|  |
| --- |
| **A serious call for a robust text restoration evaluation framework** |
|  |
| **Anonymous ACL submission** |
|  |
|  |
|  |
|  |

Abstract

A delightful abstract that is by no means generated by a robot.

Introduction

One of the fundamental tasks in ancient textual scholarship is the creation of an accurate transcription of texts under consideration. Often the physical media supporting the text has undergone decay, leaving gaps, or lacunae, in the text. Scholars are often able to restore the text in these lacunae. Given advances in computational approaches to text analysis, especially deep learning and large language models, there is hope that these methods can aid papyrologists and epigraphers to do text restoration (Assael et al., 2019; Brunello, Andrea et al., 2023).

The text restoration task

For papyrologists and epigraphers, the text restoration task is iterative: lacunae in the text are progressively restored until they are satisfied they can restore no more. They may also review previous editions of a particular text and offer new hypotheses, or readings, of missing text. Figure 1 shows a typical edition of a papyrological text from the Papyri.info of P.Flor 3 324, a contract for property sold (*Aegyptus.89.240*, n.d.). The text is marked up using the Leiden conventions (*Papyrological Conventions*, n.d.; Wilcken, Ulrich, 1932). For this paper, three things should be noted. First, text restorations are provided in brackets (for example, [ἢ ὅ]σων̣ indicates that “ἢ ὅ” has been supplied by the papyrologist). Second, some text has not been restored; this is indicated by dots (marking the approximate number of letters known to be missing) or by “-ca.?” for gaps of unknown extent. Third, some texts have *apparatus notes*, which generally point to other readings. Two apparatus notes appear in line two of the transcription: one, an alternate reading due to a different edition, the other, an alternate restoration of the text.

Acknowledgments

An example acknowledgment.

References

*Aegyptus.89.240 = HGV P.Flor. 3 324 = Trismegistos 25457 = p.flor.3.324*. (n.d.). Duke Databank of Documentary Papyri. Retrieved May 1, 2024, from http://papyri.info/ddbdp/aegyptus;89;240

Assael, Y., Sommerschield, T., & Prag, J. (2019). Restoring ancient text using deep learning: A case study on Greek epigraphy. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 6367–6374. https://doi.org/10.18653/v1/D19-1668

Brunello, Andrea, Colombi, Emanuela, Locaputo, Alessandro, Magnani, Stefano, Saccomanno, Nicola, & Serra, Giuseppe. (2023). *Usage of Language Model for the Filling of Lacunae in Ancient Latin Inscriptions: A Case Study*. *3286*. https://ceur-ws.org/Vol-3286/

*Papyrological conventions used in Duke Databank texts*. (n.d.). Retrieved May 1, 2024, from https://papyri.info/conventions.html

Wilcken, Ulrich. (1932). *Das Leydener Klammersystem*. B.G. Teubner Verlagsgesellschaft.

1. Appendices

Appendices are added after the References section by restarting the header numbering using style “A, B, C”.

1. Supplementary Material

Supplementary material also be included with the Appendices.

A screenshot of a computer

Description automatically generated

Figure 1 Leiden Transcription of P.Flor. 3 324