Federica Nicolardi – Stephen Parsons et alii

Revealing Text from a Still-rolled Herculaneum Papyrus Scroll (*PHerc.Paris.* 4)

aus: Zeitschrift für Papyrologie und Epigraphik 229 (2024) 1–13

Revealing Text from a Still-rolled Herculaneum Papyrus Scroll (*PHerc.Paris.* 4)*

In memoriam *Prof. Marcello Gigante (1923–2001), founder of the Centro Internazionale per lo Studio dei Papiri Ercolanesi, eternal teacher and visionary in the study of Herculaneum Papyri*

1. Introduction

Thanks to a global contest and a groundbreaking interdisciplinary team of computer scientists and papyrologists, complete texts are being revealed from the still-rolled Herculaneum papyrus scrolls for the first time in history. The Vesuvius Challenge,[1] which was launched in March of 2023, brought the best minds in the artificial intelligence (AI) community together with leaders in the virtual unwrapping of damaged manuscripts and experts in the study of Herculaneum papyri to accelerate efforts to discover the scrolls' hidden contents. By October, our team of expert papyrologists who were chosen to adjudicate the contest were reviewing portions of six columns of never-before-seen Greek text from within *PHerc.Paris.* 4, a completely intact scroll from the collection of the Académie des Inscriptions et Belles-Lettres at the Institut de France (Fig. 1).[2]

The initial revelation represents the most significant breakthrough in Herculaneum papyrology since their first discovery in 1752 and consequent mechanical opening procedures, warranting this initial publication. This paper describes the technical process that finally – and non-invasively – opened up *PHerc.Paris.* 4 for reading, discusses the history and features of the scroll, and provides the first diplomatic and literary transcription of the Greek text that was initially revealed (see below, Fig. 12).[3]

Fig. 1. *PHerc.Paris.* 4, AIBL, Institut de France, Paris.

* The authors wish to acknowledge the following for their support and assistance: The Andrew W. Mellon Foundation; The U.S. National Endowment for the Humanities; The U.S. National Science Foundation; The U.K. Arts and Humanities Research Council; The Académie des Inscriptions et Belles-Lettres and the Bibliothèque, Institut de France; The Biblioteca Nazionale "Vittorio Emanuele III"; The Bodleian Library at the University of Oxford; The Getty Conservation Institute and the Getty Villa, Micro Photonics, Inc.; Diamond Light Source; The Centro Internazionale per lo Studio dei Papiri Ercolanesi 'M. Gigante'.

[1] *Vesuvius Challenge* (n.d.). Retrieved January 23, 2024, from https://scrollprize.org.

[2] The numbers of *PHerc.Paris.* 3 and 4 have not been consistently cited in some prior publications. But with this work we have definitively confirmed with the Library that the scroll object of this paper (Vesuvius Challenge's scroll 1) is *PHerc.Paris.* 4.

[3] See https://scrollprize.org/firstletters for the image showing the text discussed here. The contest has continued and has so far culminated in the revelation of additional text from *PHerc.Paris.* 4, which we will be reading and interpreting in the coming months. A more complete edition of the text revealed by the Vesuvius Challenge 2023 Grand Prize is expected later this year.

2. Virtual unwrapping

The revelation of text from within *PHerc.Paris.* 4 is made possible through a multi-stage computational process called *virtual unwrapping* (VU). Virtual unwrapping was invented as a way to overcome the profound fragility that makes physical unrolling impossible.[4] In 2015, the VU technique was applied to a carbonized Hebrew scroll from En Gedi, Israel, revealing for the first time ever a complete text from inside an object so damaged that it could never be opened.[5] More recent projects include the virtual unwrapping of layered fragments from the Dead Sea Scrolls[6] and of a charred and water-logged book written in the Egyptian Coptic language and dating from the fifth or sixth century.[7]

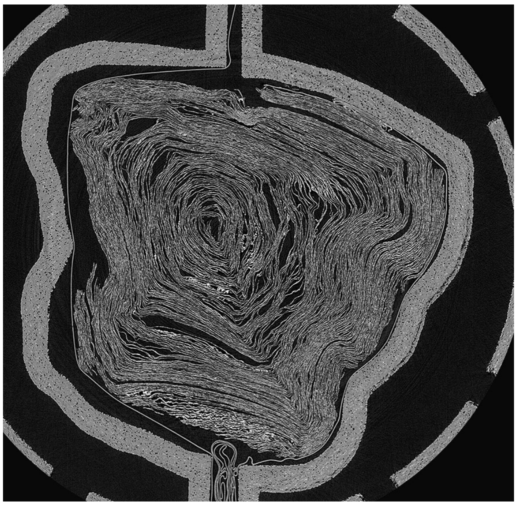

Fig. 2. Cross-section micro-CT image of *PHerc.Paris.* 4 in its form-fitting case.

Virtual unwrapping uses data acquired non-invasively via micro-computed tomography (micro-CT), an X-ray-based technique which shows the interior structure of the scroll. Rather than a flat X-ray image, however, a very high-resolution, three-dimensional volume is generated by the micro-CT scanner. This 3D X-ray recreates the entire interior structure of the scroll by relying on density measurements. At every 3D point within the micro-CT scan, a grayscale value is assigned based on the density of the material at that particular point – darker for less dense material, brighter for more dense material. The scrolls are not uniformly dense because their components, which include individual papyrus fibers, rolled-up papyrus layers, air that sits in the gaps between the layers, ink on the surface, pumice from the eruption, and assorted other elements, all vary in density. The complex 3D volume that is generated by micro-CT holds information about every item present in the scroll's interior, even if evidence of it is extremely difficult or impossible to see by the human eye in the rendered volume.

