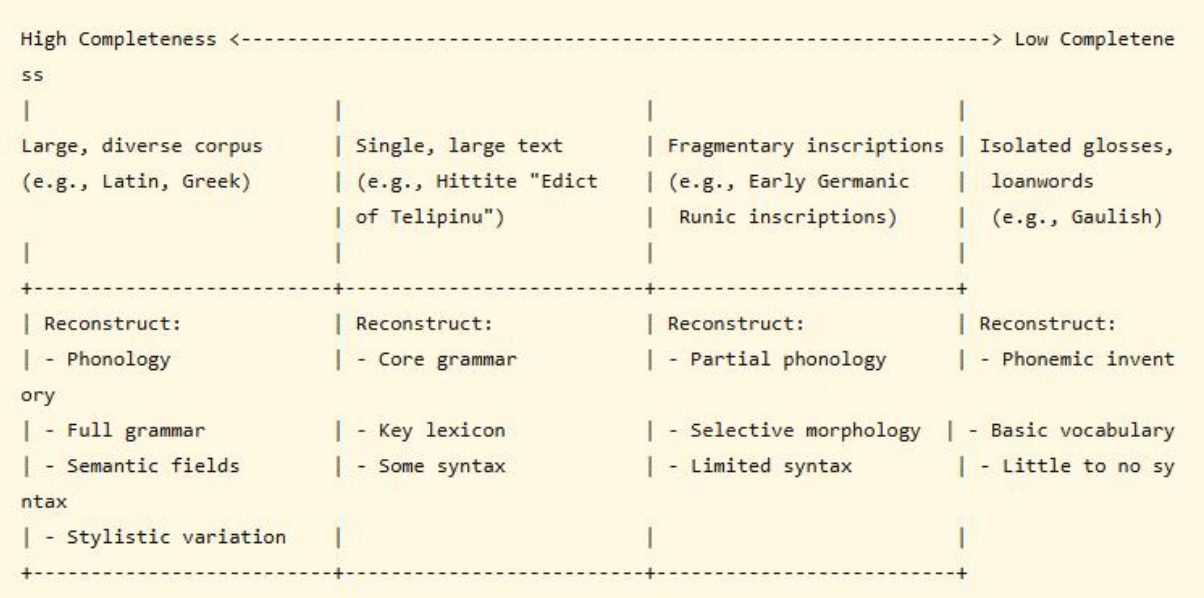

Figure 2. The spectrum of data completeness in ancient language corpora.

5. An Interdisciplinary Path Forward

Given these challenges, the most promising future for ancient language research lies in a conscious and systematic integration with other disciplines. Linguistics alone cannot fill the gaps in the record, but it can form a powerful synergy with archaeology, genetics, and anthropology.

5.1 Linguistics and Archaeology

The correlation of linguistic dispersals with material culture expansions is a classic, though fraught, endeavor. The "New Chronology" for Indo-European, linking its spread to the migration of Kurgan cultures from the Pontic-Caspian steppe, is a prime example. More nuanced approaches now seek to correlate specific lexical reconstructions (e.g., for wheeled vehicles, agriculture, social hierarchy) with archaeological evidence for their invention and diffusion. This can help test and refine both linguistic and archaeological hypotheses.

5.2 Linguistics and Paleogenetics

The explosion of ancient DNA (aDNA) studies provides an independent line of evidence for human migration and population admixture. By comparing phylogenetic trees of languages with genetic trees of populations, researchers can investigate whether language shift or demic diffusion was the primary driver of language spread. For example, aDNA evidence from the Eurasian steppe shows a massive influx of Yamnaya-related ancestry into Central Europe around 3000 BCE, providing strong circumstantial evidence for the steppe hypothesis of Indo-European origins.

Figure 3 for this diagram visualizes the ideal interdisciplinary approach. The core linguistic hypothesis is continuously tested, refined, and constrained by independent evidence from archaeology, genetics, and anthropology, with computational models providing a quantitative framework for integration. Arrows indicate a two-way flow of information and constraint.

