Automatic Suggestions of Supplements in the Herculaneum Papyri: Language Models and RESTful APIs

Angelo Mario Del Grosso[1], **Gabriele Giannessi**[2], **Simone Zenzaro**[1], **Federico Boschetti**[1]
[1]Cnr-Istituto di Linguistica Computazionale “Antonio Zampolli”, Italia
[2]Dipartimento di Informatica, Università di Pisa, Italia
{angelomario.delgrosso, simone.zenzaro, federico.boschetti}@cnr.it,
g.giannessi4@studenti.unipi.it

Abstract

This paper addresses a computational philology task focused on the automatic restoration of textual gaps (i.e., lacunae) in the Herculaneum Papyri, whose Ancient Greek texts are inherently fragmentary due to damage caused by carbonization. The objective of this work is to show the preliminary results concerning the development of a web-based suggestion service for proposing plausible supplements to fill lacunae, thereby supporting the philological process of producing new critical editions within a new web-based digital scholarly editing environment. To automatically provide such suggestions, we have developed systems that generate textual supplements in Ancient Greek, employing both neural (BERT-like) and statistical (n-gram) language modeling approaches.

Keywords: Digital Humanities; Tools, Systems, Applications; Training, Fine-tuning, Adaptation, Alignment, and Representation Learning; Web Services

1. Introduction

Language Models (LMs) are probabilistic frameworks that assign likelihoods to sequences of words, allowing prediction of upcoming words or entire sentences from context, and as such, they represent powerful machine learning systems capable of interpreting, understanding, and generating natural language (Jurafsky and Martin, 2025). Some Natural Language Processing (NLP) tasks share objectives with various Digital Humanities (DH) activities, such as dating, authorship attribution, and text restoration (Assael et al., 2022). Similar methods are also used by papyrologists to reconstruct damaged texts from ancient papyri. In fact, Masked Language Modeling (MLM), also called *fill-mask*, involves providing the language model with a tokenized input sequence in which about 10%–15% of the tokens are masked. The model must then predict, for each masked token, a probability distribution over the vocabulary to determine which token is the best replacement for the missing word. Figure 1 illustrates an example of a MLM in use, specifically the "FacebookAI/xlm-roberta-base" model (Conneau et al., 2019).

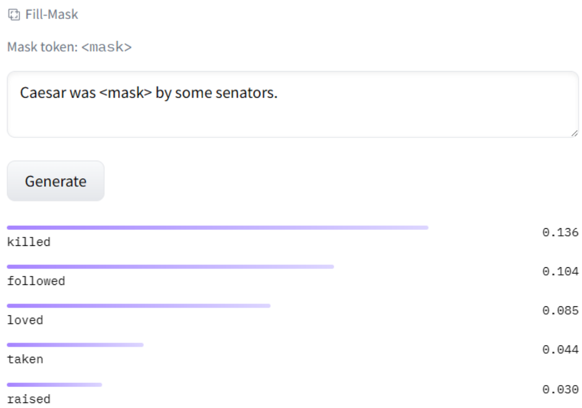

Figure 1: Providing the input sentence "Caesar was <mask> by some senators." to the "xlm-roberta-base" model. "Killed" is the predicted token

Given this alignment between MLM techniques and philological restoration needs, this work leverages these methods within a specific research context, namely the ERC Advanced Grant `885222` project (PI Prof. Graziano Ranocchia) (Del Grosso et al., 2023). The `GreekSchools` project aims to produce a new critical edition of the "*Arrangement of the Philosophers*" originally written by the Epicurean philosopher Philodemus, which has come down to us through the Herculaneum Papyri.[1] The ambitious goal of restoring the oldest existing history of Greek philosophy in Europe presents a number of challenges, ranging from the reading of the Greek texts in the papyri and their material hidden layers by means of non-invasive imaging techniques (Romano et al., 2023), to the digitization of the texts (Del Grosso et al., 2025).

Within the context of the `GreekSchools` project, we developed two tools: a textual scholarship environment called `CoPhiEditor` (Fig. 2) (Zenzaro et al., 2022) and the scholarly editions visualization platform called `CoPhiViewer`.[2] The system

[1]More information about the `GreekSchools` project is available at `https://greekschools.eu/`
[2]`https://viewer.greekschools.eu/`

Proceedings of the Fifteenth Language Resources and Evaluation Conference (LREC 2026), pages 590–598
11–16 May 2026. ©ELRA Language Resources Association (ELRA), 2026