While the micro-CT scan serves as the digital starting point, virtual unwrapping refers to the software process applied to the scan data. This set of algorithmic steps begins by tracing the surface of the papyrus. Any writing within the volume will be found on the papyrus surface, so isolating that surface is an important first step. This process, called "segmentation", uses the depicted density differences between the fibrous, denser papyrus sheets and the less dense air gaps between them to find and trace the surface of the papyrus, thereby isolating the writing substrate. The structure of the papyrus is unpredictable, and

[4] W. B. Seales – Y. Lin, Digital Restoration Using Volumetric Scanning, in *Proceedings of the 4th ACM/IEEE-CS Joint Conference on Digital Libraries*, New York 2004, 117–124 (https://doi.org/10.1145/996350.996380); W. B. Seales – C. Chapman, Technology and the Quest to Unlock the Secrets of the Herculaneum Scrolls, in *Buried by Vesuvius: The Villa dei Papiri at Herculaneum*, Los Angeles 2019, 124–132.

[5] W. B. Seales – C. S. Parker – M. Segal – E. Tov – P. Shor – Y. Porath, From Damage to Discovery via Virtual Unwrapping: Reading the Scroll from En-Gedi, *Science Advances* 2(9), 2016 (https://doi.org/10.1126/sciadv.1601247).

[6] L. Piercy, Reading the Unreadable: Seales and Team Reveal Dead Sea Scroll Text, *UKNow* 2020, July 23 (https://uknow.uky.edu/research/reading-unreadable-seales-and-team-reveal-dead-sea-scroll-text).

[7] S. Parsons – K. Gessel – C. S. Parker – W. B. Seales, Deep Learning for More Expressive Virtual Unwrapping, in *The 25th International Conference on Cultural Heritage and New Technologies*, Heidelberg 2020, 203–207.

the segmentation of every wrap and layer is difficult to achieve. Therefore, we combine human input with algorithmic refinement to verify visually that the surface estimate of the papyrus is correctly localized in the tomographic volume.

The output of the segmentation process is a set of complex 3D structures called meshes, which are made up of tiny, connected triangles, linked together to follow the geometry of the twisted 3D papyrus surfaces that are depicted in the scans.

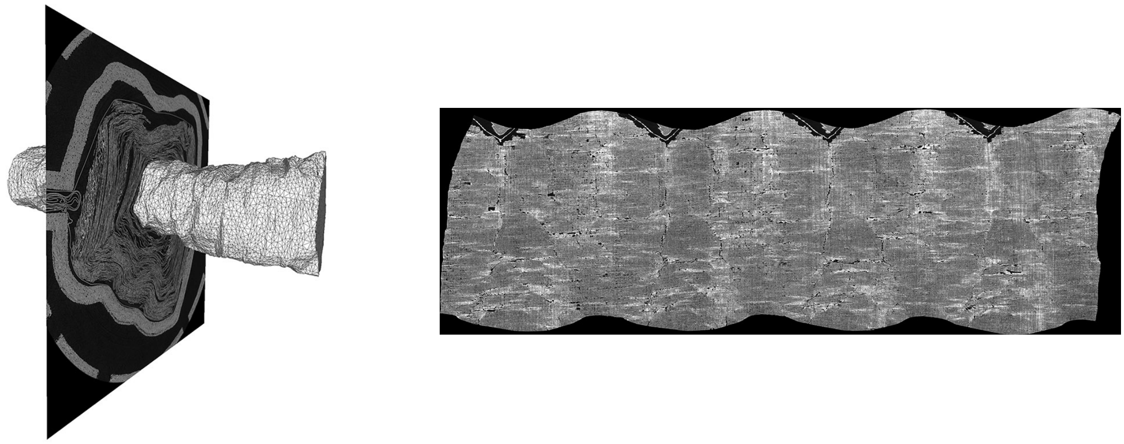

Fig. 3. The selected surface layer is first isolated via segmentation (left) and then textured and flattened (right). Ink is typically not visible on this surface.

Once the papyrus surfaces are localized in the volume and represented as triangulated meshes, it is possible to manipulate them algorithmically. Virtual unwrapping requires three such transformations: flattening, texturing, and resampling. First, the mesh is "unwrapped" or "unrolled" in the flattening step (Fig. 3), which transforms or de-warps the 3D surface shape into a flat 2D image, thereby simplifying the geometry and the visualization of the otherwise twisted 3D surface.[8]

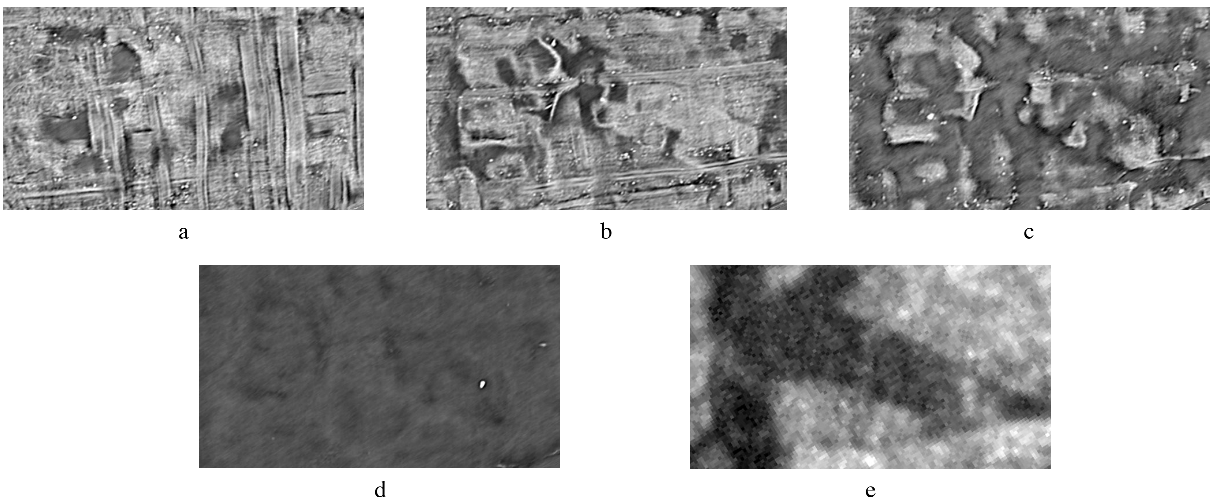

Fig. 4. Micro-CT images from a multi-layered Herculaneum fragment (*PHerc.Paris*. 1, fr. 34) showing the variations in ink signal at different depths in the papyrus. (a) Eleven microns beneath the segmented surface. Vertical papyrus fibers from the verso are prominent. (b) Three microns beneath the surface. Horizontal fibers from the recto are more prominent and ink begins to appear. (c) Two microns above the surface. Some protruding ink remains while papyrus begins to fade to air. (d) Ten microns above the segmented surface. Only air is visible. (e) Infrared photograph of same area.

