manuscript cultures

Publishing Information

Natural Sciences, Technology and Informatics in Manuscript Analysis

Edited by Oliver Hahn, Volker Märgner, Ira Rabin, and H. Siegfried Stiehl

Proceedings of the third *International Conference on Natural Sciences and Technology in Manuscript Analysis* and the workshop *OpenX for Interdisciplinary Computational Manuscript Research* that took place at the University of Hamburg, Centre for the Study of Manuscript Cultures, on 12–14 June 2018.

CONTENTS

ARTICLES

Article

Writer Identification and Script Classification: Two Tasks for a Common Understanding of Cultural Heritage

Dominique Stutzmann, Christopher Tensmeyer, Vincent Christlein | Paris, Provo, UT, Erlangen

Abstract

Writer identification and script classification are usually considered as two separate and very different tasks, in palaeography as well as in computer science. Following the ICDAR competition on the CLAMM corpus about script classification and dating, this paper proposes to reconsider the tasks and methods of *Palaeography* and *Computer Vision* applied to *Artificial Palaeography*. We argue that, when aiming at understanding past societies and written cultural heritage in their complexity, palaeography and its core tasks may be defined as discretising the historical and social continuum at different levels of granularity. In this sense, we can consider writer identification and script classification as a single task. We then transfer this hypothesis to computer science, by running two infrastructures created for script classification on a more homogeneous dataset with a focus on writer identification. The analysis of the results confirms the uniformity of these tasks and allows us to reflect in a novel way on how to demonstrate and illustrate the historical continuum while discretising it in a non-binary way.

1. Introduction

The discipline of palaeography is quite often mistakenly defined, especially in French or Spanish, as the ability to read old documents and books. Palaeographers, however, generally acknowledge three 'levels' of palaeography: (i) reading; (ii) expertise; and (iii) historical analysis, as stated in the very first sentence of Cencetti's standard work *Lineamenti: di storia della scrittura latina:*

> La 'Paleografia' è lo studio critico delle antiche scritture ed è suo scopo non solo (i) interpretare esattamente i manoscritti, (ii) ma anche datarli, localizzarli e, (iii) in generale, trarre dal loro aspetto esteriore tutti gli elementi utili allo studio del loro contenuto e in generale alla storia della cultura.[1] [numbering added].

> 'Palaeography' is the critical study of ancient scripts and its purpose is not only (i) to accurately interpret the manuscripts, (ii) but also to date them, identify their place of origin and, (iii) in general, extract from their external appearance all the elements useful for the study of their content and in general for the history of culture.

Computer science also has the tasks related to these three levels: (i) handwritten text recognition and word spotting; (ii) writer identification; and, tending towards (iii), script classification. Here, terms have to be defined. In computer science and in general language, 'script' means 'writing system', for instance in Wikipedia articles about Arabic and Latin scripts,[2] and 'script type' refers to what the palaeographer Malcolm Parkes – and most English-speaking palaeographers after him – would call a 'script'. Indeed, he gave the following definitions: 'A *script* is the model which the scribe has in his mind's eye when he writes, whereas a *hand* is what he actually puts down on the page.'[3] Marc Drogin wrote: '[Among] styles of writing [that] were constantly undergoing changes […], we can give names to certain segments of this changing pattern […]. There are obvious and relatively long-lived styles that we can recognize and term *scripts*.'[4] For instance, Textualis is such

---

[1] Cencetti 1997.

[2] 'Arabic Script' 2018; 'Latin Script' 2018.

[3] Parkes 1969, xxvi.

[4] Drogin 1989, 4.

a 'script' in palaeographical terms, 'script type' in general terms or a long-lived 'style of writing' in the words of Marc Drogin, whose characteristic features are a double bowed 'a', a loopless 'l' and a long 's' (ſ) standing on the baseline. For the sake of clarity and unless stated otherwise, this paper will consistently use the terms 'script' for the more general level of 'writing system' and 'script type' for the long-lived styles of writing.

In this paper, we argue that, when aiming at understanding past societies and the written cultural heritage in their complexity, palaeography and its core tasks may be defined as discretising the historical and social continuum. Thus, the different tasks involved in palaeography can be considered as different levels of granularity, rather than a three-stage discipline. We can list the following interconnected components: (a) distinguishing script/text from non-script/non-text, (b) distinguishing one script ('writing system') from another, (c) identifying what the text is, (d) dating or (e) localizing and pointing to a specific context of production, e.g. *scriptorium*, (f) pointing to a specific writer, (g) classifying as a script type and (h) understanding the cultural history. To perform their analytical work, palaeographers tend to agree to base their arguments on the seven aspects of a medieval hand introduced by Jean Mallon and listed here according to Mark Aussems' translation: (i) form, i.e. morphology of the letters, (ii) angle of writing, specifically pen angle, (iii) ductus, (iv) 'modulus', i.e. the dimensions of the letters, (v) contrast, i.e. the difference in thickness between hairlines and shadow lines, also called 'weight', (vi) writing support and (vii) internal characteristics.[5] The respective importance of each of these aspects, however, remains subject to debate.[6] Some of the palaeographic tasks listed above are in common with or imply a connection with other disciplines, not only those related to the internal characteristics that are connected to linguistics and history (e.g. text identification, scholarly editing, interpretation of scribal variants etc.), but also those dedicated to the 'writing support' and layout, connected with codicology, philology, diplomatics or the broader study of the '*forma mentis*'.[7] In short: there are common analytical tools for different tasks in the humanities side of palaeography,

and these tasks may be defined in various manners and at different levels.

On the other hand, computer science has developed different tools for different tasks, as evidenced by the last competitions organized in ICFHR and ICDAR: layout analysis,[8] reading or indexing,[9] writer identification,[10] and dating and script classification,[11] even if some offer a combined goal on separate 'tracks', such as layout and text recognition.[12] From a modelling perspective, however, one could argue that it is the same problem, namely to distinguish what belongs together and what does not, a question that can be addressed by the same means and in both cases entails a discretisation of a body of evidence that can be seen as a historical continuum. In the following, we will test and confirm this hypothesis. Section II discusses the fusion of script and writer identification in palaeography. Section III addresses the notion of distance in 'similarity measure' between script types, to open the discussion about historical and social continuum. Section IV applies the tools and analyses of script classification to writer identification. Finally, we conclude that the fitness of tools confirms the uniformity of the tasks.

2. Palaeography: writer identification and script classification

In palaeography, 'expertise'[13] is traditionally subdivided into two parts: on the one hand, identifying the writer or ascribing several written artefacts to one known or unknown scribe and, on the other hand, dating and localizing a manuscript artefact based on the form of its handwriting. Each task represents a different perspective.

Writer identification focuses on individual characteristics and, in its fundamental principles and goals, is related to forensics, so that, despite the divide between the fields, professors of palaeography at the École nationale des chartes and other 'chartistes' have appeared in court as experts in prominent cases such as the Dreyfus affair and the murder of Grégory Villemin. Dating and localizing, on the other

---

[5] Mallon 1952; Aussems and Brink 2009; Muzerelle 2013.

[6] Stutzmann 2015.

[7] Barret, Stutzmann, and Vogeler 2016.

[8] E.g. Simistira et al. 2017.

[9] Andreu Sánchez et al. 2017.

[10] E.g. Fiel et al. 2017.

[11] E.g. Cloppet et al. 2017.

[12] Clausner, Antonacopoulos, and Pletschacher 2017.

[13] Cf. Gilissen 1973.

hand, do not focus on individuals, but on the zeitgeist and the general characteristics of a given era, and searches for the place or time in which the written artefact would best fit (usually a country or region and a period of half a century or less).

At a medium level of granularity, palaeographers focus on specific areas and times. In this case, researchers try to single out what is common to a group of scribes and to understand how this common ground evolves over time.[14] The environment can be more or less closed, homogeneous, subject to external influences and develop more or less specific writing shapes and styles, e.g. monastic *scriptoria*, secular chanceries, or Florentine merchants. This kind of research is generally linked to the third level of palaeography as cultural history. Indeed, studying a *scriptorium* leads directly to the intellectual history of a social group and its individuals.[15] However, with its technical component of expertise, it highlights that there is no divide between the study of autographs and writer identification, on the one hand, and dating and localizing, on the other, but a continuous scale of times and places in which palaeographers will attempt to place a written artefact, from an individual to a *scriptorium* or workshop, to a school, to a milieu, to a country and century.

Here, we need to add that the social and palaeographical continuum is multidimensional. Most societies experience some sort of polygraphism, i.e. the use of several writing systems or several script types by different individuals or by the same individual in different situations. Observing the adoption and implementation of script types by scribes can indeed be meaningful for historians. For example, the Hybrida script type was used by specific religious circles of the *devotio moderna* ('Devotenbastarda'),[16] and particular script types were used for liturgical books and books of hours in the fourteenth to sixteenth centuries.[17] Polygraphism is a challenge not only for writer identification,[18] but also for palaeographical study itself, because to assess polygraphism, one needs to define the 'script types' in question before analysing the places and times in which they

may occur and their relations and reciprocal influences as evidenced by specific features. As a consequence, the notion of palaeography as cultural history cannot be a separate theoretical domain, encompassing wider historical research *in abstracto* and studying script evolutions in time and space to interpret the modes of production and reception of the written object in past societies, without getting back to the individuals and the very graphical part of the study of handwriting. Indeed, this level of study implies the combined analysis of graphic and extra-graphic factors for script evolution, and understanding the role of the individual in society and in communication processes. This task is connected to linguistics, sociology, anthropology, cognitive sciences etc., but also specifically to the question of milieu and the social uses of writing, including the study of script classifications.

Script classification is therefore, once again, a prerequisite for an in-depth study of a larger social and religious evolution. Indeed, any large-scale and statistical analysis needs to name its objects and to single out specific features or 'factors', because one has to be able to define features within each field before comparing fields. Indeed, as Marc Smith has put it, 'Devising a rational typology and nomenclature to encapsulate the dimensions and features of a vast number of scripts is a necessity, if experts are to agree on anything, but no easy task.'[19] Likewise, the palaeographer David Ganz appropriately quoted Jakob Burckhardt: 'It is the most serious difficulty of the history of civilization that a great intellectual process must be broken up into single and often into what seem arbitrary categories, in order to be in any way intelligible.'[20] In palaeography as well, any classification aims at discretizing a complex and interconnected reality that is largely a continuum.