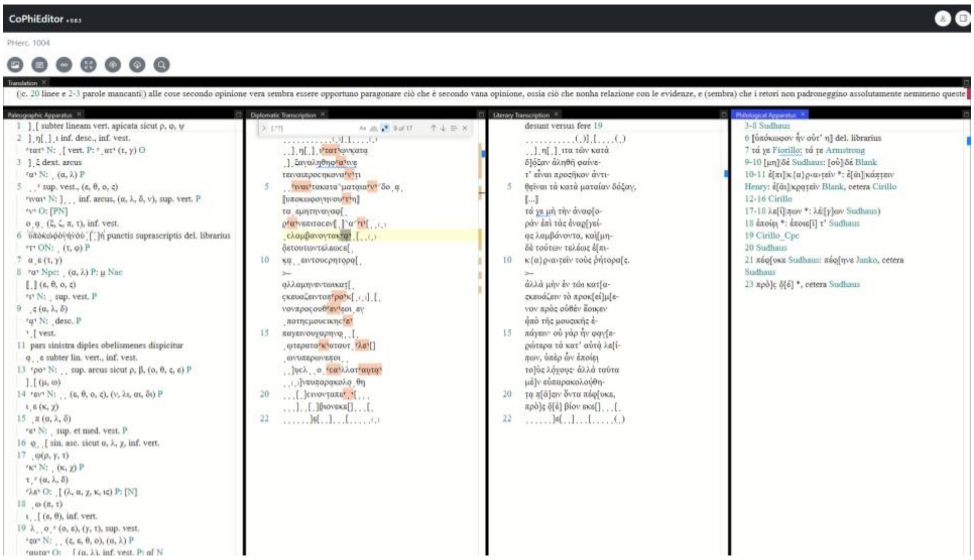

Figure 2: The `CoPhiEditor` Web Platform for Digital Scholarly Editing of the Herculaneum Papyri.

exploits a Domain-Driven Design (DDD) (Evans, 2004) approach based on Domain-Specific Languages (DSL) to offer a collaborative and cooperative environment among papyrologists to facilitate the production of digital scholarly editions. The environment provides suggestions for editing activities. This work addresses the creation of customized language models for philological supplements and describes the RESTful API service that makes integration with `CoPhiEditor` possible. While initially tested on the Herculaneum Papyri and Ancient Greek, our workflow is highly adaptable. Given suitable training data and/or language models, this methodology seamlessly transfers to other languages and scholarly traditions. To foster a broader application, the underlying editing environment will be released under an open-source license, allowing other research groups to adapt the tool for their own philological domains.

2. Related Works

Currently, various NLP tools are available that simplify operations such as the retrieval of classical document collections and language modeling.

NLTK (Bird, 2006) is a widely used toolkit for textual analysis. It provides a broad range of features, including language models, metrics, and tools for different stages of text processing, such as tokenization, POS tagging, and lemmatization. In addition to NLTK, CLTK (Johnson et al., 2021) enables the analysis of documents written in ancient languages, offering a library with pipelines and models for approximately 20 languages, including Ancient Greek. Complementing these tools, spaCy (Honnibal et al., 2020) is a modern NLP library optimized for industrial applications. This library offers a third-party module for Ancient Greek.[3]

Several systems have been developed to address tasks such as the textual restoration of ancient texts. Although traditional n-gram models (Jurafsky and Martin, 2025) established an early baseline, their performance has been greatly exceeded by neural network architectures (Mikolov et al., 2010).

Large Language Models (LLMs) based on Transformers (Vaswani et al., 2017) have recently achieved promising results: Pythia (Assael et al., 2019) and its successor Ithaca (Assael et al., 2022) not only perform text restoration but also assign geographical and chronological contexts to inscriptions/epigraphic texts.

BERT-like models such as AristoBERTo (Myerston, 2022), GreBERTa (Riemenschneider and Frank, 2023), and Logion (Cowen-Breen et al., 2023) are trained on Ancient Greek corpora, leveraging BERT's native Masked Language Modeling (MLM) (Devlin, 2018) to predict missing words.

Models such as GreTa (Riemenschneider and Frank, 2023) and Philo-1-Preview (Ferrara, 2025) tackle text restoration by integrating denoising strategies into their generative process to reconstruct corrupted or partially observed inputs. This approach significantly improves robustness to noisy, distorted, or incomplete text. Importantly, such denoising objectives are orthogonal to the choice of model architecture and can be paired with both encoder-only MLM models and encoder–decoder seq2seq models.