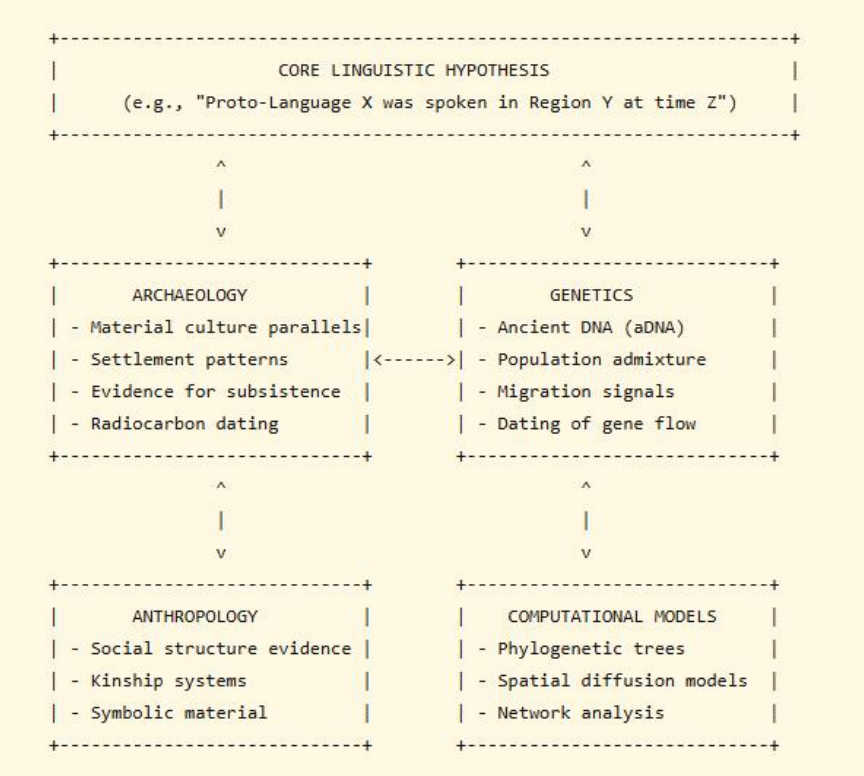

Figure 3. The multi-disciplinary feedback loop in ancient language research.

5.3 A Case for Cautious Optimism: The Anatolian Language Context

The interaction of these disciplines is beautifully illustrated in the study of ancient Anatolia. The Hittite language, deciphered in the early 20th century, provided the first direct evidence for an Anatolian branch of Indo-European. Archaeological excavations at Hattusa yielded the royal archives. Now, genetic studies of skeletons from Anatolian sites are beginning to trace the population dynamics that underpinned the arrival and establishment of Hittite speakers alongside the indigenous Hattic population. No single discipline could have painted this picture alone [16].

5.4 The Role of Epigraphy and Materiality in Linguistic Reconstruction

A truly interdisciplinary approach must also integrate the material study of texts-**epigraphy** and **palaeography**-into the linguistic model. The physical medium of a text is not a neutral container but an integral part of its meaning and linguistic context. The language used in a hastily scratched graffiti on a Pompeian wall is sociolinguistically distinct from the meticulously carved formal dedication on a marble temple frieze, even if they are contemporaneous.

The study of writing materials, scripts, and scribal practices can provide direct linguistic insights. For example:

•**Orthographic Variation:** "Misspellings" or non-standard orthography in informal inscriptions are invaluable for reconstructing phonological changes in Vulgar Latin, such as the loss of final /-m/ and the collapse of vowel length distinctions, which presaged the development of the Romance languages (Väänänen, 1981).

•**Scribal Errors:** The analysis of common scribal mistakes in manuscript traditions can reveal phonological mergers that had occurred in the spoken language of the scribe but were not yet reflected in the literary standard.

•**Material Constraints:** The physical size of a clay tablet or a sheet of papyrus could influence syntactic complexity, potentially leading to a higher frequency of parataxis (clause coordination) over hypotaxis (subordination) in space-constrained texts.

By treating the text as a material artifact, linguists can extract sociolinguistic and phonological data that the text's content alone does not explicitly provide. This bridges the gap between the abstract linguistic system (*langue*) and its concrete instantiation in a specific historical and material context (*parole*).

5.5 Ethical Dimensions and Community Engagement in Ancient Language Studies

The 21st century has also brought a heightened awareness of the ethical responsibilities inherent in studying ancient languages, particularly those of indigenous, colonized, or otherwise marginalized peoples. The field is increasingly

recognizing that these languages are not merely academic data points but are often central to the cultural heritage and identity of descendant communities.

The decipherment of Mayan glyphs, for instance, was a triumph of modern linguistics, but it also raised questions of ownership and interpretation. The knowledge produced by Western academics did not always align with or benefit the living Maya communities. A new paradigm of community philology is emerging, which involves collaborative research with descendant communities, respecting their traditional knowledge and ensuring that research outcomes are accessible and beneficial to them.

This ethical dimension extends to genetic research as well. The extraction of aDNA from ancient human remains must be conducted in consultation with relevant communities and in accordance with strict ethical guidelines. The story of the Ancient One (Kennewick Man) in North America is a cautionary tale about the conflict between scientific curiosity and Indigenous rights. A responsible, interdisciplinary approach to ancient language research must therefore include not only archaeologists and geneticists but also anthropologists, ethicists, and, crucially, community representatives. This ensures that the reconstruction of the linguistic past is not just an academic exercise but a respectful dialogue with the past that acknowledges its living legacies.