Discretising is a difficult task *per se*. It is difficult even to differentiate between writing systems, and which script was used in some famous artefacts, such as the Namara inscription bearing the epitaph of king Imru'l-Qays, is itself an object of debate. The genealogy and distances between writing systems are likewise a field of research.[21] Polygraphism and the use of a script within another system (e.g. of Greek characters within a document otherwise written in the Latin alphabet)

[14] Nievergelt, Gamper, and Bernasconi 2015.

[15] E.g. Ganz 1990.

[16] Oeser 1992.

[17] Stutzmann 2017, 2019.

[18] De Robertis 2013.

[19] Smith 2016.

[20] Burckhardt 1982; Ganz 1990, 121.

[21] Hosszú 2017.

also brings about situations in which it is difficult to state to which writing system a particular sign may belong. For the study of more closed milieus, such as monastic *scriptoria*, with specific styles within a script type,[22] we need to identify specific features that appear consistently in several books and ascertain a chronology in order to explain the evolution within the abbey, in relation to successive abbots or to a specific cultural change. Indeed, the scribal landscape is like a huge maelstrom in which so-called canonised scripts (as defined by Cencetti)[23] form some islands and emerge with fixed morphological features.[24]

As a consequence of these difficulties and the diverging valuation of the features under scrutiny, for now there is no generally accepted classification.[25] Following Albert Derolez's proposal for a general classification of Gothic book scripts, Marc Smith discussed the arbitrariness of any such classification and that it is difficult to assess the extent to which a nomenclature, however based on objective features, truly reflects historical realities.[26] According to him, scholars should therefore avoid thinking in terms of fixed, arbitrary, sometimes 'elementary' nomenclatures, rather than perceiving the complex evolutionary mechanisms.

These needs and shortcomings together show that all tasks involved in forensic examination in palaeographical research are really a single task on different levels and, depending on historical reality itself, the focus can be set on virtually any point on a continuous scale:

a. inner individual variation < individual < closed group of individuals < larger group or region in a given time < world at a given time

b. script instances, 'scribal hands' < varieties of script types < script types < writing systems < visual semiotic systems[27]

3. Image analysis and script classification: introducing distance visualisation as heuristics

Parallel to the tasks of palaeography, in recent years there have been two different series of competition for the classification of historical scripts. One is on writer identification, the other on script classification. The former is by far the most prominent. It aims to identify several writing samples as having been produced by a single person,[28] even in different writing systems or a 'multi-script environment'.[29] In these competitions, the measured dissimilarity of the handwriting is generally considered to be the consequence of difference between individuals, but some research addresses the notion of script style and other biologically or sociologically defined groups.[30] Regarding the latter, it must be noted that gender, i.e. a social difference, is probably a better explanation than biological difference. Studies combining image analysis, genetics and social sciences would be welcome. In studies supposedly demonstrating the influence of hormones,[31] the difference in the expression of masculine and feminine traits is also demonstrated and probably to be linked to social behaviour, as in historical scripts, where the social division of male and female and their distinct association with script types has been observed (e.g. Cursiva for men, Textualis for women[32]).

The notion of the division of script types is rarely addressed in image analysis, even when, as in the latest competition on historical writer identification,[33] the corpus covers such a large chronological span (with 720 writers) that the differences between writing samples may reflect script types rather than merely individual variety.

The other strand, started in 2016, is the competition on Latin handwriting classification,[34] based on the CLAMM corpus.[35] The competitors were required to provide a normalized belonging matrix (a CSV file recording the

---

[22] Palaeographers sometimes call these styles 'types', sometimes using the German word 'Schrifttypen'.

[23] Cencetti 1978, 1997.

[24] Stutzmann 2018.

[25] Stutzmann 2015.

[26] Smith 2004.

[27] Here, following our vocabulary, we introduce the notion of 'variety of a script type' to render what Malcolm Parkes called 'the varieties of a single script'; Parkes 2008, 152.

[28] Fiel et al. 2017.

[29] Djeddi et al. 2015.

[30] Maadeed et al. 2012; Djeddi et al. 2015, 2016.

[31] Beech and Mackintosh 2005.
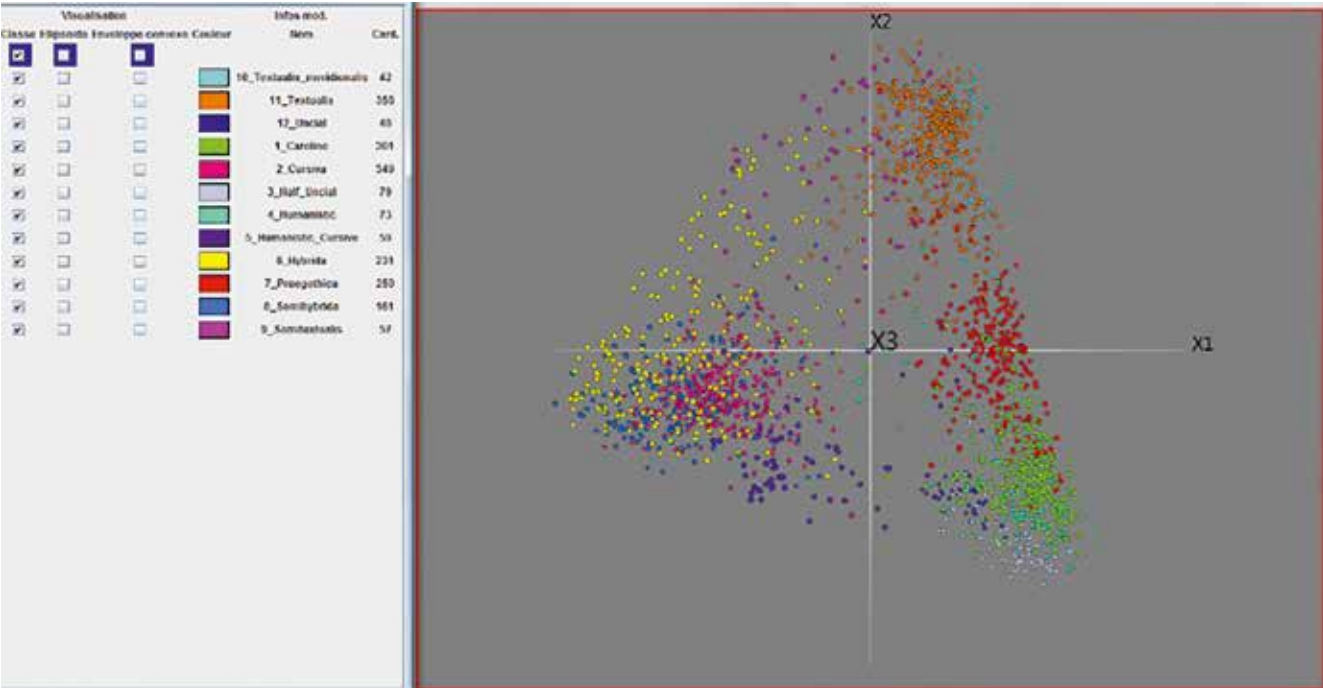
[32] Stutzmann 2012.

[33] Fiel et al. 2017.

[34] Cloppet et al. 2016, 2017.

[35] Stutzmann 2016.