[3] `https://spacy.io/universe/project/grecy`

3. Data

To create language models, a dataset composed of texts from Ancient Greek corpora must be compiled. The MAAT dataset (Fitzgerald and Barney, 2024)[4] was used to convert the data available in the corpora from EpiDoc format (Bodard, 2010) into a *machine-actionable* format (Fitzgerald and Barney, 2024). For this work, the MAAT corpus was specifically enriched with the Perseus corpus (Smith et al., 2000), the First1KGreek corpus (Crane et al., 2014), and the new digital editions made available within the `GreekSchools` project[5], as these corpora are similar in nature to the texts within the historical-literary domain represented by the Herculaneum papyri. After collection, a pre-processing phase is carried out to normalize the text content, handle gaps and editorial conventions, including normalization of diacritics, removal of modern editorial marks, and standardization of textual variants, as well as to define the sentence splits used for training, development, and testing. At this stage, sentences are extracted from the processed texts to create the training and validation sets. To optimize the models' hyperparameters, a held-out *development* set consisting of sentences from the MAAT corpus specifically belonging to the Herculaneum papyri was created. The final evaluation is conducted on a *test* set of previously unpublished-in-print examples (released in our dataset)[6] selected directly from the digital editions of papyrus texts available on the `CoPhiViewer` platform. These datasets have been made publicly available on both the GitHub[7] and HuggingFace[8] platforms. The overall composition of the datasets in terms of tokens and sentences distribution across the training, development, and test sets is summarized in Table 1. The limited size of the test set (150 sentences) is due to the limited availability of new scholarly editions produced within `GreekSchools` project; efforts to expand this sample are already underway.

[4]https://github.com/WMU-Herculaneum-Project/maat
[5]https://github.com/CoPhi/gs-public-editions/releases
[6]https://github.com/CoPhi/gs-suggestions-dataset/blob/57d3a93c62f429251c795b78cf4c6ab71eae2ce1/test_abs.csv
[7]https://github.com/CoPhi/gs-suggestions-dataset/tree/57d3a93c62f429251c795b78cf4c6ab71eae2ce1/data
[8]https://huggingface.co/datasets/CNR-ILC/gs-maat-corpus

Set	Tokens	Sentences
Training	17.208.506	93.172
Development	688.971	4.640
Test	10.068	150
Total	**17.907.545**	**97.962**

Table 1: Distribution of tokens and sentences across the training, development, and test sets in the compiled dataset.

4. Models and Services

In order to address the issue of restoring the `GreekSchools` project's damaged texts and to develop related restoration services, both language models and RESTful APIs have been implemented. We designed the service input by representing each damaged word in square brackets, preserving the readable letters and using a dot to mark each missing or illegible character (e.g., [κατ.σκευαζειν]). The system is expected to output the full word with the missing letters restored (e.g., κατασκευαζειν). Two principal approaches have been adopted for the language models: (a) the statistical approach, based on n-gram language models (Jurafsky and Martin, 2025) that estimate the probability of a word from a preceding context window; and (b) the neural approach, based on Transformer architectures (Vaswani et al., 2017), capable of processing sequences more flexibly and learning richer linguistic representations.

To represent the statistical approach, the n-gram models were chosen for two main reasons: (i) to provide a baseline; and (ii) to compare their performance with more sophisticated approaches, which can be competitive in low-resource settings such as for the Ancient Greek language. In addition to serving as a baseline, n-gram models were included for their high degree of interpretability, a valuable feature for scholars seeking to understand the rationale behind a given suggestion.

We train two n-gram models:

- a global model m_g, trained on the entire *training* set;
- a more specialized model m_d, trained only on the Herculaneum papyri included in the training set, which constitutes a subset of the full corpus.

The creation of the specialized model aims to adapt the probability distributions to the domain of the Herculaneum papyri (domain adaptation). In this context, the probabilities produced by the system are calculated through a linear interpolation scheme (Jurafsky and Martin, 2025) that weights the domain-specific model more heavily (controlled by the λ parameter).

Given the target word w and the context c, the probability is computed as follows:

$$P_{\text{interp}}(w \mid c) = \lambda \cdot P_{m_d}(w \mid c) + (1 - \lambda) \cdot P_{m_g}(w \mid c) \quad (1)$$

For the n-gram models, a search algorithm was implemented following the *beam search* decoding strategy to find the best supplements based on a heuristic (see Figure 3). The heuristic combines the model score computed over the preceding context window with constraints extracted from the incomplete string, namely the preserved prefix (head) and suffix (tail) of the damaged word, as well as the expected length of the proposed supplement.