[8] C. S. Parker – W. B. Seales – P. Shor, *Quantitative Distortion Analysis of Flattening Applied to the Scroll from En-Gedi*, 2020 (arXiv:2007.15551) (https://doi.org/10.48550/arXiv.2007.15551).

Once flattened, “texturing” of the mesh follows, whereby the grayscale density values from the micro-CT 3D volume are transferred to their corresponding locations on the flattened mesh. For a Herculaneum scroll, this texturing step produces a flat image showing the grid-like fiber pattern of a papyrus sheet. Individual fibers can be seen clearly, as well as surface defects and other interesting features. However, the ink remains, at best, difficult to see.

Finally, the segmented mesh is used as a guide to flatten and resample the 3D volume into a simplified form called a surface volume. That is, the algorithms inspect the papyrus surface with some depth, looking not only at the thin boundary following the surface, but also above and below the papyrus surface in order to capture those places containing the subtle effects of ink presence. This re-sampling step is important since ink penetrates the papyrus surface and because the spatial resolution in the volume is high enough to capture a number of samples through the cross section of every papyrus layer (Fig. 4). Without this fine-toothed sampling, evidence of ink can be missed.

Rather than representing the papyrus layer as a single image, this process regularizes the surface, presenting it as a set of parallel, flattened images that capture all ink evidence through the entire thickness of the papyrus from top to bottom. This regularization step preserves evidence of the ink while conditioning the images to be easier and more efficient to process in subsequent steps.

It is at this point that the Herculaneum papyri prove more challenging than prior virtually unwrapped texts. For some manuscripts, such as those from the medieval period that use metal-containing inks like iron gall, the density of the ink differs enough from that of the writing surface that the ink appears clearly after the flattening and texturing steps. In the case of Herculaneum papyri, the ink and papyrus are both made primarily of carbon, resulting in extremely similar density values that cause the materials to look the same in the micro-CT data.[9] Thus, the “data signal” of ink found in Herculaneum scrolls is much more subtle, requiring more sophisticated detection methods than those based on density differences alone.

To address this problem of carbon ink “invisibility”, machine learning-based ink detection algorithms can be applied to the textured meshes to reveal the location of ink (and therefore the text) on the writing surface.[10] By taking photographs of papyrus fragments that show visible text on the surface and aligning them with micro-CT scans of those same papyrus samples, a neural network can be trained to recognize the data configuration that signals the presence of ink (Fig. 5).[11] This approach was used to develop the proof of concept that Herculaneum ink can be recovered from micro-CT. Since the photographic ink labels serve as “ground truth”, the outputs of the machine learning process can be validated quantitatively for correctness, establishing machine learning as a viable ink detection method.

9 On the discovery of metal in inks from Herculaneum, however, see W. B. Seales. Lire sans détruire les papyrus carbonisés d’Herculanum, *Comptes rendus des séances de l’Académie des Inscriptions et Belles-Lettres*, 153, 2, 2009, 907–923; E. Brun – M. Cotte – J. Wright – M. Ruat – P. Tack – L. Vincze – C. Ferrero – D. Delattre – V. Mocella, Revealing Metallic Ink in Herculaneum Papyri, *Proceedings of the National Academy of Sciences* 14.113, 2016, pp. 3751–3754; P. Tack – M. Cotte – S. Bauters – E. Brun – D. Banerjee – W. Bras – C. Ferrero – D. Delattre – V. Mocella – L. Vincze, Tracking Ink Composition on Herculaneum Papyrus Scrolls: Quantification and Speciation of Lead by X-ray Based Techniques and Monte Carlo Simulations, *Scientific Reports* 6, 2016, 20763; O. Bonnerot – G. Del Mastro – J. Hammerstaedt – V. Mocella – I. Rabin, XRF Ink Analysis of Some Herculaneum Papyri, *ZPE* 216, 2020, pp. 50–52.

10 C. Parker – S. Parsons – J. Bandy – C. Chapman – F. Coppens – W. B. Seales, From Invisibility to Readability: Recovering the Ink of Herculaneum, *PLOS ONE* 14.5, May 2019, pp. 1–17 (https://doi.org/10.1371/journal.pone.0215775).

11 S. Parsons, *Hard-Hearted Scrolls: A Noninvasive Method for Reading the Herculaneum Papyri. Theses and Dissertations – Computer Science* 138, 2023 (https://uknowledge.uky.edu/cs_etds/138).