Table 1: ICDAR2017 PCA on Christlein's belonging matrix.
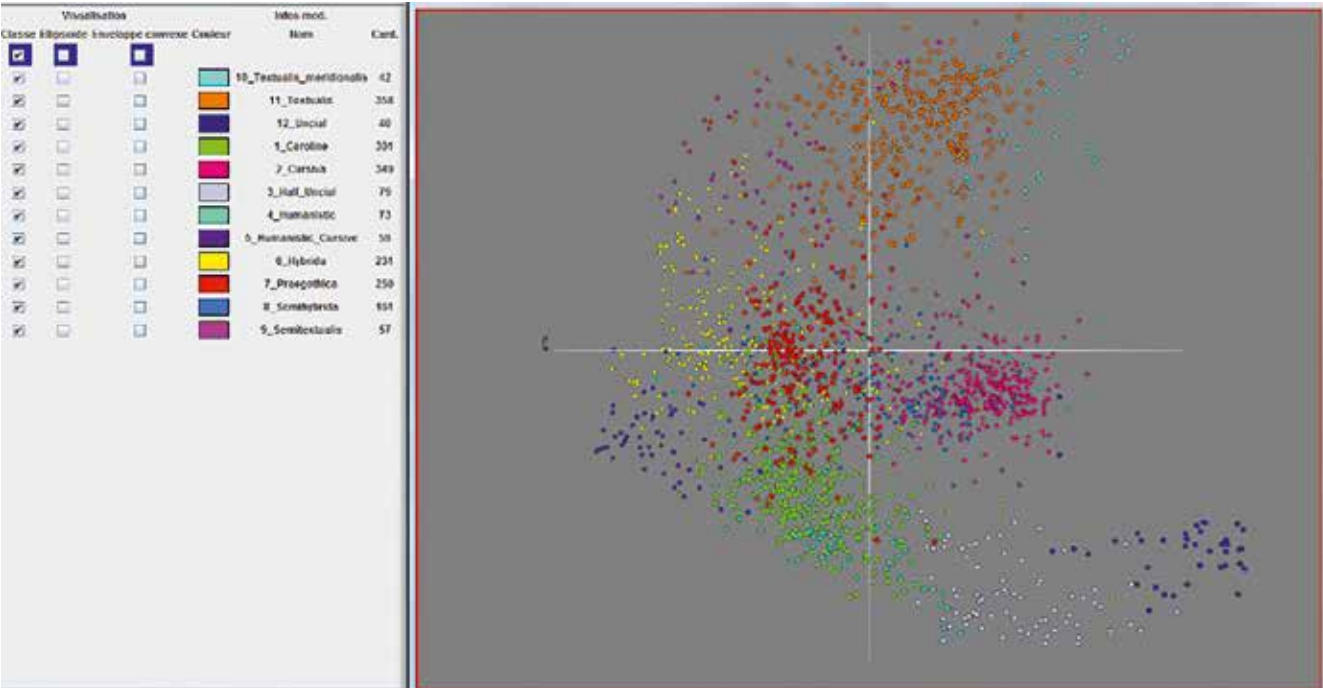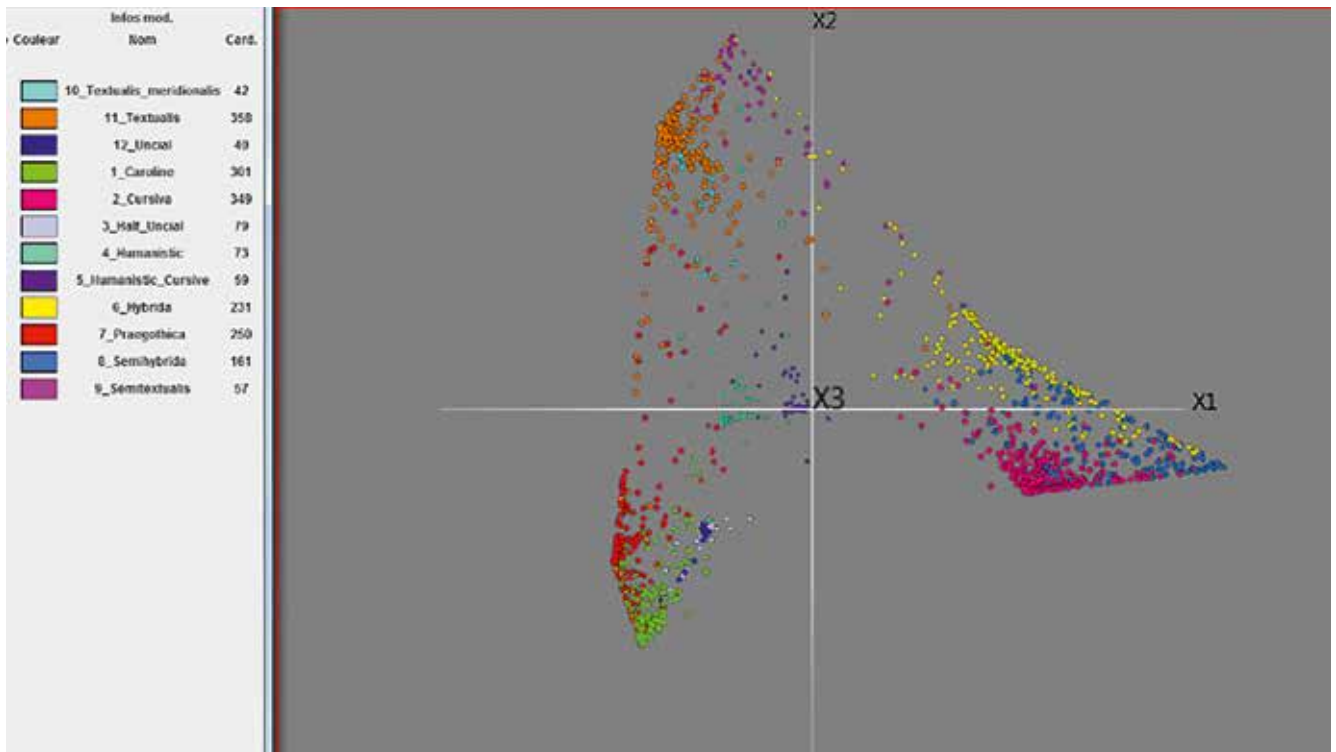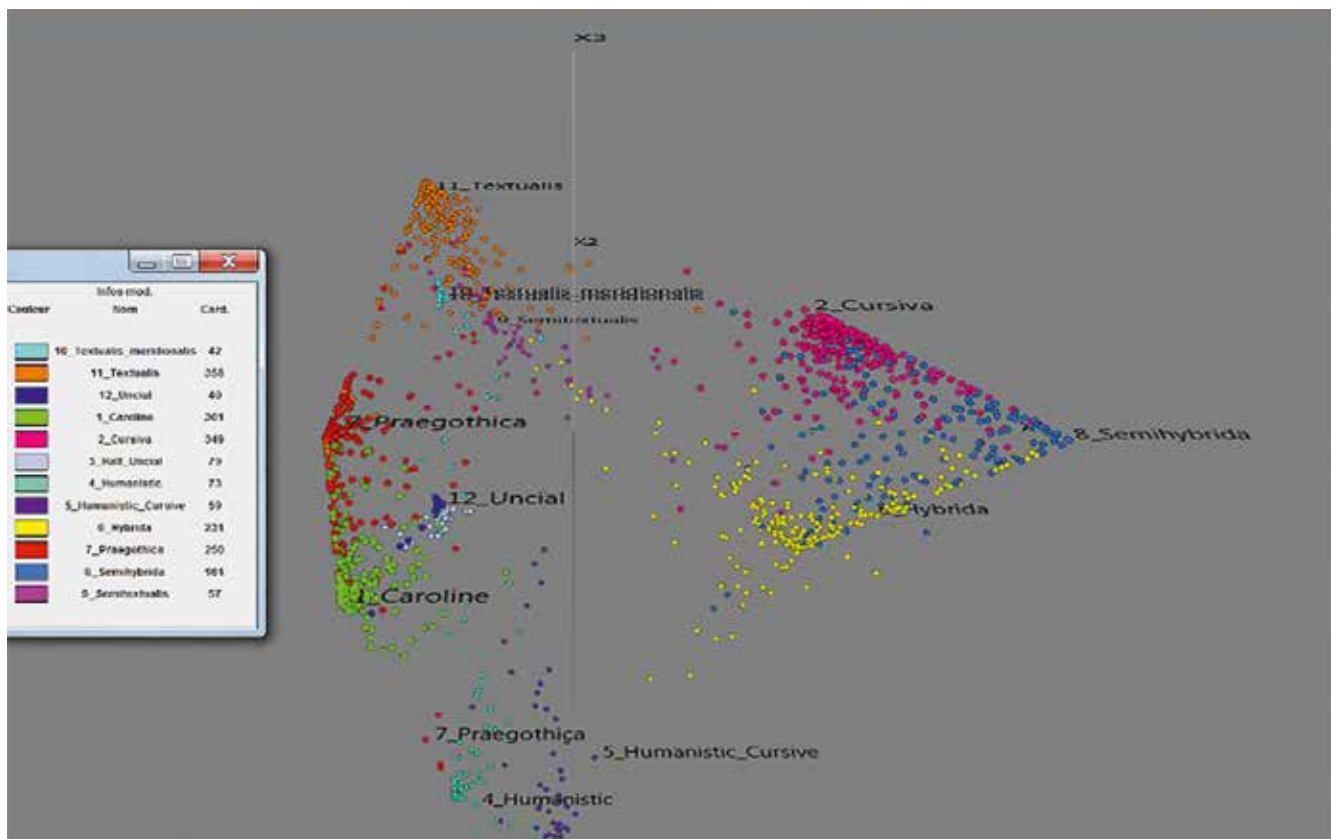
Axes 1 and 2

Axes 2 and 3

Table 2: ICDAR2017 PCA on Tensmeyer's belonging matrix.

Axes 1 and 2

Axes 1 and 3

relative score of all twelve proposed script type labels for each image) and a distance matrix describing all images. The evaluation was based on the belonging matrix, in which the script type with the highest score had to match the ground-truth label. A first palaeographical interpretation was given based on the confusion matrices, showing similarities and recurring confusions between some of the script types.[36]

Based on the belonging and distance matrices, we can establish a new heuristic and interpretation mode, as we suggested in 2015 and then implemented and commented with Mike Kestemont.[37] Principal component analysis (PCA) of the belonging matrix easily reduces the 12 dimensions to a three-dimensional space. Here, we use the Explorer3D software.[38] For the following interpretation, we compare the data from the two best-performing systems in the ICDAR 2017 competition. One was developed by Vincent Christlein,[39] the other by Christopher Tensmeyer.[40] The results are quite different.

In Christlein's results (Table 1), the first two axes of the PCA analysis can be represented as a triangle, and the points gather: (a) Uncial, Half-Uncial, Caroline and Humanistic, (b) Textualis, Textualis meridionalis and Semitextualis; (c) Cursiva, Hybrida and Semihybrida. One script type, namely Praegothica, forms the edge between poles (a) and (b), as does Humanistic Cursive between (a) and (c), while Semitextualis connects with Hybrida. A view along axes 2 and 3 highlights the fact that the script types in the three poles are also well defined.

In Tensmeyer's results (Table 2), the best visualisation combines the first three axes and gives a different view of the relations between scripts, with the position of the Humanistic group and that of Praegothica being greatly changed within a space with four main zones: (a) Uncial, Half-Uncial, Caroline and most Praegothica samples belong in one corner; (b) Cursiva, Semihybrida and Hybrida in another corner; (c) Textualis, Semitextualis, and Textualis meridionalis in a compact zone with two corners; and (d) Humanistic minuscule and Humanistic Cursive in a distinct

zone that appears to be almost completely autonomous (except for the fact that, in this zone, Humanistic minuscule is, as expected, on the side of Caroline and Textualis, while Humanistic Cursive is closer to Cursiva).

These results, despite or because of the differences, are very valuable for freshening our perspective on the history of the Latin writing system. The training data set contained only labels and no statement about the distance between scripts in history. Despite this, not only do the results correspond to our historical knowledge, but the discrepancies between the two sets of results also correspond to on-going palaeographical debates. They force historians to reconsider the respective positions of each script. They also illustrate how entangled the different script types are. At the same time, coeval script families may be either very clearly separated (e.g. Textualis and Cursiva) or completely integrated (e.g. Cursiva and Semihybrida). This is a first illustration of the historical continuum, in which some script families emerge and separate from a common stem.