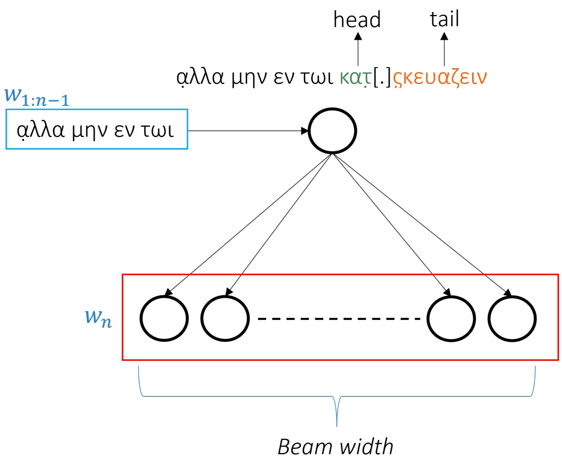

Figure 3: Basic diagram illustrating beam search decoding strategy. Supplements are generated from context window $w_{1:n-1}$, with candidate selection based on available information including the head and tail of the lacuna (shown in square brackets).

For the neural approach, BERT models were considered for two key reasons: 1) BERT's pre-training task (Masked Language Modeling) aligns directly with the restoration objective, namely generating supplements to fill gaps in damaged passages; and 2) due to design constraints imposed by the `GreekSchools` project, computational resource usage must be limited, and BERT, in its base version with 110M parameters, offers an excellent balance between efficiency and computational cost.

A fine-tuning process was conducted on BERT models tailored for Ancient Greek. AristoBERTo (Myerston, 2022) is initialized with the weights of GreekBERT (Koutsikakis et al., 2020) and further pre-trained on an Ancient Greek corpus of approximately 900 MB obtained through web scraping. GreBerta (Riemenschneider and Frank, 2023) is a variant of RoBERTa (Liu et al., 2019) adapted for Ancient Greek, while Logion (Cowen-Breen et al., 2023) is trained on a corpus of over 70 million words in Modern Greek and features an extensive vocabulary (50K tokens).

Fine-tuning, specifically *continual pre-training* (Jurafsky and Martin, 2025), was performed on AristoBERTo, GreBerta, and Logion using the *training set* discussed in Section 3, over a total of 10 epochs. Figure 4 illustrates the reduction in overall loss, calculated at the end of each epoch, for both the *training set* and the *development set*. The fine-tuning processes were carried out on an NVIDIA A10 Large GPU. Table 2 shows the duration for each BERT model.

BERT Model	**Time (min)**
AristoBERTo	1h 43m
GreBerta	2h 48m
Logion	1h 32m

Table 2: Fine-tuning duration (rounded to the nearest minute) for the BERT models adapted for Ancient Greek.

In addition to developing MLM models, a RESTful API has been implemented and deployed to provide microservices that can be integrated into the `CoPhiEditor` environment.[9] The microservice relies on the Python FastAPI framework and the OpenAPI specification, utilizing the integrated Swagger UI to expose technical documentation. Figure 5 shows the current release of the API, which includes endpoints for creating, deleting, and retrieving models, either all available models or a specific one by its identification number, as well as an endpoint for invoking the prediction service related to the fill-mask task. Figure 6 illustrates the parameters used by the prediction service. The parameters are: (1) model ID, (2) text fragment with gap (dots indicate missing characters), (3) number of provided tokens, and (4) expected number of tokens to generate (n-gram models only).

5. Results and Discussion

We evaluate the system by comparing generated suggestions against a specific printed edition (gold standard). A prediction is counted as correct only if it exactly matches the reference; therefore, plausible alternative supplements are counted as errors in the automatic evaluation. In this framework, automatic metrics quantify agreement with the reference edition rather than philological plausibility; we therefore complement them with expert assess-

[9]The current development prototype is available at: `https://6016.greekschools.eu/doc` and codebase is available at: `https://github.com/CoPhi/gs-suggestions-dataset.git`