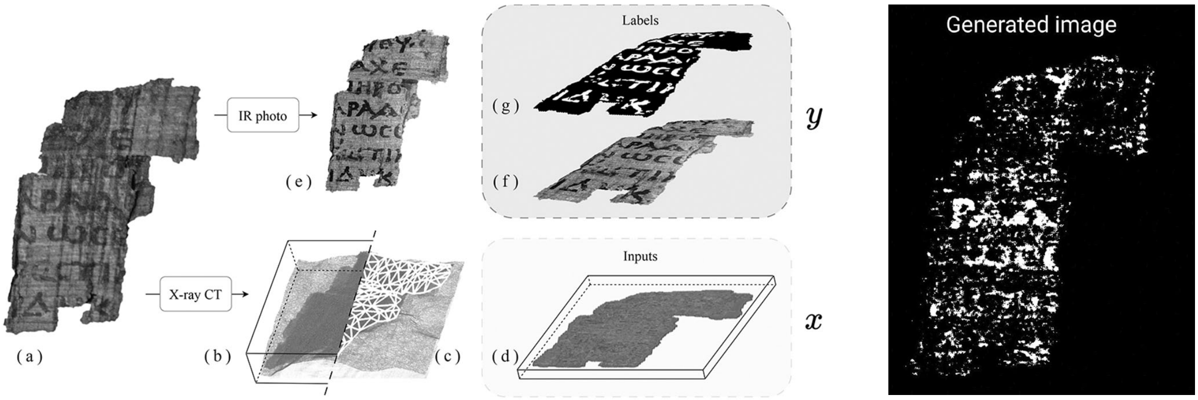

Fig. 5. Training the inkID neural network using *PHerc.Paris*. 2, fr. 47. (a) Photo of fragment with visible text. (b) Volumetric X-ray CT image. (c) Surface segmentation. (d) “Flattened” surface volume. (e) Infrared photograph of fragment. (f) Infrared photograph aligned to surface volume. (g) Ink labels aligned with surface volume and used for training. On the right: Generated image from inkID network.

In other cases, one can train a network using the faintly visible hints of ink that can occasionally be perceived upon scrutinizing the tomography within intact scrolls (Fig. 6). This approach was used to produce the results presented here. Ink in tomography may vaguely appear as peeling and flaking paint or dry, cracked mud.[12] Such examples represent the more exaggerated ink features that are sometimes seen by eye; in most cases, however, the ink signal is less clear. The features are clearly detectable from a combination of evidence in both shape and density, although the relative contributions of that evidence is difficult to determine.

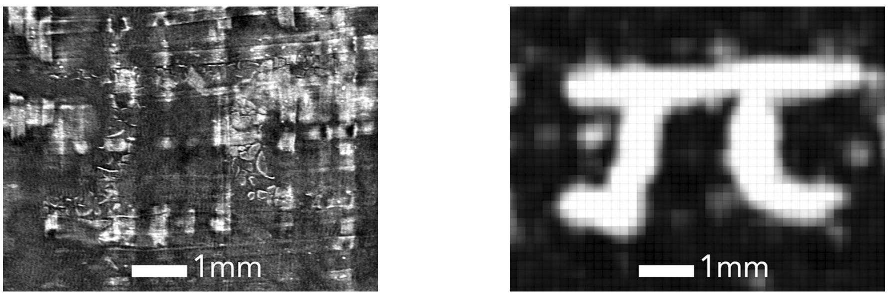

Fig. 6. *PHerc.Paris*. 4. (a) X-ray CT showing crackle pattern. (b) Machine-learning ink predictions of same area.

Despite some imprecision in the contributions of ink evidence, what is clear is that the networks learn to detect some type of subtle difference between papyrus and carbon ink, and then exaggerate it for the purpose of better visibility. A key point to note about both approaches, however, is that the algorithms are not trained to detect letterforms: in other words, they are not searching for something that looks like an α, β or γ. In fact, they examine data regions much too small to contain a complete character. Instead, they learn to inspect tiny 3D regions of micro-CT data for signals indicating that ink is present and then to issue a verdict, or prediction, on whether ink is present at that particular point. Aggregating these predictions on the textured mesh enhances the image, making the writing visible. The identification of letterforms in the output images is, therefore, the result of human interpretation and not of algorithmic hallucination. Further details regarding the software process will be made available in a forthcoming technical publication.

12 See https://scrollprize.org/firstletters.

3. A brief history of *PHerc.Paris.* 4

Of the hundreds of carbonized papyri found at Herculaneum between 1752 and 1754, six are today preserved at the Institut de France. They were in fact offered to First Consul Napoleon Bonaparte in 1802 by Ferdinand IV, King of Naples, along with numerous other archaeological objects, in return for the promise that they would be unrolled and published. Two publications by Aubin Louis Millin de Grandmaison bear witness to this diplomatic gift of more than 220 years ago.[13]

These six rolls belong to the only library to have come down to us from antiquity. They were recorded as 'whole papyri' in the ancient inventories of the Officina dei Papiri at Portici under the numbers 148, 171, 184, 185, 205 and 1009.[14] They arrived in Paris in a softwood crate lined with coarse cotton wool and were catalogued as 'Objet 59' in the library of the Institut. Two rolls were opened in partnership with Neapolitan conservators between 1986 and 1987 (*PHerc.Paris.* 1 and 2).[15] Two rolls survive intact (*PHerc.Paris.* 3 and 4). The remaining two rolls are fragmentary, making their identification more problematic.[16]

In July 2009 the two intact rolls were imaged non-invasively for the first time using a SkyScan computed tomography system that was moved to the Institut de France specifically for this purpose.[17] The first-ever 3D tomographic view of the interior of both scrolls showed clearly why it would be physically impossible to unroll the papyrus. The interior structure of both scrolls proved to be very irregular, with papyrus wrinkles, tears, delamination, and compressions in all directions. *PHerc.Paris.* 3 was discovered to contain a complete *umbilicus* in its centre, although broken in several places, while *PHerc.Paris.* 4 was simply rolled upon itself without an *umbilicus*. Clear evidence of ink was not apparent in the tomographic data that was acquired at that time.

As available imaging methods matured, researchers made further attempts to achieve clear ink contrast, both through higher spatial resolutions and contrast-enhancement methods such as phase contrast. In 2013 *PHerc.Paris.* 4 (the subject of this present study) was imaged again with computed tomography, this time using the ESRF synchrotron in Grenoble.[18] The hope was that propagation-based phase-contrast imaging could produce clearer evidence of ink in the tomography.