Another visualisation of this continuum is to represent the 2000 × 2000 distance matrix of all images. This is also done with the Explorer3D software and uses classical multidimensional scaling (MDS). The first observation is that Christlein's and Tensmeyer's distance matrices are surprisingly different. In Christlein's method, a major step is the definition of the features of the learnt classes,[41] and the distances between the images themselves do not initially appear to be very different (Table 3). Indeed, a 3D projection shows a compact, populated kernel and some secluded groups. The images of each script are not distributed in coherent classes with a small intra-class distance, but rather along axes either starting at the centre or traversing it and developing mostly in one direction each. This happens in such a way that two otherwise clearly separate scripts, such as Cursiva and Textualis, are distributed in axes pointing almost in the same direction, even if the subsequent classification distinguishes them without mistake. This is an intriguing view of the history of script. Each family comprises samples that are closer to the other families than to its siblings. And, here, 'intriguing' does not mean 'false': on the contrary, it corresponds perfectly to the experience of palaeographers who can always find script images for which the so called 'first impression' is misleading or may correspond to other features than letter shapes.

---

[36] Cloppet et al. 2016, 2017.

[37] Stutzmann 2015; Kestemont, Christlein, and Stutzmann 2017.

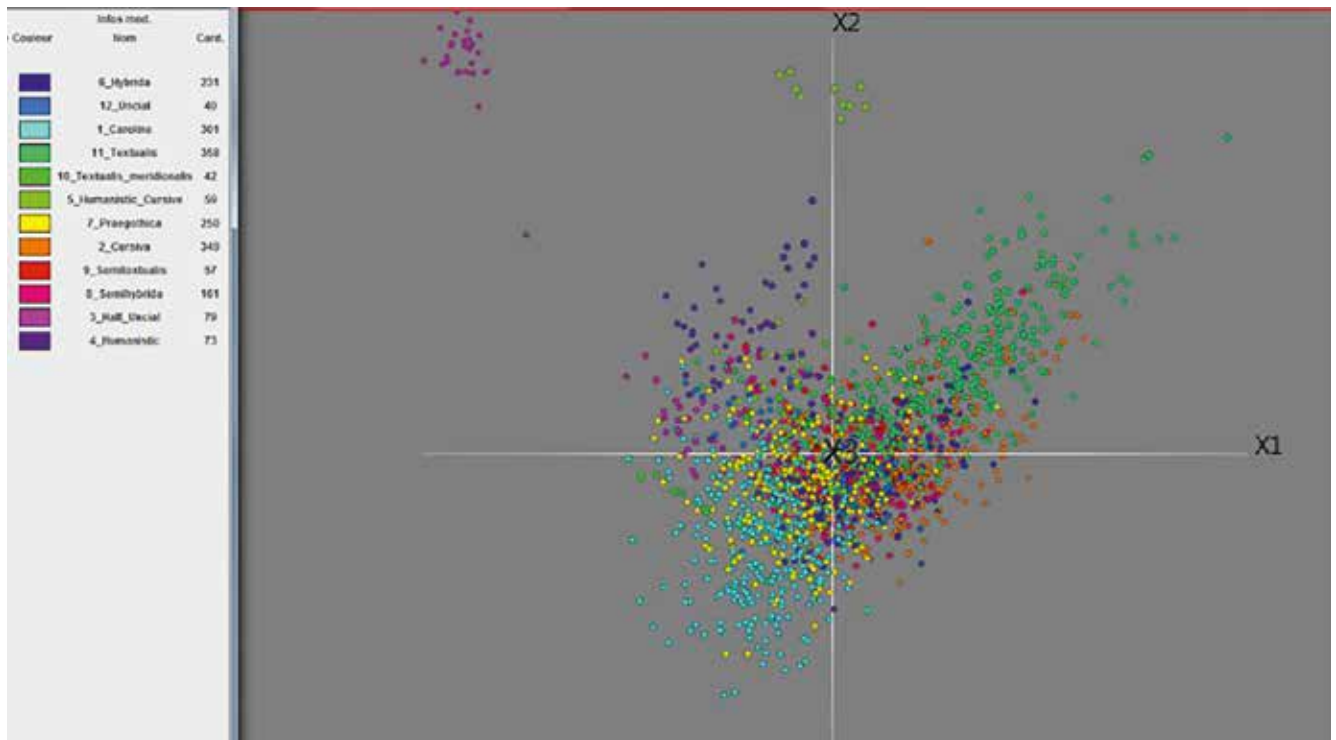[38] Exbrayat and Martin 2017.

[39] Christlein et al. 2017; Christlein [2017] 2018.

[40] Tensmeyer, Saunders and Martinez 2017; Christopher Tensmeyer 2017.

[41] Kestemont, Christlein, and Stutzmann 2017.

Table 3: ICDAR2017 Christlein's distance matrix.

Axes 1 and 2

Axes 2 and 3

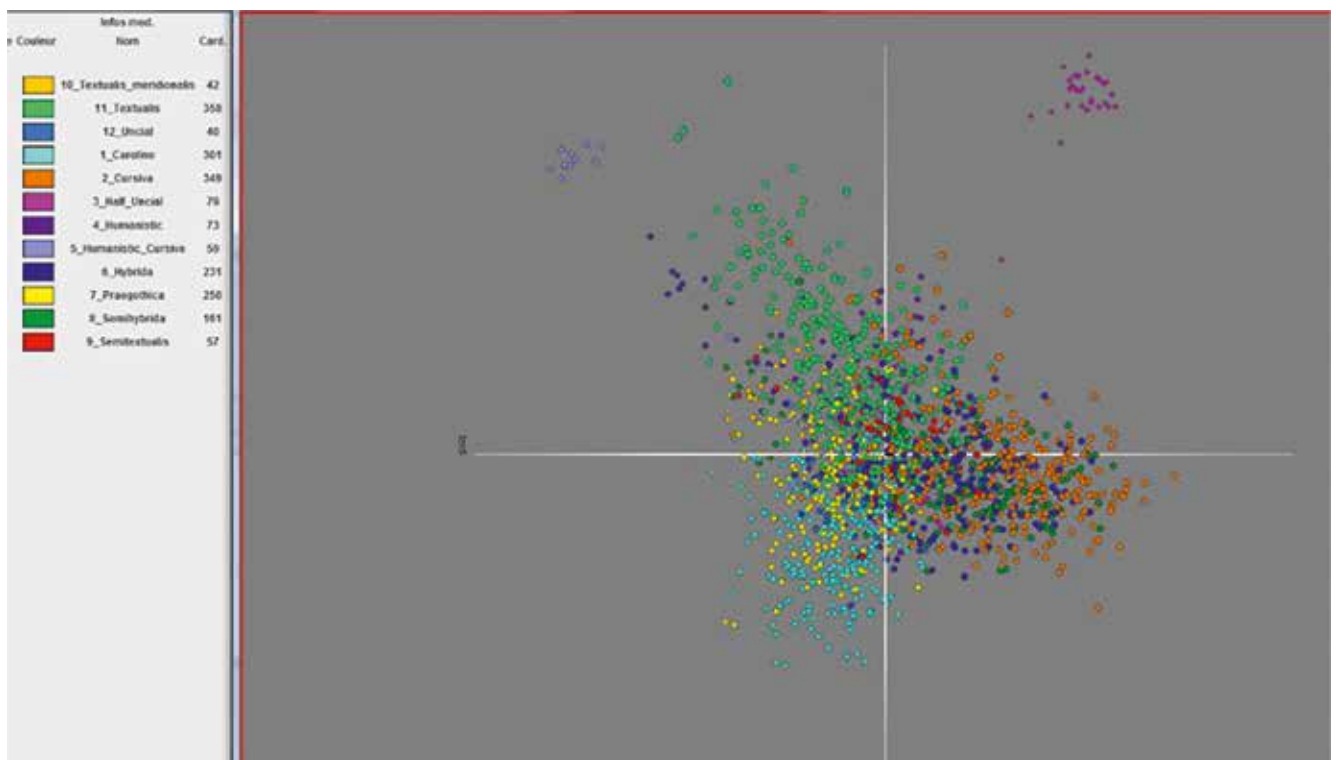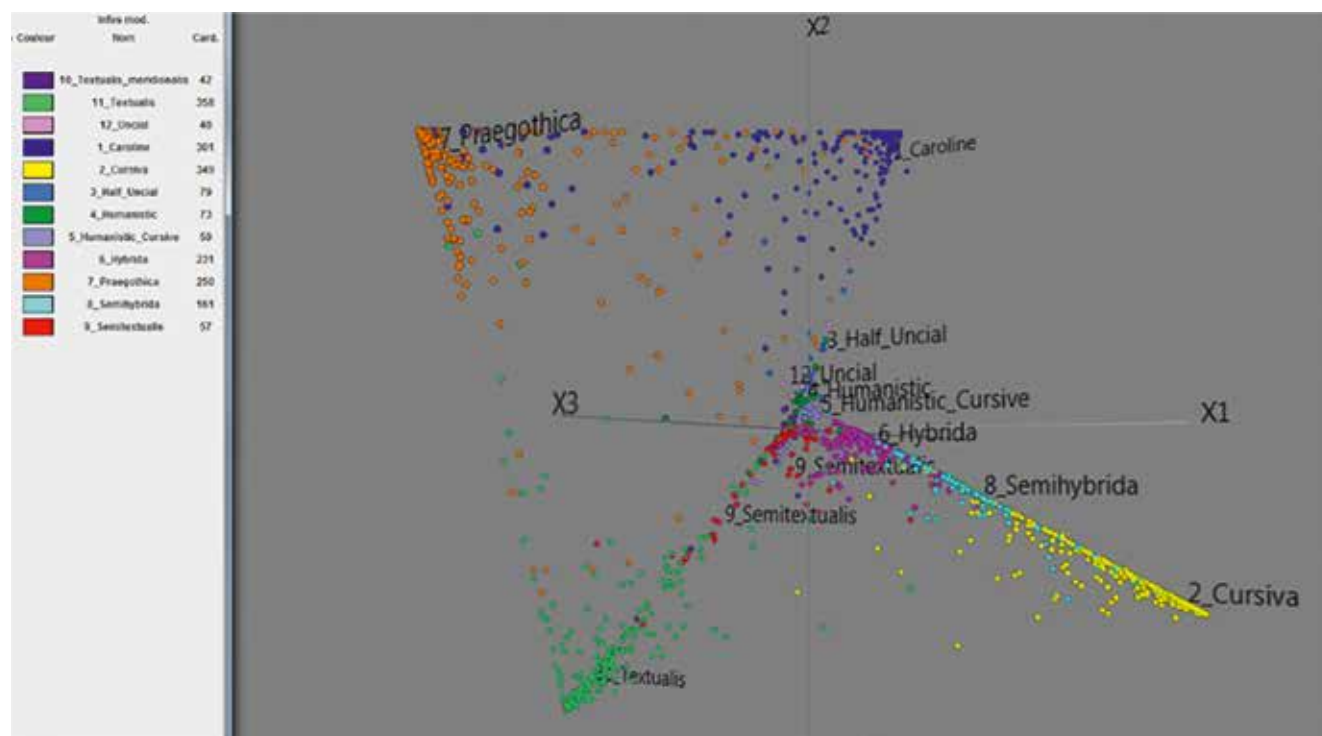Table 3: ICDAR2017 Christlein's distance matrix.