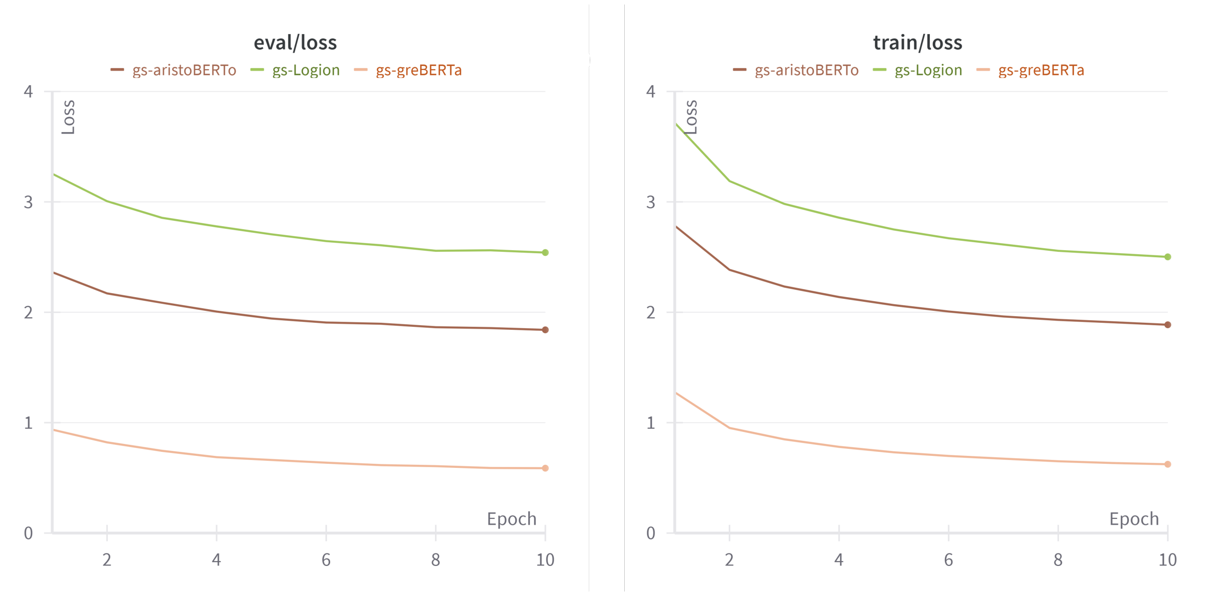

Figure 4: Loss curve during fine-tuning.

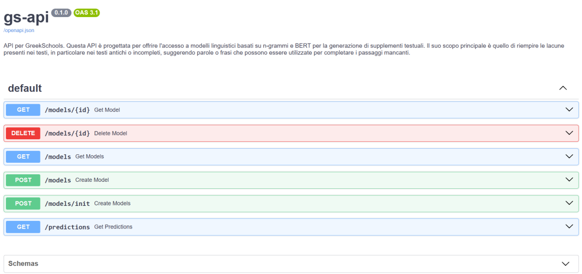

Figure 5: OpenAPI-compliant Swagger UI documentation for the text restoration task in the `GreekSchools` project.

Figure 6: OpenAPI-compliant Swagger interface of the Prediction Service, including its configurable parameters.

ment in future work, including an analysis of inter-annotator agreement to ensure consistency.

For evaluation, two metrics are used to assess the quality of language models: cross-entropy loss and top-K accuracy.

Cross-entropy loss is an objective function employed by learning algorithms such as gradient descent (Jurafsky and Martin, 2025) to guide model optimization during training. In this context, it measures how well the model predicts probability distributions over the masked tokens M in a sentence composed of a set of tokens (*embeddings*) $\mathbf{x} = \{x_1, \ldots, x_n\}$.

Given M, the cross-entropy loss is calculated as follows:

$$\mathcal{L}_{CE} = -\frac{1}{M} \sum_{i \in M} \log \hat{y}_i \tag{2}$$

In Equation 2, $\hat{y}_i$ represents the probability assigned by the language model to the token x_i.

Top-K accuracy is particularly well-suited to text restoration in philological contexts, where multiple plausible completions may exist for a given gap, thereby leaving the philologist to determine or conjecture the most appropriate reconstruction.

For instance, the following are two cases of successful conjectural suggestions: μαχομένων (machoménōn) for [...]ομενω[.] ([...]omenō[.]) as a conjectural reconstruction in PHerc. 1004, col. 36, line 12 (Fig. 7),[10] and ἀνάγκην (anánkēn) for [...]γκ[..] ([...]nk[.]) in PHerc. 1020, col. 103, lines 20–21 (Fig. 8).[11]

The context of [...]ομενω[.] ([...]omenō[.]), in interpretive transcription, is: ὅτι δυεῖν οὐσῶν βοηθειῶν [...]ομένω[.] δ᾽ ἀλλήλαις (hóti dyeîn ousôn boētheiôn [...]oménō[.] d’ allḗlais), an excerpt that means: “[...] because, since there are two auxiliaries (scil. philosophy and rhetoric), yet [...]-ing with each other [...]”.

The sequence -ομένω[.] (-oménō[.]) is most probably part of a participial *suffix* agreeing with βοηθειῶν (boētheiôn), the genitive plural of the noun βοήθεια (boḗtheia), “aid, auxiliary troops”.

[10] https://viewer.greekschools.eu/editions/col-36-pherc-1004

[11] https://viewer.greekschools.eu/editions/col-103-pherc-1020

The suggested supplement (Fig. 7) [μαχ]ομένω[ν] ([mach]oménō[n]), "fighting, contending", as proposed in the conjectural reconstruction by Essler, is well grounded, as it satisfies both the syntactic (present participle, genitive feminine plural) and semantic-pragmatic (a military metaphor within a philosophical context) requirements of the passage.