The most recent imaging, which is the basis of the virtually unwrapped text in this paper, was conducted in 2019 at the Diamond Light Source in Oxford. The monochromatic beam produced results containing implicit propagation-based phase shift data, but did not prioritize the amplification of this shift, thus

---

[13] The more substantial of both appeared already in 1803, in *Le Magasin Encyclopédique ou Journal des Sciences, des Lettres et des Arts*, Paris 1803, pp. 96–98. The second, containing a simple reference to the gift of the scrolls as well as to the arrival of J. Hayter in Portici, figures in the *Dictionnaire des Beaux-Arts* (Paris 1806, vol. II, *s.v.* "Herculaneum", p. 39), also by Millin de Grandmaison.

[14] See the reference in the hand-written Catalogue of the Herculaneum Papyri of 1822, preserved in the Officina dei Papiri in the Biblioteca Nazionale Vittorio Emanuele of Naples, which specifies on p. 49, opposite no. 1099: "papiro intero : per ord(in)e superiore mandato al I° Console di Francia Napoleone Buonaparte" ("whole papyrus: sent by order of higher authority to the First Consul of France, Napoleon Bonaparte"). In the oldest Catalogue (before 1786, dubbed "Piaggio's Catalogue"), the dimensions given for this roll are as follows: length, 9.2 once with an upper diameter of 2 and a half once, i.e. a length of 20.286 cm with an upper diameter of 5.512 cm (1 oncia = 2.205 cm); see D. Blank – F. Longo Auricchio, Inventari antichi dei Papiri Ercolanesi, *CErc* 34, 2004, p. 81. This catalogue, unfortunately incomplete, begins only with no. 312.

[15] D. Delattre, Le retour du PHerc. Paris 2 à l'Institut de France: un rouleau épicurien inédit en 279 fragments, *Académie des Inscriptions et Belles-Lettres. Comptes Rendus* 2004, pp. 1351–1391, with illustrations and bibliography.

[16] See D. Delattre, Cronistoria dei papiri ercolanesi conservati a Parigi (1802–2012), *CErc* 44, 2014, pp. 129–144, esp. 134–141.

[17] The first imaging was the result of collaboration between Daniel Delattre, commissioned by the Institut to work on these carbonized scrolls, and Brent Seales. The use made of the data collected by the American team was the subject of a detailed report: W. B. Seales – D. Delattre, Virtual Unrolling of Carbonized Herculaneum Scrolls: Research Status (2007–2012), *CErc* 43, 2013, pp. 191–208. See also W. B. Seales, Lire sans détruire les papyrus carbonisés d'Herculanum, *Comptes rendus des séances de l'Académie des Inscriptions et Belles-Lettres* 153, 2009, pp. 907–923.

[18] V. Mocella – E. Brun – C. Ferrero – D. Delattre, Revealing Letters in Rolled Herculaneum Papyri by X-ray Phase-contrast Imaging, *Nature Communications* 20 January 2015 (https://doi.org/10.1038/ncomms6895). Other experiments on Herculaneum papyri were also conducted by I. Bukreeva et al., Virtual Unrolling and Deciphering of Herculaneum Papyri by X-ray Phase-contrast Tomography, *Scientific Reports* 6, 2016, 27227 (https://doi.org/10.1038/srep27227). Despite early claims of textual discovery, these experiments, which used a different approach than the technique presented here, have not yet led to further textual discoveries.

resembling standard X-ray micro-CT. These images and their processing, released in EduceLab-Scrolls,[19] instead focus on the salient cues that had become known by then to be crucial to ink detection: the highest achievable resolution, precise segmentation, and accurate labeling.[20] These factors combine to create a dataset in which machine learning-based methods can detect the presence of ink, even without strong visual contrast.

After a 270-year history of preservation and then non-invasive tomographic imaging (with at least three different systems), the combination of extremely high precision spatial resolution, a monochromatic incident beam, and implicit propagation-based phase contrast is the basis for the current work. The acquired data combined with an artificial-intelligence approach for carefully enhancing the evidence of the ink has succeeded in producing textual evidence that is substantial enough to support the following scholarly papyrological examination.

4. Palaeography, layout of the text, position in the scroll

The script of *PHerc.Paris.* 4 is a formal, medium-sized capital. Excluding those breaking bilinearity (*rho*, *phi*), the letters are 3–4 mm high. The interlinear spacing is regular, measuring approximately 3 mm. *Omicron* is usually slightly smaller than the other letters and often floats above the base line. The strokes tend to be curved: *pi* bends its second upright to the right, while *kappa*, *my* and *ny* bend their first upright to the left (sometimes this also happens to the vertical stroke of *rho*). *Epsilon* sometimes shows the middle stroke detached from the body of the letter. Other peculiarities involve: *alpha* with first and second strokes linked in a closed loop; *my* which sometimes merges the second upright with the second diagonal (coll. -15, l. 2; -14, l. 5; -13, l. 3); and small serifs at the base of *kappa*, at the foot of *pi* and, above all, of *tau*, in which the left part of the upper stroke curves with an elegant curl. It is possible to observe a διπλῆ (col. -15, l. 2, with the second stroke longer than the first and curving downwards) and a παράγραφος two lines below.

The column-to-column width measures approximately 79 mm, with an average column width of 62 mm and an average intercolumnium of 17 mm. The number of letters per line varies between 14 and 17, with an average of 16 letters. This is a slightly lower number than is normally observed in the Herculaneum papyri. The final letters of the line are often compressed to avoid difficult word-divisions. Maas's Law is slightly observed.