Table 4: ICDAR2017 Tensmeyer's distance matrix .

Axes 2 and (1 and 3)

Tensmeyer's distance matrix (Table 4), however, can be represented in a shape that is much more similar to our view of the belonging matrix. The shape is a triangular pyramid, whose edges are populated very differently, as well as diagonals to the centre of gravity. The corners are Caroline, Praegothica, Textualis and Cursiva. The centre of gravity or central node joins, in an unexpected way, Humanistic and Humanistic Cursive to (a) Uncial and Half-Uncial (very close, but rather on the diagonal to Caroline), (b) Hybrida and Semihybrida on the diagonal to Cursiva or slightly dispersed on the plane between Cursiva and Textualis and (c) Semitextualis and Textualis meridionalis on the diagonal to Textualis. The whole representation is striking, because one can make some sense of it, but only in a very fuzzy way. The centrality of Humanistic, Humanistic Cursive, Uncial, Half-Uncial and, to some extent, Semitextualis may correspond to what students perceive as the readability or legibility of a script, but it clearly does not correspond to any historical evolution, nor to morphological features at the level of single letters (according to which Humanistic should be located between Caroline and Praegothica). In this sense, we are quite uncertain how to interpret the distance matrix. It is interesting to note that this system also attempts to analyse the non-textual components of the manuscript pages to classify the script type of the page and that there is a significant bias about which scripts could be identified from non-textual content (e.g. Praegothica, Humanistic).[42] It is not clear how much this implementation influences the present output.

Nevertheless, this kind of visualisation is very closely connected to what one palaeographer envisioned in the 1970s as being a key model for a 'Cartesian nomenclature',[43] that is, a 3D space in which each zone would correspond to a specific script – with the only exception that axes would be defined according to a limited set of allographs.

Distance visualisation on the results of script classification is a way to gain access to the notion of a historical continuum. It does not abolish the need for or the validity of existing classifications, but it allows us to go back to the most confusing artefacts and then either reshape our criteria or understand the features that influence our perception and, at the same time, better understand what the system has learnt.

---

[42] Tensmeyer, Saunders, and Martinez 2017.

[43] Gumbert 1976.

Table 5: Tensmeyer's belonging matrix for duplicate images (red = tif ; blue = jpg).

For three documents, the thumbnails show that jpg and tif are not classified identically.

Table 6: Christlein's belonging matrix.
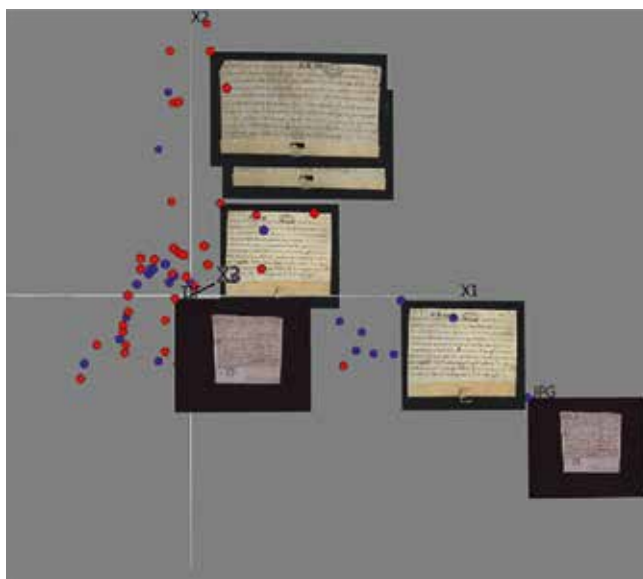
4. Image analysis and writer identification

Section III has demonstrated that systems designed to identify script classes may produce belonging and distance matrices in which a palaeographer may try not only to validate the labelling, but also to explore the distances between script classes and the distribution of script samples within the class and in respect to other classes. Closeness may be interpreted in morphological and phylogenetic-historical terms.

In this section, we examine the same systems applied to a new data set built for writer identification. The systems have not been retrained with new labels, but only applied to new images as if they were part of the original test set. The new data set encompasses 377 images of written documents and books produced or received by the Cistercian abbey of Fontenay during the twelfth century. Among them are thirty-seven pairs of duplicate images, present as tif and jpg files, to test the impact of format on the systems. For the documents, we created a ground truth based on our study
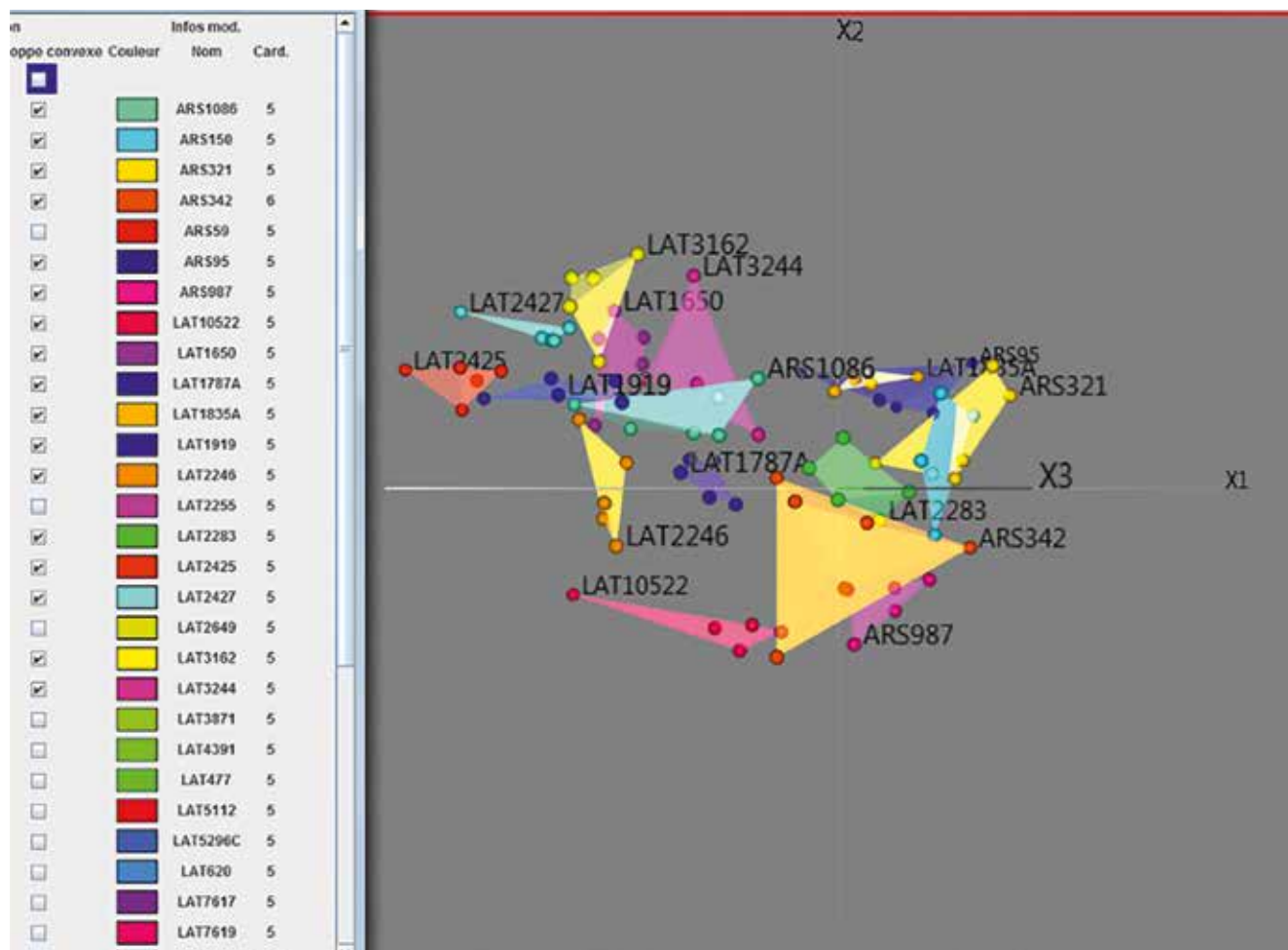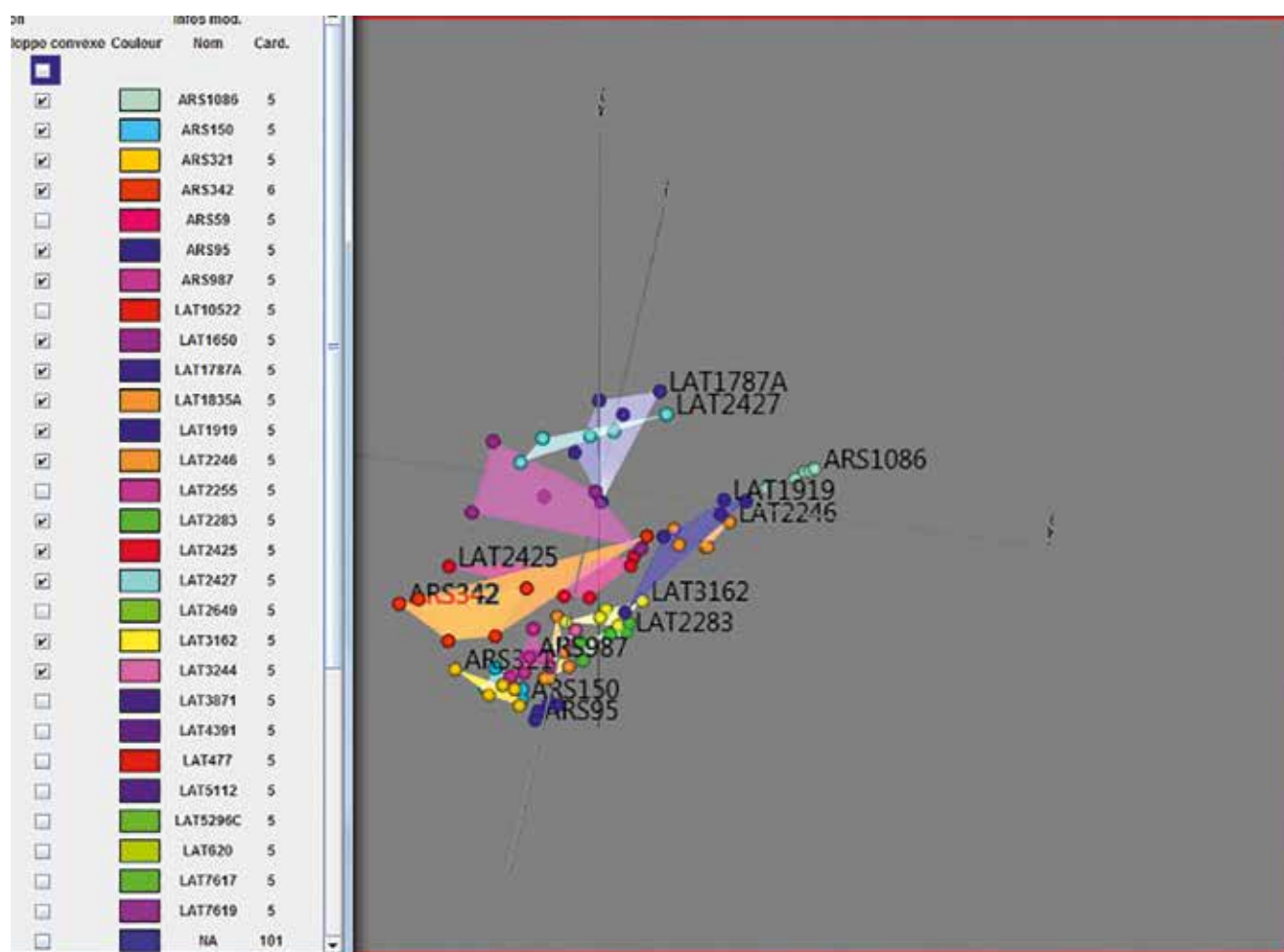
Table 7: Tensmeyer's belonging matrix.

