Siam Bhayro
Peter E. Pormann
William I. Sellers

Imaging the Syriac Galen Palimpsest: preliminary analysis and future prospects

Introduction

In the last issue of this journal, two of the present authors, Siam Bhayro and Peter E. Pormann, together with Robert Hawley and Grigory Kessel, drew the attention of the scholarly community to the presence of a remarkable palimpsest containing the Syriac version of Galen's *On simple drugs* by Sergius of Rēš ʿAynā.[1] The same four scholars also published a more in-depth yet still preliminary study in the *Journal of Semitic studies* at the beginning of this year.[2] Both these articles resulted from collaborative work that was undertaken in the wake of a workshop held at the University of Manchester in May 2012. To summarise the results in the briefest possible terms: the Syriac Galen Palimpsest (henceforth SGP) probably contains at least books six to nine, but probably six to eleven of Sergius' translation of Galen's *On simple drugs*. It preserves a number of interesting readings which at times prove to be superior to those contained in the other manuscript, London, British Library, MS Add. 14661 (henceforth BL), which only contains books six to eight. Since SGP comprises roughly twice as much text as BL, and since Sergius' translation of Galen's *On simple drugs* is one of the very few versions that survive today from the sixth-century Greek-Syriac translation movement, SGP is of crucial importance for the study of this translation movement. Moreover, it also sheds new light on the ninth-century Graeco-Syriac and Graeco-Arabic translation movements. Our preliminary research suggests, for instance, that Ḥunayn ibn Isḥāq (d. c. 873) was much more indebted to Sergius than his *Epistle* (*Risāla*) would suggest.[3]

This research, however, was only made possible because the anonymous owner of the palimpsest made high-quality, multi-spectral images available free of charge under a creative commons licence.[4] Although these spectral images—and especially the composite so-called pseudo-colours and "sharpies"—allow scholars to work with the undertext and read it much better than would be possible with the naked eye, in many instances the undertext remains only partially readable, notably in the areas where it is covered directly by overtext. The question of how one can potentially improve the reading of the undertext through computational means is addressed in the present note. This work draws again on the results of a Manchester workshop, this time organised by the John Rylands Research Institute and the Centre for Heritage Imaging and Collection Care, both of the University of Manchester, in October 2013. We present some preliminary results here and also suggest possible avenues for future research in this area.

Computational analysis of multi-spectral images

In order to explore how sophisticated computational algorithms can enhance the image quality of the undertext of SGP, we chose to focus on fol. 102^{v}–107^{r}, because it contains a fairly clear gutter area and featured in last year's article. The aim was to see to what extent additional information could be extracted from the set of multispectral images available at the web site.[5] We wanted to test whether the overtext is completely opaque to all the wavelengths used in the multispectral image, or whether post-processing would be able to extract information from beneath the overtext. Bill Sellers performed a statistical analysis of the different

1. S. Bhayro, R. Hawley, G. Kessel, P. E. Pormann, "Collaborative research on the digital Syriac Galen Palimpsest", *Semitica et Classica* 5, 2012, pp. 261-264.
2. S. Bhayro, R. Hawley, G. Kessel, P. E. Pormann, "The Syriac Galen Palimpsest : progress, prospects and problems", *Journal of Semitic studies* 58, 2013, pp. 131-148.
3. For more details, see Bhayro *et al*. (*supra*, n. 2), pp. 139-143; see also P. E. Pormann, "The development of translation techniques from Greek into Syriac and Arabic : the case of Galen's *On the faculties and powers of simple drugs*, Book Six", in *Medieval Arabic thought : essays in honour of Fritz Zimmermann*, ed. by R. Hansberger, M. Afifi al-Akiti and Ch. Burnett (Warburg Institute studies and texts 4), London, Warburg Institute – Turin, Aragno, 2012, pp. 143-162.
4. http://www.digitalgalen.net.
5. *Ibid*.