There are, among the Herculaneum papyri, hands that present a graphic language similar to our papyrus: for example *PHerc.* 1507 (*De bono rege secundum Homerum*) and *PHerc.* 1050 (Philodemus, *De morte*) show letters with a very similar shape (*ny*, *rho*, *tau*, *omega*); however, the general appearance of the script allows us to exclude that this is by the same scribe. Among the non-Herculaneum papyri some useful comparisons can be made with *PRyl.* 3, 540 (TM 60260, Hom. *Il.* I), *PAmst.* inv. I (TM 61163, Hom. *Il.* 4) and *PRyl.* 1,22v (TM 63125, *Homerica*; less elegant but with some similar features). These comparisons allow us to date *PHerc.Paris.* 4 between the second half of the 1st century BCE and, more probably, the first half of the 1st century CE.

From the segmentation of the scroll we know that the first identifiable column, of which only a few letters from the ends of two lines are visible, is the sixteenth, counting backwards from the end of the scroll. Consequently, we have chosen to assign negative numbers to the columns, ranging from -16 to -11. Even though the segmentation of the whole scroll is not yet complete, and the exact number of columns from the beginning to col. -16 is still unknown, counting the wraps of the entire remains of the scroll (140–150) and measuring them in the scan allowed us to estimate the total length of the scroll as approximately 13.2 meters. This means the *volumen* preserves approximately 165 columns.

[19] S. Parsons – C. S. Parker – C. Chapman – M. Hayashida – W. B. Seales, EduceLab-Scrolls: Verifiable Recovery of Text from Herculaneum Papyri Using X-ray CT, *arXiv* preprint *arXiv*:2304.02084, 2023 (https://doi.org/10.48550/arXiv.2304.02084); S. Parsons, Hard-Hearted Scrolls, *cit.* n. 11.

[20] C. Parker – S. Parsons *et. al.*, From Invisibility to Readability, *cit.* n. 10.

5. Publishing a virtually unwrapped Herculaneum scroll: Material and digital damages

Publishing a partial text from a virtually unwrapped Herculaneum papyrus for the first time required us to address new challenges and to question traditional conventions in order to be able to confirm their validity in completely new circumstances. Above all, one must acknowledge that no original is available for autopsy. The object in our possession will never physically appear in the state in which we perceive it via the AI-generated image that offers the opportunity partially to read its text, and, as was noted earlier, in images that will soon allow us to read it in its entirety.

Upon close examination of the images, a wavy outline at the top and bottom of the virtually unwrapped papyrus strip is immediately evident. Since this distortion could not be attributed to recurring material damage, as these curves represent the edges of the segmented area rather than those of the fragments, we sought to determine whether these waves could serve as a guide for identifying the circumferences in the scroll. In addition to the waves, two types of vertical fractures can be observed.

Comparing the distance between wave crests (solid line arrows in Fig. 7) with that between corresponding fractures (dotted line arrows for fractures 'A', and dashed line arrows for fractures 'B'), we note that the results are compatible:

Distances between crests:	61 mm	59 mm	57 mm	55 mm	53 mm	51 mm
Distances between fractures 'A':	61 mm	59 mm	57 mm	55 mm	54 mm	×
Distances between fractures 'B':	×	59 mm	58 mm	57 mm	56 mm	×

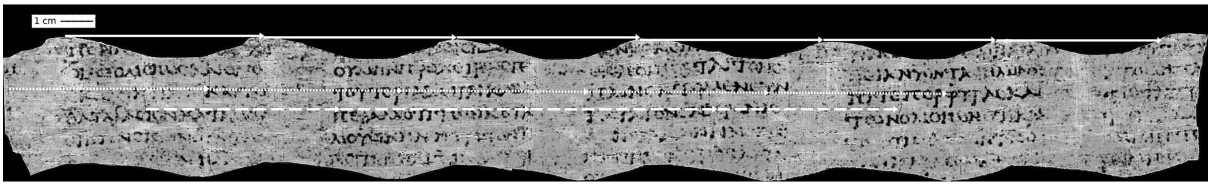

Fig. 7. Distances between wave crests (solid line arrows),
fractures 'A' (dotted line arrows) and fractures 'B' (dashed line arrows).

This confirms that they are all "guide damages", meaning that the distance between two corresponding distortions is equal to the width of a circumference of the scroll. However, in the first case, the damages are specifically 'digital', stemming from variations in height introduced by the segmentation procedure, wherein the scroll was cut at different heights on its two sides. This is due to the scroll not being exactly vertical in the scanner, but slightly tilted.

Like the waves, the gaps that appear are also exclusively present in the digital form: that which appears to be missing before the first visible letters in the image, for instance, or after col. -11, is not 'missing' in reality. On the contrary, it is safely preserved – though not visible – within the unopened scroll. While understanding that these identified gaps are only digital and temporary, as progress in the virtual unwrapping will reveal new portions of text, we decided to use the usual Leiden square brackets for them. Similarly, we chose to use sublinear dots, as usual in papyrological editions, even if the degree of 'legibility' is highly dependent on the degree of success that the AI has achieved thus far in identifying and revealing consistent ink traces and is therefore susceptible to considerable improvement. What can happen with further progress is comparable to the situation where a papyrus, initially captured only through visible light, is later reimaged using infrared or other techniques that reveal additional ink.

A more challenging situation required special consideration, viz. when AI suggests traces that are impossible or contrary to what one would – with a high degree of certainty – expect.

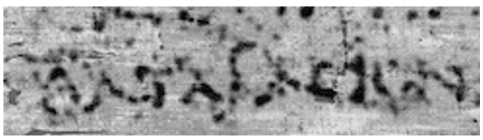

Fig. 8. Col. -15, l.4.