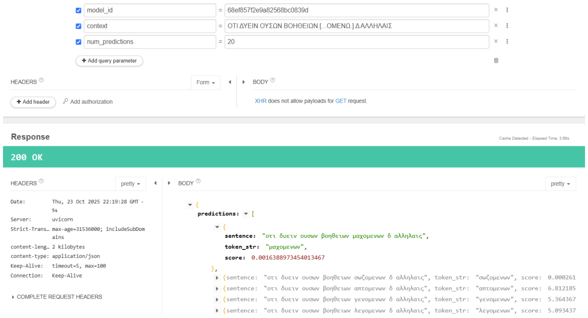

Figure 7: Greek supplement suggestions for [μαχ]ομενω[ν] in οτι δυειν ουσων βοηθειων <mask> δ αλληλαις (in capital letters).

The context of [...]γκ[..] ([...]nk[.]), in interpretive transcription, is: καθὸ ἀτύφως φύσει ἀνθρωπικῆι κατ' [..]γκ[..] χωρὶς ὑπαντᾶι τῶν ἀγαθῶν ἐξαπατωμένων μηδ'έν (kathò atýphōs physei anthrōpikêi kat' [..]nk[..] chōrìs hypantâi tōn agathôn exapatōménōn m<ēd>én), an excerpt that means: "because, without arrogance, it agrees with human nature, by necessity in an independent way, since good people are deceived in nothing". The suggested supplement (Fig. 8) ἀνάγκην ([aná]nk[ēn]), is the accusative of ἀνάγκη (anánkē) "necessity", as proposed by Alessandrelli-Ranocchia. The conjecture is syntactically appropriate in the context and semantically compatible with the idea of an intrinsic order of human nature.

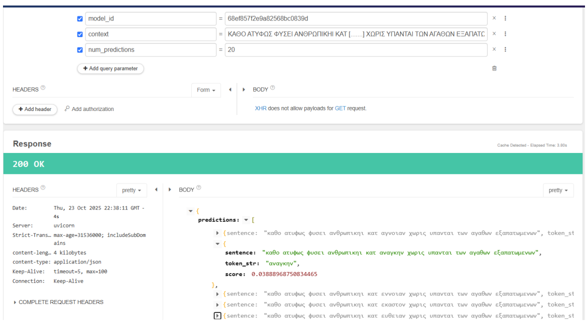

Figure 8: Greek supplement suggestions for αναγκην in καθο ατυφως φυσει ανθρωπικηι κατ <mask> χωρις υπανται των αγαθων εξαπατωμενων (in capital letters).

Notably, Morpheus[12] generates over 130 potential Ancient Greek inflected forms to fill the gaps for [...]ομενω[.] ([...]omenō[.]): e.g. ἀγρομένων (agroménōn), λεγομένωι (legoménōi), and φερομένων (pheroménōn) and the inflected forms candidate to fill the gaps for [...]γκ[..] ([...]nk[.]) are more than 70, e.g. ἐνέγκει (enénkei), πλαγκτή (planktḗ), σκίγκων (skínkōn), words syntactically and semantically unrelated to the context. This demonstrates that our conjecturer acts effectively to filter out the manifold inflected forms compatible with the gaps to fill.

The Top-K accuracy measures whether the gold token (label) appears among the K suggestions returned by the RESTful API (Fig. 7 and Fig. 8). This aligns closely with the needs of philological analysis, as we discussed throughout the two examples above.

Since BERT's architecture does not natively support character-level constraints on a *masked* word, the system applies a post-processing filter. The API retrieves the top-K predictions from the BERT model and cross-references them with a morphological lexicon to retain only those forms that match the visible head, tail, and character length of the lacuna. This ensures that the high-level semantic capabilities of the Transformer-based model are constrained by the material evidence of the papyrus.

In the n-gram models, the configurations for training are shown in Tables 3 and 4. Additionally, Figure 9 shows preliminary results obtained from the evaluation of the top-K accuracy metric on language models. In particular, BERT models adapted through the fine-tuning process are denoted by the prefix 'gs', which indicates that the models belong to the `GreekSchools` project (Del Grosso et al., 2023). The BERT models outperform the n-gram models with the configurations chosen in Tables 3 and 4. This improvement is further confirmed by cross-entropy loss values (Table 5), where fine-tuned models (gs-prefix) consistently show lower loss compared to their base versions.