of the scriptorium.[44] For the books, we used five pages of each volume and considered that each of these samples of five pages could represent one anonymous 'hand' – that is, a single writer in a single context. The relative distance/closeness between the different sets of anonymous hands can be a starting point for suggesting the identity of separate writer sets, for which there is, however, no formalized ground truth.

A first comparison is based on the duplicate images in jpg and tif and between grey-scale and colour images (Table 5). Both systems deal accurately with both grey-scale and colour images. Tensmeyer's system uses two separate analysis mechanisms for tif and jpg, so that the two spaces overlap but are not identical, and duplicate images in different formats may be somewhat separated

However, for both Tensmeyer and Christlein, if we compare images in the same format, we see very coherent and compact spaces for each manuscript (same scribe, same moment). If we add the script classes used in documents, written either by different scribes or by a single scribe at different moments and in different sizes and layouts, the spaces are slightly less homogeneous, but still offer a very neat classification. The following figures (Tables 6 and 7), representing the results of PCA on the belonging matrix, are reduced to the distribution of all five images from only sixteen different manuscripts. The full data is too compact to be easily illustrated, but each group of images pertaining to one writer is sufficiently compact to be distinguished from the other groups.

Here, we will not go into the details of the analysis. However, we can stress that the belonging matrix keeps the labels of the original task, which are totally irrelevant for the present task. The classification of the images in the present data set into different groups was based mainly on

---

[44] Stutzmann 2009.

morphological features,[45] which are not at all in line with those used for describing the long-term history of scripts. Christlein and Tensmeyer's systems, however, were trained to classify writing samples produced from the sixth to the sixteenth centuries. It is therefore striking and unexpected that they are able to 'zoom in' onto a very homogeneous *scriptorium* production and adequately separate different manuscripts and slightly differing styles within the closed milieu of a monastic *scriptorium*, when they rely on learnt features that we would not deem applicable to the present data set.

Once again, the visualisations based on the data generated by each system do not render exactly the same landscape, but both measures are in line with palaeographical expertise and may shed light on the debate about respective dates of production (given that none of the manuscripts is dated or signed).

5. Conclusion

This paper reunites writer identification and script classification as a single task, from a theoretical perspective not only in palaeography, but also in computer science. The hypothesis of unity is corroborated by the experiment made for this research: applying programs trained for script classification for the medieval millennium to writer identification in the closed and homogeneous production of a monastic *scriptorium*. Moreover, we also propose to use and analyse distance metrics as a way to show the historical and social continuum of written production. This opens a non-binary solution to the task of script or writer identification and a more balanced view of model (what is in the 'mind's eye'), influence, imitation and evolution. It also opens the discussion about the features used in image analysis, the meaning of measured distance and visualisation. Further experiments may now be organized. The first one should measure how well a unique scribe can be ascertained across different scripts and script types. The corpus under scrutiny would be composed of the autographs of known writers using several script types (*di-* or *polygraphism*). A second one should extend the notion of *continuum* that we used in this paper and test the measures of distance between different script systems (e.g. Latin, Greek, Hebrew, Arabic, Chinese etc.).

REFERENCES

Andreu Sánchez, Joan, Verónica Romero, Alejandro Héctor Toselli, Mauricio Villegas, and Enrique Vidal (2017), 'Competition on Handwritten Text Recognition on the READ Dataset', in *14th International Conference on Document Analysis and Recognition (ICDAR)*, vol. 1 (Institute of Electrical and Electronics Engineers: IEEE Xplore Digital Library), 1383–1388. doi: <10.1109/ICDAR.2017.226>.

'Arabic Script' (2018), in *Wikipedia*. <https://en.wikipedia.org/w/index.php?title=Arabic_script&oldid=857390673.8.2.2019> (accessed 19 November 2019).

Aussems, Mark, and Axel Brink (2009), 'Digital Palaeography', in Malte Rehbein, Patrick Sahle, and Torsten Schaßan (eds), *Kodikologie und Paläographie im digitalen Zeitalter – Codicology and Palaeography in the Digital Age* (Norderstedt: Books on Demand; Schriften des Instituts für Dokumentologie und Editorik 2), 293–308.

Barret, Sébastien, Dominique Stutzmann, and Georg Vogeler (eds) (2016), 'Introduction', in *Ruling the Script in the Middle Ages: Formal Aspects of Written Communication (Books, Charters, and Inscriptions)* (Turnhout: Brepols; Utrecht Studies in Medieval Literacy 35), 1–24.

Beech, John R., and Isla C. Mackintosh (2005), 'Do Differences in Sex Hormones Affect Handwriting Style? Evidence from Digit Ratio and Sex Role Identity as Determinants of the Sex of Handwriting', *Personality and Individual Differences,* 39.2: 459–468. doi: <10.1016/j.paid.2005.01.024>.

Burckhardt, Jacob (1982), *Über das Studium der Geschichte: Der Text der "Weltgeschichtlichen Betrachtungen" auf Grund d. Vorarbeiten von Ernst Ziegler nach d. Hs. hrsg. von Peter Ganz* (München: C.H. Beck).

---

[45] Stutzmann 2014.

Cencetti, Giorgio (1978), *Paleografia latina* (Roma: Jouvence; Guide allo studio della civiltà romana 10,3).

—— (1997), *Lineamenti di storia della scrittura latina, dalle lezioni di paleografia*, 2nd rev. edn, ed. by Gemma Guerrini Ferri (Bologna: Pàtron).

Christlein, Vincent, Martin Gropp, Stefan Fiel, and Andreas Maier (2017), 'Unsupervised Feature Learning for Writer Identification and Writer Retrieval', in *14th International Conference on Document Analysis and Recognition (ICDAR)*, vol. 1 (Institute of Electrical and Electronics Engineers: IEEE Xplore Digital Library), 991–997. doi: <10.1109/ICDAR.2017.165>.

Christlein, Vincent [2017] (2018), *Unsupervised Feature Learning for Writer Identification and Writer Retrieval: VChristlein/Icdar17code*, Python. <https://github.com/VChristlein/icdar17code> (accessed 19 November 2019).

Clausner, Christian, Apostolos Antonacopoulos, and Stefan Pletschacher (2017), 'ICDAR2017 Competition on Recognition of Documents with Complex Layouts – RDCL2017', in *14th International Conference on Document Analysis and Recognition (ICDAR)*, vol. 1 (Institute of Electrical and Electronics Engineers: IEEE Xplore Digital Library), 1404–1410. doi: <10.1109/ICDAR.2017.229>.

Cloppet, Florence, Véronique Églin, Marlène Helias-Baron, Cuong Kieu, Nicole Vincent, and Dominique Stutzmann (2017), 'ICDAR2017 Competition on the Classification of Medieval Handwritings in Latin Script', in *14th International Conference on Document Analysis and Recognition. (ICDAR)*, vol. 1 (Institute of Electrical and Electronics Engineers: IEEE Xplore Digital Library), 1371–1376. doi: <10.1109/ICDAR.2017.224>.