6. Conclusion

The study of ancient languages stands at a fascinating crossroads. The traditional tools of the Comparative Method and Internal Reconstruction remain indispensable, providing the foundational proofs of genetic relationship and the initial sketches of proto-languages. However, they are no longer sufficient on their own. The computational turn, despite its ongoing growing pains, offers a powerful new suite of quantitative tools for modeling language change and divergence on a grand scale. Simultaneously, the digitization of texts allows us to analyze linguistic variation with a granularity and scale previously unimaginable.

The future of ancient language research is not in a single methodology, but in a convergent, multi-disciplinary synthesis that is as ethically informed as it is methodologically rigorous. By weaving together, the enduring threads of the Comparative Method with the powerful new tools of computational analysis, and by contextualizing linguistic data within the rich tapestry of archaeological evidence, genetic history, and material culture, we can hope to mute the "tyranny of the fragmentary record." The goal is no longer just to reconstruct a proto-form like **kwón- for 'dog,' but to understand the society of the people who spoke that word, the path their descendants took across continents, the other languages they encountered, and the complex story of contact and change that shaped the languages of the world today. In doing so, we do not merely reconstruct languages; we strive to give voice, however partially, to the unwritten human past.

References

[1] Windfuhr, G., & Windfuhr, G. (Eds.). (2009). The Iranian Languages (1st ed.). Routledge. https://doi.org/10.4324/9780203641736
[2] J.-L. L. Quellec, “Prehistory in North Africa after the Middle Palaeolithic,” in The Cambridge World Prehistory, C. Renfrew and P. Bahn, Eds. Cambridge: Cambridge University Press, 2014, pp. 151–164
[3] Iosif Lazaridis et al., The genetic history of the Southern Arc: A bridge between West Asia and Europe. Science377,eabm4247(2022). https://doi.org/10.1126/science.abm4247
[4] Greenhill SJ, Blust R, Gray RD. The Austronesian Basic Vocabulary Database: From Bioinformatics to Lexomics. Evolutionary Bioinformatics. 2008;4. https://doi.org/10.4137/EBO.S893
[5] Haak, W., Lazaridis, I., Patterson, N. et al. Massive migration from the steppe was a source for Indo-European languages in Europe. Nature 522, 207–211 (2015). https://doi.org/10.1038/nature14317
[6] Pagel, M. Human language as a culturally transmitted replicator. Nat Rev Genet 10, 405–415 (2009). https://doi.org/10.1038/nrg2560
[7] Szemerenyi, Oswald J L, Introduction to lndo-European Linguistics (Oxford, 1997; online edn, Oxford Academic, 31 Oct. 2023), https://doi.org/10.1093/oso/9780198240150.001.0001
[8] Watkins, Calvert, How to Kill a Dragon: Aspects of Indo-European Poetics (New York, NY, 1995; online edn, Oxford Academic, 31 Oct. 2023), https://doi.org/10.1093/oso/9780195085952.001.0001
[9] Dunn, M., Greenhill, S., Levinson, S. et al. Evolved structure of language shows lineage-specific trends in word-order universals. Nature 473, 79–82 (2011). https://doi.org/10.1038/nature09923
[10] Gray, R., Atkinson, Q. Language-tree divergence times support the Anatolian theory of Indo-European origin. Nature 426, 435–439 (2003). https://doi.org/10.1038/nature02029
[11] V. Heyd, “Kossinna’s smile,” Antiquity, vol. 91, no. 356, pp. 348–359, 2017. https://doi.org/10.15184/aqy.2017.21
[12] Y. Matras, Language Contact. Cambridge: Cambridge University Press, 2009. https://doi.org/10.1017/CBO9780511809873
[13] Remco Bouckaert et al. ,Mapping the Origins and Expansion of the Indo-European Language Family.Science337,957-960(2012). https://doi.org/10.1126/science.1219669
[14] J. Clackson, Indo-European Linguistics: An Introduction. Cambridge: Cambridge University Press, 2007. https://doi.org/10.1017/CBO9780511808616
[15] Hock, Hans Henrich. Principles of Historical Linguistics, Berlin, Boston: De Gruyter Mouton, 2021. https://doi.org/10.1515/9783110746440
[16] D. Ringe and J. F. Eska, Historical Linguistics: Toward a Twenty-First Century Reintegration. Cambridge: Cambridge University Press, 2013. https://doi.org/10.1017/CBO9780511980183