Hyperparameter	**Value**
Dimension (N)	3
λ	0.2
α	1
β	1
δ	1

Table 3: Configuration of the hyperparameters of the n-gram model with MLE scoring type.

[12]Morpheus is a morphological parser and form generator developed by Gregory Crane at the Perseus Project, available at https://github.com/perseids-tools/morpheus; derivative lists from a dump of Morpheus are available in https://github.com/CoPhi/grc-words

Hyperparameter	Value
Dimension (N)	3
λ	0.2
α	1
β	1
δ	1
γ	0.001

Table 4: Configuration of the hyperparameters of the n-gram model with smoothing (Add-K smoothing) as the scoring type.

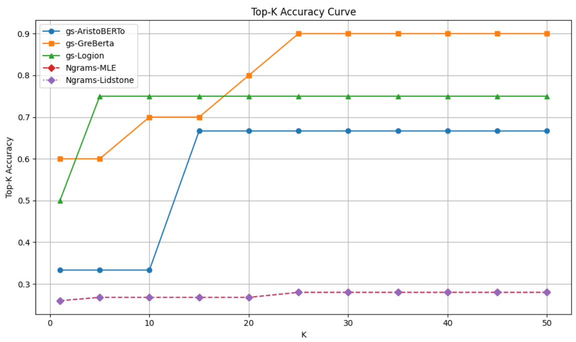

Figure 9: Graph of top-K accuracy calculated on language models created by varying the K parameter.

6. Conclusion and Future Work

This contribution addressed the problem of supplement generation, formulated as a machine learning fill-mask task. Several language models were developed to generate textual suggestions for missing portions of text. As a baseline, n-gram models were implemented alongside three BERT-based language models trained on Ancient Greek corpora. The ongoing development phase includes the deployment of a web service that provides access to these models through the `CoPhiEditor` platform, offering philologists an effective tool for textual restoration. The involvement of domain experts in testing has facilitated the creation of new and more refined models.

Current evaluation demonstrates that the BERT models achieve substantially higher performance in terms of top-K accuracy. In particular, fine-tuning has enabled these models to adapt specifically to ancient texts, thereby improving the quality of the generated supplements. By contrast, the n-gram models exhibit limitations inherent to their statistical approach.

Although the results are promising, the evaluation presented here requires further validation due to the limited size of the current test set. Consequently, future work will focus on building a larger test set, including texts extracted from the new edition of the Herculaneum papyri, to enable more

Model	$\mathcal{L}_{CE}$
AristoBERTo	3.06
gs-aristoBERTo	**2.39**
Logion	4.43
gs-logion	**3.61**
GreBerta	3.18
gs-greBerta	**2.44**

Table 5: Comparison of the cross-entropy loss metric value calculated on the test set. Models with the prefix ‘gs’ represent models specialized on the dataset through fine-tuning.

robust assessment.

In future work, we will also complement model-centric metrics with a human-centered evaluation that directly measures the tool’s impact on philological practice (A/B testing). We plan a controlled between-subjects study in which two groups of philologists perform the same restoration tasks on identical texts: a baseline group working only with traditional resources and an experimental group working with both traditional resources and the suggestion tool. Participants will be randomly assigned to conditions and matched for experience level to ensure group comparability. An independent adjudication panel, blind to the experimental condition associated with each response, will assess the scholarly plausibility and overall quality of the proposed conjectures against predefined criteria; inter-rater agreement will be measured via Cohen’s κ. Operational efficiency will be quantified through time-to-completion and interaction/iteration counts, though care will be taken not to interpret faster completion as a quality indicator in the absence of a corresponding measure of critical engagement with the system’s suggestions. Finally, we will systematically vary the top-K size returned by the system to determine whether increasing the number of candidates improves, stabilizes, or degrades overall performance in terms of both conjecture quality and cognitive load.

7. Acknowledgments

This work was supported by the `GreekSchools` project, The project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020, Excellent Science (Grant agreement No. `885222`, PI: Graziano Ranocchia). The authors also thank Gloria Mugelli and Alessia Oliverio for their valuable assistance in the evaluation task.

8. Bibliographical References

Yannis Assael, Thea Sommerschield, and Jonathan Prag. 2019. Restoring ancient text using deep learning: a case study on greek epigraphy.

Yannis Assael, Thea Sommerschield, Brendan Shillingford, Mahyar Bordbar, John Pavlopoulos, Marita Chatzipanagiotou, Ion Androutsopoulos, Jonathan Prag, and Nando De Freitas. 2022. Restoring and attributing ancient texts using deep neural networks. *Nature*, 603(7900):280–283.

Steven Bird. 2006. Nltk: the natural language toolkit. In *Proceedings of the COLING/ACL 2006 interactive presentation sessions*, pages 69–72.