Cloppet, Florence, Véronique Églin, Van Cuong Kieu, Dominique Stutzmann, and Nicole Vincent (2016), 'ICFHR2016 Competition on the Classification of Medieval Handwritings in Latin Script', in *15th International Conference on Frontiers in Handwriting Recognition (ICFHR)* (Institute of Electrical and Electronics Engineers: IEEE Xplore Digital Library), 590–595. doi: <10.1109/ICFHR.2016.0113>.

De Robertis, Teresa (2013), 'Una mano tante scritture. Problemi di metodo nell'identificazione degli autografi', in Nataša Golob (ed.), *Medieval Autograph Manuscripts. Proceedings of the XVIIth Colloquium of the Comité International de Paléographie Latine, held in Ljubljana, 7–10 September 2010* (Turnhout: Brepols; Bibliologia, 36), 17–38. doi: <10.1484/M.BIB.1.101466>.

Djeddi, Chawki, Somaya Al-Maadeed, Abdeljalil Gattal, Imran Siddiqi, Abdellatif Ennaji, and Haikal El Abed (2016), 'ICFHR2016 Competition on Multi-Script Writer Demographics Classification Using 'QUWI' Database', in *15th International Conference on Frontiers in Handwriting Recognition (ICFHR)* (Institute of Electrical and Electronics Engineers: IEEE Xplore Digital Library), 602–606. doi: <10.1109/ICFHR.2016.0115>.

Djeddi, Chawki, Somaya Al-Maadeed, Abdeljalil Gattal, Imran Siddiqi, Labiba Souici-Meslati, and Haikal El Abed (2015), 'ICDAR2015 Competition on Multi-Script Writer Identification and Gender Classification Using 'QUWI' Database', in *13th International Conference on Document Analysis and Recognition. (ICDAR)* (Institute of Electrical and Electronics Engineers: IEEE Xplore Digital Library), 1191–1195. doi: <10.1109/ICDAR.2015.7333949>.

Drogin, Marc (1989), *Medieval Calligraphy: Its History and Technique* (New York: Dover Publications).

Exbrayat, Matthieu, and Lionel Martin (2017), 'Explorer3D', Université d'Orléans, Laboratoire d'informatique Fondamentale d'Orléans. <https://www.univ-orleans.fr/lifo/software/Explorer3D/> (accessed 19 November 2019).

Fiel, Stefan, Florian Kleber, Markus Diem, Vincent Christlein, Georgios Louloudis, Nikos Stamatopoulos, and Basilis Gatos (2017), 'ICDAR 2017 Competition on Historical Document Writer Identification (Historical-WI)', in *14th International Conference on Document Analysis and Recognition (ICDAR)*, vol. 1 (Institute of Electrical and Electronics Engineers: IEEE Xplore Digital Library), 1377–1382. doi: <10.1109/ICDAR.2017.225>.

Ganz, David (1990), *Corbie in the Carolingian Renaissance* (Sigmaringen: Thorbecke; Beihefte der Francia, 20). <http://www.perspectivia.net/content/publikationen/bdf/ganz_corbie> (accessed 19 November 2019).

Gilissen, Léon (1973), *L'expertise des écritures médiévales. Recherche d'une méthode avec application à un manuscrit du XIe siècle: le lectionnaire de Lobbes (Codex Bruxellensis 18018)* (Gand: E. Story-Scientia; Les publications de "Scriptorium", 6).

Gumbert, Johan Peter (1976), 'A Proposal for a Cartesian Nomenclature', in Johan Peter Gumbert and Max J.M. Haan (eds), *Essays Presented to G.I. Lieftinck*, vol. 4: *Miniatures, Scripts, Collections* (Amsterdam: van Gendt; Litterae Textuales, 4), 45–52.

Hosszú, Gábor (2017), 'Phenetic Approach to Script Evolution', in Hannah Busch, Franz Fischer, and Patrick Sahle (eds), *Kodikologie und Paläographie im digitalen Zeitalter – Codicology and Palaeography in the Digital Age 4* (Norderstedt: Book on Demand; Schriften des Instituts für Dokumentologie und Editorik 11), 179–252, <http://kups.ub.uni-koeln.de/7787/> (accessed 19 November 2019).

Kestemont, Mike, Vincent Christlein, and Dominique Stutzmann (2017), 'Artificial Paleography: Computational Approaches to Identifying Script Types in Medieval Manuscripts', *Speculum,* 92.S1: S86–S109. doi: <10.1086/694112>.

'Latin Script' (2018), in *Wikipedia*. <https://en.wikipedia.org/w/index.php?title=Latin_script&oldid=858277087. 8.2.2019> (accessed 19 November 2019).

Maadeed, Somaya Al, Wael Ayouby, Abdelâali Hassaïne, and Jihad Mohamad Aljaam (2012), 'QUWI: An Arabic and English Handwriting Dataset for Offline Writer Identification', in *2012 International Conference on Frontiers in Handwriting Recognition (ICFHR)* (Institute of Electrical and Electronics Engineers: IEEE Xplore Digital Library), 746–751. doi: <10.1109/ICFHR.2012.256>.

Mallon, Jean (1952), *Paléographie Romaine* (Madrid: Consejo superior de investigaciones cientificas; Scripturae. Monumenta et Studia, 3).

Muzerelle, Denis (2013), 'Jeux d'angles et jeux de plume. I. Retour sur l'hypothèse du biseautage de la plume', *Gazette du livre médiéval,* 60.1: 1–27. doi: <10.3406/galim.2013.2032>.

Nievergelt, Andreas, Rudolf Gamper, Marina Bernasconi Reusser, Birgit Ebersperger, and Ernst Tremp (eds) (2015), *Scriptorium: Wesen, Funktion, Eigenheiten, Comité International de Paléographie Latine, XVIII. Kolloquium, St. Gallen, 11.–14. September 2013* (München: C.H. Beck; Veröffentlichungen der Kommission für die Herausgabe der mittelalterlichen Bibliothekskataloge Deutschlands und der Schweiz).

Oeser, Wolfgang (1992), 'Beobachtungen zur Entstehung und Verbreitung schlaufenloser Bastarden. Eine Studie zur Geschichte der Buchschrift im ausgehenden Mittelalter', *Archiv für Diplomatik, Schriftgeschichte, Siegel- und Wappenkunde,* 38: 235–343.

Parkes, Malcolm B. (1969), *English Cursive Book Hands 1250–1500* (Oxford: Clarendon Press; Oxford Palaeographical Handbooks).

—— (2008), *Their Hands Before Our Eyes: A Closer Look at Scribes. The Lyell Lectures Delivered in the University of Oxford 1999* (Aldershot: Ashgate).

Simistira, Fotini, Manuel Bouillon, Mathias Seuret, Marcel Würsch, Michele Alberti, Rolf Ingold, and Marcus Liwicki (2017), 'ICDAR2017 Competition on Layout Analysis for Challenging Medieval Manuscripts', in *14ᵗʰ IAPR International Conference on Document Analysis and Recognition (ICDAR)*, 1361–70. doi: <10.1109/ICDAR.2017.223>.

Smith, Marc H. (2004), 'Review of Albert Derolez (2003), *The Palaeography of Gothic Manuscript Books: From the Twelfth to the Early Sixteenth Century* (Cambridge: Cambridge University Press)', *Scriptorium*, 58: 274–279.

——— (2016), 'Learning to Look at Texts', in Marc H. Smith and Laura Light (eds), *Primer 9*: Script (New York: Les Enluminures), 2–7.

Stutzmann, Dominique (2009), *Écrire à Fontenay: Esprit cistercien et pratiques de l'écrit en Bourgogne (XIIe-XIIIe siècles)* (Thèse de doctorat Université Paris 1 Panthéon-Sorbonne).

Stutzmann, Dominique (2012), 'Paléographie latine et vernaculaire (livres et documents)', *Annuaire de l'École pratique des hautes études (EPHE), Section des sciences historiques et philologique, Résumés des conférences et travaux*, 143: 152–57.

——— (2014), 'Conjuguer diplomatique, paléographie et édition électronique: Les mutations du XIIe siècle et la datation des écritures par le profil scribal collectif', in Antonella Ambrosio, Sébastien Barret, and Georg Vogeler, *Digital Diplomatics: The Computer as a Tool for the Diplomatist* (Cologne: Böhlau; Archiv für Diplomatik, Beiheft ,14), 271–290.

——— (2015), 'Clustering of Medieval Scripts through Computer Image Analysis: Towards an Evaluation Protocol', *Digital Medievalist*, 10.<https://journal.digitalmedievalist.org/articles/10.16995/dm.61/> (accessed 19 November 2019).

——— (2016), *Competition on the Classification of Medieval Handwritings in Latin Script* (Paris: Institut de Recherche et d'Histoire des Textes). <http://clamm.irht.cnrs.fr/>. (accessed 19 November 2019)

——— (2017), 'Les "manuscrits datés", base de données sur l'écriture', in Teresa De Robertis and Nicoletta Giovè Marchioli (eds), *Catalogazione, storia della scrittura, storia del libro: I Manoscritti datati d'Italia vent'anni dopo* (Florence: SISMEL – Edizioni del Galluzzo), 155–207.

——— (2018), 'Variability as a Key Factor For Understanding Medieval Scripts: The ORIFLAMMS Project (ANR-12-CORP-0010)', in Stewart J. Brookes, Malte Rehbein, and Peter A. Stokes (eds), *Digital Palaeography* (London: Routledge, Taylor & Francis Group; Digital Research in the Arts and Humanities). <https://halshs.archives-ouvertes.fr/halshs-01778620> (accessed 19 November 2019).

——— (2019), 'Résistance au changement? Les écritures des livres d'heures dans l'espace français (1290–1550)', in Eef Overgaauw and Martin J. Schubert, *Change in Medieval and Renaissance Scripts and Manuscript: Proceedings of the 19ᵗʰ Colloquium of the Comité international de paléographie latine (Berlin, September 16-18, 2015)* (Turnhout: Brepols; Bibliologia, 50), 97–116.