The most evident case of such an issue is at col. -15, l. 4 (Fig. 8), where the word θαυμάσιο̣ν can clearly be read. At a first glance, the reading is disturbed by a sinusoidal stroke through the right part of the *my*, which is, however, still visible in its entirety. After further consideration, including exchanges with the Vesuvius Challenge and virtual unwrapping teams, it appeared most reasonable to conclude that the sinusoidal stroke is a hallucination caused by noise in the image, which the AI misinterpreted as ink. How to treat such cases? Printing the *my* with a sublinear dot would suggest that it is uncertain and that the visible traces could point to a different reading, which is not the case here. The temptation to establish a convention to deal with such special circumstances that arise when working with an AI-generated image was strong, not only to account for the divergence of the image, but also to avoid the risk of inadvertently encouraging the network to recognize noise as ink by providing the technical team with a transcription showing a simple *my*. After careful consideration, however, we decided to avoid superfluous augmentation to an already complex ecdotic system, such as that of papyrological editions, particularly of Herculaneum papyri.[21] Alerting the technical team is sufficient to prevent misuse of this *my* in model training, and this brief comment is intended to clarify why the sinusoidal stroke has no place in our transcription.

Some other ambiguous cases occur in the images. We are confident that improvements of the images will address them. Below we list the most notable examples:

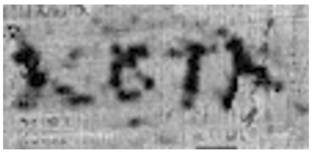

Fig. 9. Col. -14, l.4.

– At col. -14, l. 4 (Fig. 9) at the place of the second letter of the sequence κα̣τα, a rounded stroke, open to the right, and an oblique stroke, descending from left to right and going through the rounded shape, are visible. At present, it is not possible to determine with absolute certainty whether the scribe corrected the text tracing an *alpha* onto a rounded letter, or, as appears more plausible to us, the rounded stroke is a result of noise misinterpreted by the AI.

Fig. 10. Col. -14, l. 5.

– The small grey rectangle at the top left of the *ypsilon* in col. -14, l. 5 (Fig. 10) seems to be a segmentation imperfection, but we cannot exclude for now that it is a small hole in the papyrus.

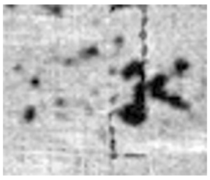

Fig. 11. Col. -14, l. 5.

– The oblique tick that appears before a *kappa* at col. -15, l. 3 (Fig. 11) could be interpreted as a punctuation mark, but could also simply be an AI hallucination.

Taking into consideration the temporary nature of the gaps, the fact that the limited readability or complete illegibility of the text in specific regions of the image is due to AI's current inability to identify the ink consistently, and the expectation of new results and images from the awarding of the Vesuvius Challenge Grand Prize,[22] we have chosen to publish an extremely cautious text, which includes only readings on which we all agree, and a very limited apparatus. In addition, we decided to postpone any comments on and discussions of the text to a future, more comprehensive article as was mentioned in the introduction.

[21] Scholars can now refer to the *Guidelines for editing papyri*, published online on the website of the Association Internationale de Papyrologues (https://aip.ulb.be//PDF/Guidelines_for_editing_papyri.pdf), which also propose new rules for rendering certain data not taken into account by the Leiden system.

[22] As the Grand Prize results were announced when we submitted our paper to this journal and we did not have the chance to make full use of the new images, we did not use them to increase the quantity of transcribed text. In the few cases where the new images suggested reconsideration of our reading of the papyrus, we noted the new reading in a footnote.

6. The text

col. -16

– – –
[..............]..(.)
[................]
[..............].
– – –

– – –
[..............]..
[................]
[..............].
– – –

col. -15

– – –
περι.[......]....
> ....ομοιωϲκαιο..
..........κα..ο
¯θαυμαϲι.νκατα...
....(.)ν...(.)ν......
.[.....]........
– – –

– – –
περι.[......]....
> ....ομοίωϲ καὶ ον̣..
......τ̣ι̣.... καὶ̣ τ̣ὸ
¯θαυμάϲιο̣ν κατα...
....(.)ν̣...(.)ν......
.[.....]........
– – –

2 ο̣ἱ̣ δ̣ὲ̣ ὁμοίωϲ Del Mastro, Fowler, Janko, Nicolardi, Reinhardt : π̣α̣ν̣ομοίωϲ Delattre   3 ʹ̣ *ante* και̣ *ut videtur*

… in like manner and … and the marvelous …

col. -14

– – –
.[....]..ειϲ...
ουμηνπροχειρωϲγε
τουτ.ρ.α..ταικα..
περαλλοτιτωνκ.τα
μουϲικ.ν.τριτ.νυ
π.τ...ν.ι..[.]....
– – –

– – –
.[....]..ειϲ...
οὐ μὴν προχείρωϲ γε
τοῦτ’ ἐ̣ργ̣ά̣ζ̣ε̣ται, καθ̣ά̣-
περ ἄλλο τι τῶν κα̣τὰ
μουϲικ̣ήν, η̣ τρίτο̣ν ὑ-
πο̣τ...να̣ι..[.]....
– – –

5 ἢ̣ *vel* ἣ̣ *vel* ἦ̣   5–6 ὑ|πο̣τα̣γ̣ῆ̣να̣ι *exempli gratia*[23]

… Yet this is not achieved readily, like some other aspect of the matters that pertain to music, or/which/to which … to have been ranked third …

col. -13

– – –
..νκε.....[......]
...(.)ποιηϲεταιτον.
πα.ν.ντομεν.υν
κ.τα.ον.νχρ.νι.
...........(.)ν...
...τ............
..[.........]....
– – –

– – –
..νκε.....[......]
...(.)ποιήϲεται τὸν ἔ̣-
παι̣νο̣ν. τὸ μὲν ο̣ὖν
κα̣τὰ τ̣ὸν ἐ̣νχρο̣νιϲ̣-
........ν̣..(.)ν...
...τ............
ο̣υ̣[.........]....
– – –

4–5 ἐ̣νχρο̣νιϲ̣|μ̣ό̣ν̣ *vel* ἐ̣νχρο̣νιϲ̣|θ̣έ̣ν̣τ̣α̣

[23] The new images now show clearly ὑ|ποτιθέναι.