Doi: 10.1484/J.SEC.1.103745

wavelengths to see what the separation limit was from the data available. He sampled a series of 200 points approximately uniformly across the page from four different categories: 50 from unwritten areas, 50 from areas where only the overtext was written, 50 from areas where only undertext was written, and 50 from areas where it could be judged with reasonable certainty that there was both over- and undertext overlapping. As mentioned above, this last group presents the key difficulty in reading this manuscript. The existing "sharpie" approach separates the two texts from the underlying parchment extremely well, but is unable to show when there is undertext beneath the overtext.

Bill Sellers arrived at the following preliminary conclusions. The overtext, undertext and parchment have clearly different spectra, and we should expect little difficulty in separating them. The spectra for the overtext and for the combined over- and undertext areas, however, are almost indistinguishable. The only real difference occurs in the infrared wavelengths, as the divergence between the overtext and the combined text seems to increase with increasing wavelength. This suggests that wavelengths longer than the current maximum of 940 nm may result in better separation. The 940 nm image shows no distinguishable difference between the overlapping text areas and the overtext areas, although a clear statistical difference can be identified in the sample.

To explore the segmentation of the image further, Bill Sellers used the 200 points mentioned above as input to a Canonical Variates Analysis. This statistical technique maximises the between group separation of multidimensional data.[6] In this case, each spectral image can be considered a different dimension so that each point has 23 intensity values associated with it. Canonical Variate Analysis shows that almost all the variation in the data can be explained using the first three canonical axes. We can plot pairs of canonical variate axes to illustrate the separation that can be achieved from the data. The separation between parchment, overtext and undertext is complete using the first two canonical variate axes, but this leaves the overtext and the combined text overlapping. By using the third canonical variate analysis, we can see limited separation between the overtext and the combined text, but as can be seen from the plot, there is still considerable overlap.

Canonical Variate Analysis generates the set of linear equations that maximises separation. Since we have three axes that contain useful information, we can use the equations for these axes to combine the 23 multispectral images and produce three Canonical Variate (henceforth CV) values for each pixel. These CV values can then be mapped to the red, green and blue channels of a colour image and will result in a combined image that maximises the colour separation of the four classes. The resulting image possesses, however, relatively low contrast because it includes the full spread of all the pixels in the image which includes elements that would not be classified into one of the four groups. Figure 1 shows part of an image where the contrast is increased because the range of values has been set from the range of the 200 test points rather than the whole image.

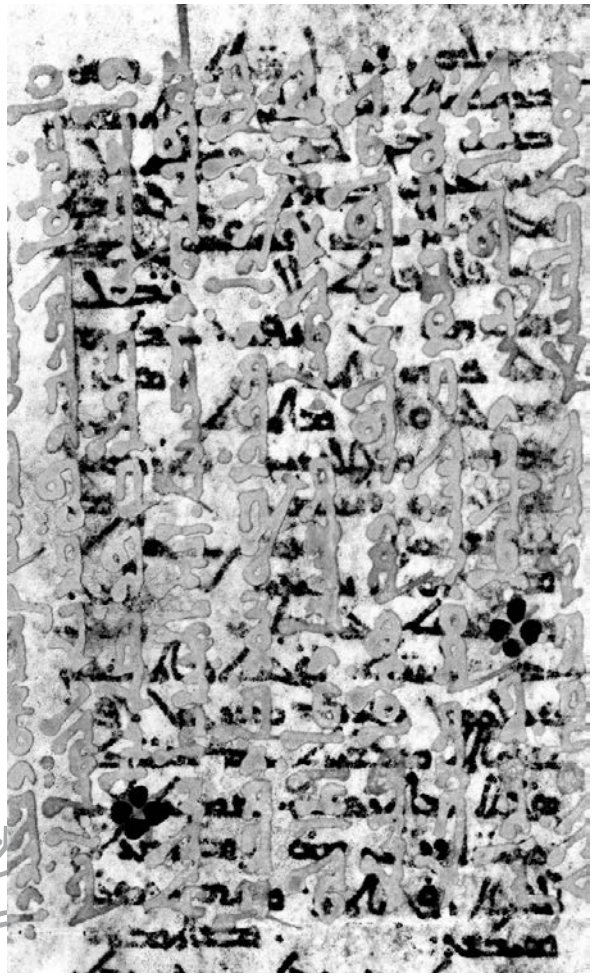


Figure 1 - Partial black-and-white version of the image resulting from Canonical Variates Analysis with enhanced contrast; the full range of images produced during the Canonical Variate Analysis can be found on Manchester eScholar at https://www.escholar.manchester.ac.uk/uk-ac-man-scw:211265.