Tensmeyer, Chris, Daniel Saunder, and Tony Martinez (2017), 'Convolutional Neural Networks for Font Classification', in *14ᵗʰ IAPR International Conference on Document Analysis and Recognition (ICDAR)*, 985–990. doi: <10.1109/ICDAR.2017.164>.

——— (2017), *Submission for CLaMM Competition as Part of ICDAR 2017: Ctensmeyer/Clamm_2017*, Python. <https://github.com/ctensmeyer/clamm_2017>.(accessed 19 November 2019).

Written Artefacts as Cultural Heritage

Ed. by Michael Friedrich and Doreen Schröter

Written Artefacts as Cultural Heritage was established in 2020. The series is dedicated to the double role of written artefacts as representations and generators of humankind's cultural heritage. Its thematic scope embraces aspects of preservation, the identity-defining role of artefacts as well as ethical questions.

The mix of practical guides, colloquium papers and project reports is specifically intended for staff at libraries and archives, curators at museums and art galleries, and scholars working in the fields of manuscript cultures and heritage studies.

Every volume of *Written Artefacts as Cultural Heritage* has been peer-reviewed and is openly accesible. There is an online and a printed version..

If you wish to receive a copy or to present your research, please contact the editorial office:
https://www.csmc.uni-hamburg.de/publications/cultural-heritage.html

manuscript cultures (mc)

Editors: Michael Friedrich and Jörg B. Quenzer
Editorial office: Irina Wandrey

CSMC's academic journal was established as newsletter of the research unit 'Manuscript Cultures in Asia and Africa' in 2008 and transformed into a scholarly journal with the appearance of volume 4 in 2011. *manuscript cultures* publishes exhibition catalogues and articles contributing to the study of written artefacts. This field of study embraces disciplines such as art history, codicology, epigraphy, history, material analysis, palaeography and philology, informatics and multispectral imaging.

manuscript cultures encourages comparative approaches, without regional, linguistic, temporal or other limitations on the objects studied; it contributes to a larger historical and systematic survey of the role of written artefacts in ancient and modern cultures, and in so doing provides a new foundation for ongoing discussions in cultural studies.

Every volume of *manuscript cultures* has been peer-reviewed and is openly accessible:

https://www.csmc.uni-hamburg.de/publications/mc.html

If you wish to receive a copy or to present your research in our journal, please contact the editorial office:

irina.wandrey@uni-hamburg.de

manuscript cultures 14

manuscript cultures 13

manuscript cultures 11

manuscript cultures 10

manuscript cultures 9

manuscript cultures 8
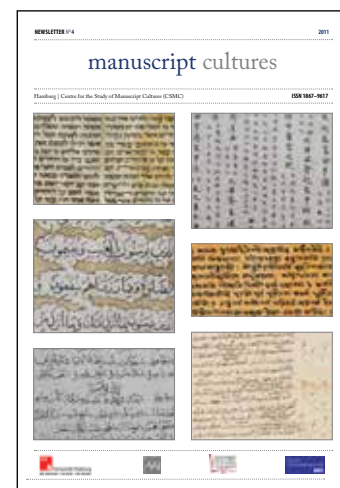

manuscript cultures 7

manuscript cultures 6

manuscript cultures 5

manuscript cultures 4

Studies in Manuscript Cultures (SMC)

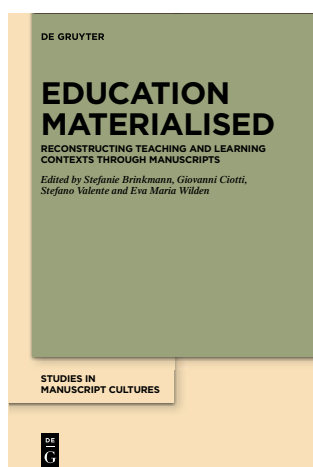Ed. by Michael Friedrich, Harunaga Isaacson, and Jörg B. Quenzer

From volume 4 onwards all volumes are available as open access books on the De Gruyter website:
https://www.degruyter.com/view/serial/43546
https://www.csmc.uni-hamburg.de/

DE GRUYTER

Forthcoming

DE GRUYTER

EDUCATION MATERIALISED

RECONSTRUCTING TEACHING AND LEARNING
CONTEXTS THROUGH MANUSCRIPTS

Edited by Stefanie Brinkmann, Giovanni Ciotti,
Stefano Valente and Eva Maria Wilden
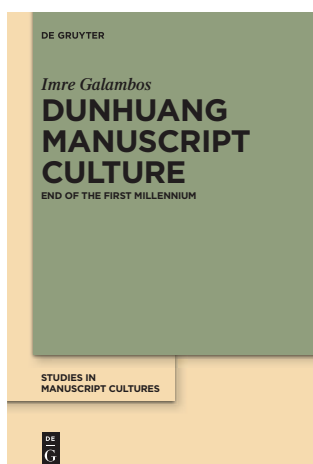
**STUDIES IN
MANUSCRIPT CULTURES**

23 – *Education Materialised: Reconstructing Teaching and Learning Contexts through Manuscripts,* edited by Stefanie Brinkmann, Giovanni Ciotti, Stefano Valente and Eva Maria Wilden

Manuscripts have played a crucial role in the educational practices of virtually all cultures that have a history of using them. As learning and teaching tools, manuscripts become primary witnesses for reconstructing and studying didactic and research activities and methodologies from elementary levels to the most advanced.

The present volume investigates the relation between manuscripts and educational practices focusing on four particular research topics: educational settings: teachers, students and their manuscripts; organising knowledge: syllabi; exegetical practices: annotations; modifying tradition: adaptations.

The volume offers a number of case studies stretching across geophysical boundaries from Western Europe to South-East Asia, with a time span ranging from the second millennium BCE to the twentieth century CE.

New release

DE GRUYTER

Imre Galambos

DUNHUANG MANUSCRIPT CULTURE

END OF THE FIRST MILLENNIUM

**STUDIES IN
MANUSCRIPT CULTURES**

22 – *Dunhuang Manuscript Culture: End of the First Millennium,* by Imre Galambos

Dunhuang Manuscript Culture explores the world of Chinese manuscripts from ninth–tenth century Dunhuang, an oasis city along the network of pre-modern routes known today collectively as the Silk Roads. The manuscripts have been discovered in 1900 in a sealed-off side-chamber of a Buddhist cave temple, where they had lain undisturbed for for almost nine hundred years. The discovery comprised tens of thousands of texts, written in over twenty different languages and scripts, including Chinese, Tibetan, Old Uighur, Khotanese, Sogdian and Sanskrit. This study centres around four groups of manuscripts from the mid-ninth to the late tenth centuries, a period when the region was an independent kingdom ruled by local families. The central argument is that the manuscripts attest to the unique cultural diversity of the region during this period, exhibiting – alongside obvious Chinese elements – the heavy influence of Central Asian cultures. As a result, it was much less 'Chinese' than commonly portrayed in modern scholarship. The book makes a contribution to the study of cultural and linguistic interaction along the Silk Roads.

Studies in Manuscript Cultures (SMC)

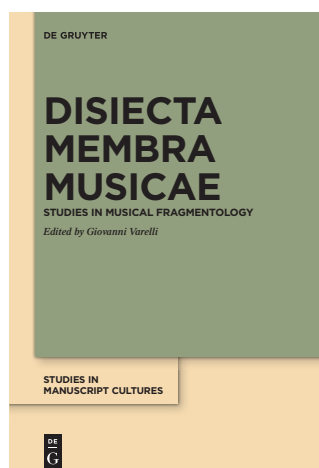Ed. by Michael Friedrich, Harunaga Isaacson, and Jörg B. Quenzer

DE GRUYTER

New release

21 – *Disiecta Membra Musicae: Studies in Musical Fragmentology*, edited by Giovanni Varelli

Although fragments from music manuscripts have occupied a place of considerable importance since the very early days of modern musicology, a collective, up-to-date, and comprehensive discussion of the various techniques and approaches for their study was lacking. On-line resources have also become increasingly crucial for the identification, study, and textual/musical reconstruction of fragmentary sources. *Disiecta Membra Musicae. Studies in Musical Fragmentology* aims at reviewing the state of the art in the study of medieval music fragments in Europe, the variety of methodologies for studying the repertory and its transmission, musical palaeography, codicology, liturgy, historical and cultural contexts, etc. This collection of essays provides an opportunity to reflect also on broader issues, such as the role of fragments in last century's musicology, how fragmentary material shaped our conception of the written transmission of early European music, and how new fragments are being discovered in the digital age. Known fragments and new technology, new discoveries and traditional methodology alternate in this collection of essays, whose topics range from plainchant to *ars nova* and fifteenth- to sixteenth-century polyphony.

New release
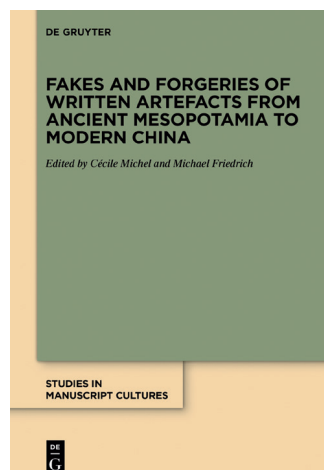
20 – *Fakes and Forgeries of Written Artefacts from Ancient Mesopotamia to Modern China,* edited by Cécile Michel and Michael Friedrich

Fakes and forgeries are objects of fascination. This volume contains a series of thirteen articles devoted to fakes and forgeries of written artefacts from the beginnings of writing in Mesopotamia to modern China. The studies emphasise the subtle distinctions conveyed by an established vocabulary relating to the reproduction of ancient artefacts and production of artefacts claiming to be ancient: from copies, replicas and imitations to fakes and forgeries. Fakes are often a response to a demand from the public or scholarly milieu, or even both. The motives behind their production may be economic, political, religious or personal – aspiring to fame or simply playing a joke. Fakes may be revealed by combining the study of their contents, codicological, epigraphic and palaeographic analyses, and scientific investigations. However, certain famous unsolved cases still continue to defy technology today, no matter how advanced it is. Nowadays, one can find fakes in museums and private collections alike; they abound on the antique market, mixed with real artefacts that have often been looted. The scientific community's attitude to such objects calls for ethical reflection.