Gabriel Bodard. 2010. Epidoc: Epigraphic documents in xml for publication and interchange. *Latin On Stone: epigraphic research and electronic archives*, pages 101–18.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Unsupervised cross-lingual representation learning at scale. *CoRR*, abs/1911.02116.

C. Cowen-Breen, C. Brooks, J. Haubold, and B. Graziosi. 2023. Logion: Machine learning for greek philology.

Gregory Ralph Crane, Monica Berti, Annette Geßner, Matthew Munson, and Tabea Selle. 2014. Open greek and latin project.

Angelo Mario Del Grosso, Simone Zenzaro, Federico Boschetti, and Graziano Ranocchia. 2023. Greekschools: Making traditional papyrology machine actionable through domain-driven design. In *2023 7th IEEE Congress on Information Science and Technology (CiSt)*, pages 621–626. IEEE.

Angelo Mario Del Grosso, Simone Zenzaro, Federico Boschetti, and Graziano Ranocchia. 2025. Bridging Traditional and Digital Papyrology with Domain-Specific Languages: The GreekSchools Case Study. In Nicola Reggiani, editor, *The Digital Critical Edition of Greek Papyri: Issues, Projects, and Perspectives*, pages 125–152. De Gruyter, Berlin, Boston.

Jacob Devlin. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Eric Evans. 2004. *Domain-driven design: tackling complexity in the heart of software*. Addison-Wesley Professional.

Giuseppe Ferrara. 2025. Philo-1-preview: Un modello T5-Base per l'emendazione dei testi antichi. In *Atti del XIII Convegno Annuale AIUCD 2025*.

Will Fitzgerald and Justin Barney. 2024. A new machine-actionable corpus for ancient text restoration. In *Proceedings of the 1st Workshop on Machine Learning for Ancient Languages (ML4AL 2024)*, pages 56–60, Hybrid in Bangkok, Thailand and online. Association for Computational Linguistics.

Matthew Honnibal, Ines Montani, Sofie Van Landeghem, and Adriane Boyd. 2020. spaCy: Industrial-strength Natural Language Processing in Python.

Kyle P Johnson, Patrick J Burns, John Stewart, Todd Cook, Clément Besnier, and William JB Mattingly. 2021. The classical language toolkit: An nlp framework for pre-modern languages. In *Proceedings of the 59th annual meeting of the association for computational linguistics and the 11th international joint conference on natural language processing: System demonstrations*, pages 20–29.

Daniel Jurafsky and James H. Martin. 2025. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, 3rd edition. Prentice Hall. Online manuscript released January 12, 2025.

John Koutsikakis, Ilias Chalkidis, Prodromos Malakasiotis, and Ion Androutsopoulos. 2020. Greek-bert: The greeks visiting sesame street. In *11th Hellenic Conference on Artificial Intelligence*, SETN 2020, page 110–117. ACM.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach.

Tomas Mikolov, Martin Karafiát, Lukas Burget, Jan Cernockỳ, and Sanjeev Khudanpur. 2010. Recurrent neural network based language model. In *Interspeech*, volume 2, pages 1045–1048. Makuhari.

Jacobo Myerston. 2022. aristoberto.

Frederick Riemenschneider and Anette Frank. 2023. Exploring large language models for classical philology. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 15181–15199, Toronto, Canada. Association for Computational Linguistics.

Francesco P. Romano, Enzo Puglia, Claudia Caliri, Danilo P. Pavone, Michele Alessandrelli, Andrea Busacca, Claudia G. Fatuzzo, Kilian J. Fleischer, Carlo Pernigotti, Zdenek Preisler, Christian Vassallo, Gertjan Verhasselt, Costanza Miliani, and Graziano Ranocchia. 2023. Layout of ancient Greek papyri through lead-drawn ruling lines revealed by Macro X-Ray Fluorescence Imaging. *Scientific Reports*, 13(1):6582.

David A Smith, Jeffrey A Rydberg-Cox, and Gregory R Crane. 2000. The perseus project: A digital library for the humanities. *Literary and Linguistic Computing*, 15(1):15–25.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need.

Simone Zenzaro, Angelo Mario Del Grosso, Federico Boschetti, Graziano Ranocchia, et al. 2022. Verso la definizione di criteri per valutare soluzioni di scholarly editing digitale: il caso d'uso greekschools. *Copyright© 2022 AIUCD Associazione per l'Informatica Umanistica e la Cultura Digitale*, page 20.

9. Language Resource References

Fitzgerald, Will AND Barney, Justin. 2024. *Machine-Actionable Ancient Text Corpus (MAAT)*.