6. R. A. Fisher, "The use of multiple measurements in taxonomic problems", *Annals of eugenics* 7, 1936, pp. 179-188; Zh. Guang and A. L. Maclean, "A comparison of canonical discriminant analysis and principal component analysis for spectral transformation", *Photogrammetric engineering & remote sensing* 66, 2000, pp. 841-847.

Evaluation of the preliminary data

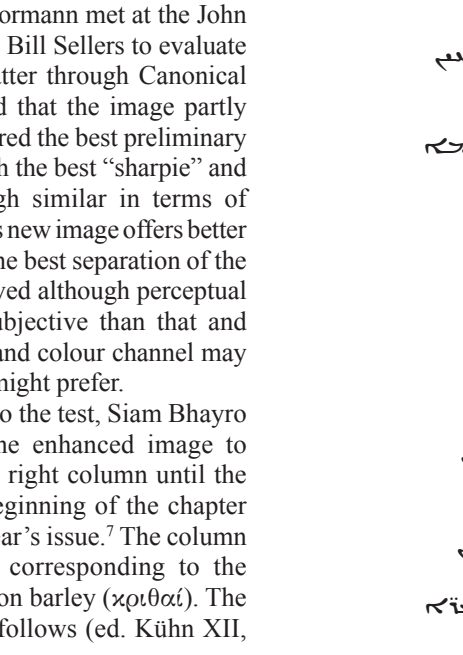
Siam Bhayro and Peter E. Pormann met at the John Rylands Research Institute with Bill Sellers to evaluate the images generated by the latter through Canonical Variate Analysis. It was agreed that the image partly reproduced as figure 1 here offered the best preliminary result. Moreover, compared with the best "sharpie" and pseudo-colour images, although similar in terms of legibility, the high contrast of this new image offers better readability. Statistically this is the best separation of the four elements that can be achieved although perceptual preferences are much more subjective than that and manual adjustment of contrast and colour channel may produce results that the reader might prefer.

To put this first impression to the test, Siam Bhayro and Peter E. Pormann used the enhanced image to transcribe the beginning of the right column until the gutter area, that is, until the beginning of the chapter on *κρίμνον*, published in last year's issue.[7] The column begins with a new sentence, corresponding to the second sentence of the chapter on barley (*κριθαί*). The underlying Greek text runs as follows (ed. Kühn XII, p. 44, line 12–p. 45, line 4):

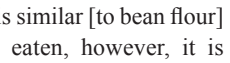
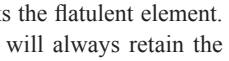
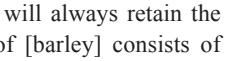
> εἰσὶ δὲ τοῦ τῶν κυάμων ἀλεύρου τοῦ χωρὶς τοῦ λέμματος βραχεῖ τινι ξηραντικώτεραι. τὰ δ' ἄλλα πάντα παραπλήσια χρωμένοις ἔξωθεν. ἐσθιόμεναι δὲ ταύτῃ πλεονεκτοῦσι κυάμων, ὅτι τὸ φυσῶδες ἀποτίθενται. κυάμῳ δὲ ὅπως ἂν ἑψηθῇ παραμένει τὸ φυσῶδες. παχυμερεστέρας γάρ ἐστιν οὐσίας ἢ κατὰ κριθήν, διὰ τοῦτο καὶ τροφιμώτερος αὐτῆς ἐστιν. ἐπεὶ δ' ὀλίγον ἄμφω τῆς μεσότητος ἀποκεχωρήκασι, διὰ τοῦτό εἰσι πολύχρηστοι. τὰ γὰρ τοιαῦτα φάρμακα πολλὰ ἑτέροις μίγνυται, καθάπερ τινὲς ὕλαι, καὶ διὰ τοῦτο καὶ ὁ κηρὸς καὶ τὸ ἔλαιον οὐκ ὀλίγοις ἐπιπλέκεται φαρμάκοις. τὰ δ' ἄλφιτα πολὺ καὶ τῶν κριθῶν αὐτῶν ἐστι ξηραντικώτερα.