… will make the praise … Therefore, the … according to the (prolonged use?/one whose time was wasted?) …

col. -12

– – –	– – –
....[.....]......α	....[.....]... π̣α̣ρ̣α-
.εικνυοντα.ημιν.	δ̣εικνύονται̣ ἡμῖν ἐ̣-
πιτεπορφυραςκαι	πί τε πορφύρας καὶ
τωνομοιωνοτικ..	τῶν ὁμοίων ὅτι κ..
τ.(.).υ.......(.)τ....	τω̣. υπ̣......(.)τ..ο̣.
....	....
[.....]...[......]	[.....]...[......]
– – –	– – –

1–2 π̣α̣ρ̣α|δ̣εικνύονται̣ *vel* π̣α̣ρ̣α|δ̣εικνύοντας̣[24]

*… they are compared (*or *who compare) to us in the case of purple (dye/clothing/…?) and similar things because …*

col. -11

– – –	– – –
[...]..τ̣...[.....]	[...]..τ̣...[.....]
.το.ηπα[.......]	.τομ̣η πα[.......]
τ..τ...[.......]	τ..τα̣..[.......]
τ......[.......]	τ......[.......]
.....τ.[.......]	....ντο̣[.......]
ρ......[.........]	ρ......[.........]
– – –	– – –

7. Conclusion

The "black box" nature of AI-enabled virtual unwrapping may give some scholars pause for thought regarding the trustworthiness of the revealed text. However, all three teams – the papyrologists, the Vesuvius Challenge technologists, and the virtual unwrapping inventors – have developed protocols for reviewing the results and confirming the authenticity of the text. As noted above, the papyrologists were careful to identify problem areas in the revealed text and work with the technical teams to understand the reason behind those distortions.

In addition, the unique approach of the Vesuvius Challenge itself enables an extremely rigorous validation process. To confirm their accuracy and veracity, the results undergo a thorough technical review that entails several actions. First, the individual steps of the pipeline are subject to an ongoing communal review. Throughout the contest, thousands of community members were validating the algorithmic steps both by auditing the *methods* (e.g. examining the program code) and by independently reproducing intermediate *results*. This process also gradually raised the validation standard by contributing a stream of incremental improvements. Second, multiple independent teams discovered textually consistent results, even using slightly different approaches, lending credibility to the revealed text. Third, the Vesuvius Challenge administrative team independently reproduced the findings by running the final pipeline submitted by the contests on the original data once again. Finally, in keeping with the principle of open science, the data and code were released to the larger contest community, enabling others to validate the findings.

[24] The new images now show clearly παρα|δεικνύοντας.

Another benefit of producing an open algorithmic pipeline is that it reveals a born-digital object's "digital provenance", meaning any component of the final result can be inspected in its intermediate stages or in the original data. For example, when inspecting a particular character, scholars can look at the corresponding location in the original CT scan to answer questions about the papyrus surface. Such provision of a digital provenance chain that reveals the process used in the revelation of text has been a key component of virtual unwrapping since its inception, and the virtual unwrapping team continues to make the issue a priority by pushing the envelope regarding the transparency of algorithmic applications.

Federica Nicolardi, Università degli Studi di Napoli Federico II/CISPE 'M. Gigante'
federica.nicolardi@unina.it

Stephen Parsons, Vesuvius Challenge/University of Kentucky
stephen@scrollprize.org

Daniel Delattre, CNRS-IRHT Paris/CISPE 'M. Gigante'
phercparis@gmail.com

Gianluca Del Mastro, Università della Campania "L. Vanvitelli"/CISPE 'M. Gigante'
gianluca.delmastro@unicampania.it

Robert L. Fowler,The Herculaneum Society
Robert.Fowler@bristol.ac.uk

Richard Janko, University of Michigan/CISPE 'M. Gigante'
rjanko@umich.edu

Tobias Reinhardt, University of Oxford
tobias.reinhardt@classics.ox.ac.uk

C. Seth Parker, University of Kentucky
c.seth.parker@uky.edu

Christy Chapman, University of Kentucky
cychapman@uky.edu

W. Brent Seales, University of Kentucky
seales@uky.edu

contributing authors:
James H. Brusuelas (University of Kentucky), JP Posma (Vesuvius Challenge), Nat Friedman (Vesuvius Challenge), Daniel Gross (Vesuvius Challenge), Luke Farritor (Vesuvius Challenge Contestant), Youssef Nader (Vesuvius Challenge Contestant), Ben Kyles (Vesuvius Challenge), David Josey (Vesuvius Challenge), Konrad Rosenberg (Vesuvius Challenge), Julian Schilliger (Vesuvius Challenge Contestant), Roger T. Macfarlane (Brigham Young University)

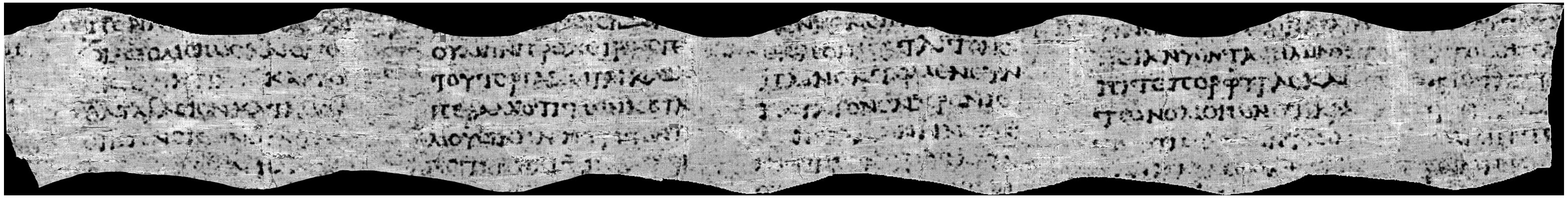

Fig. 12. Text revealed from *PHerc.Paris*. 4, shown on top of the papyrus sheet it is written on.