> [Barley] is slightly more drying than bean flour without the husks. In all other respects, it is similar [to bean flour] when applied externally. When eaten, however, it is superior to beans, because it lacks the flatulent element. However one cooks beans, they will always retain the flatulent element. The essence of [barley] consists of thicker parts than that found in wheat. For this reason, the former is more nourishing than the latter. Because both [barley and wheat] are removed from the mean, they have many uses. These drugs are often mixed with others, such as other substances; for this reason, wax and oil are frequently combined with many other drugs. Barley-groats [*ἄλφιτα*] are much more drying than barley itself.

Here is the diplomatic transcription and literal translation of the Syriac text in SGP:



> And in all of these others it is like those that are used externally. And those that are eaten surpass beans because they do not make flatulence. And the bean if boiled always possesses that which causes flatulence. For its substance swells more than barley and therefore also nourishes more than these. And because both of them are a little distant from balance and moderation, because of this they are most useful. For medicines such as these are mixed with many others (like) some juices; and therefore also wax and oil are concocted with many remedies. And the flour of barley which is tender, which is called *álphita*, also dries more than that of barley.

In comparison with BL, there are two different readings. First, in lines 15-16, BL reads  for SGP's  . The BL reading is preferable, hence our restoration of "(like)" in the above translation. Second, in line 16, BL reads  for SGP's  . In this case, the reading in SGP is preferable, as the extra conjunction appears to be superfluous. There is a clear difference between the assumed underlying Greek version and the Syriac version, with the Syriac asserting that barley is the less nourishing substance.

Possible workflow and future enhancements

The process of selecting 200 points for the Canonical Variate Analysis and of conducting this analysis is straightforward and does not take much

7. Bhayro *et al*. (*supra*, n. 1), p. 262.

time. It can therefore easily be repeated for each page of the palimpsest, and each page will have its own optimised colour separation. The analysis also identifies the scope of possible text separation, and shows that with the current spectral wavelengths there is very little information below the overtext. There is clearly a tendency for areas where there is undertext to affect the amount of red and infrared light that is reflected but the spatial resolution of this effect is relatively low and does not seem to help make the undertext more legible. The other advantage of this approach is that it is completely general. It can use any number of spectral images, and will use all the information present. The training stage allows the reader to select the regions that are of interest for any particular task.

The statistical output shows how much separation is possible for a given set of images. It also allows the spectral characteristics of particular parts of the image to be estimated, which may inform further work. Thus whilst currently there does not appear to be enough information in the infrared images to read obscured undertext, the spectra are beginning to diverge at these wavelengths so further data collection using a camera that is sensitive at longer wavelengths might prove useful. Paradoxically, because the problem appears to be the opacity of the ink used for the overtext, it may be that much shorter wavelengths could also be usefully tried, since these are able to penetrate to greater depths. The efficacy of these approaches, however, depends on the chemistry of the pigments used and this is currently unknown.

Conclusion

Ultimately, the goal of any image enhancement exercise is to present readers with a picture that best enables them to read the text. The approach presented allows us to take the information from the full set of multispectral images and produce a single colour image that contains almost all the useful information present which should therefore provide an excellent starting point. It also allows us to evaluate the information content of the multispectral images objectively so that we can decide what is likely to be possible with the information currently available, and may inform subsequent re-imaging using different techniques. Giving scholars the ability to manipulate and adjust the existing images for themselves, and according to their own preferences, is probably the most cost-effective way of facilitating the further decipherment of SGP. The production of software that enables this is a strong desideratum.

s.bhayro@exeter.ac.uk
University of Exeter

peter.pormann@manchester.ac.uk
University of Manchester

William.Sellers@manchester.ac.uk
University of Manchester