An Interpretable Deep Learning Approach for Morphological Script Type Analysis

Malamatenia Vlachou-Efstathiou[1,2][0000−0002−9397−356], Ioannis Siglidis[2][0009−0002−2278−5825], Dominique Stutzmann[1][0000−0003−3705−5825], and Mathieu Aubry[2][0000−0002−3804−0193]

[1] Institut de Recherche et d'Histoire des Textes, Paris, Île-de-France, France
{malamatenia.vlachou, dominique.stutzmann}@irht.cnrs.fr
[2] LIGM, Ecole des Ponts, Univ Gustave Eiffel, CNRS, Marne-la-Vallée, France
{mathieu.aubry, ioannis.siglidis}@enpc.fr
https://learnable-handwriter.github.io/

Abstract. Defining script types and establishing classification criteria for medieval handwriting is a central aspect of palaeographical analysis. However, existing typologies often encounter methodological challenges, such as descriptive limitations and subjective criteria. We propose an interpretable deep learning-based approach to morphological script type analysis, which enables systematic and objective analysis and contributes to bridging the gap between qualitative observations and quantitative measurements. More precisely, we adapt a deep instance segmentation method to learn comparable character prototypes, representative of letter morphology, and provide qualitative and quantitative tools for their comparison and analysis. We demonstrate our approach by applying it to the *Textualis Formata* script type and its two subtypes formalized by A. Derolez: Northern and Southern *Textualis*.

Keywords: Latin Palaeography · Computer Vision · Palaeographical Analysis · Character Prototypes · Textualis Formata

1 Introduction

The concept of *script type* is of central importance to palaeography, which studies handwritten documents in relation to their context of production such as date, origin, and scribal hands, to support historical discourse. Adapting M. Parkes, we define a *script type* as "the model which the scribe has in mind's eye when they write" [43,56], that is, if we restrict ourselves to characters, the set of prototypical forms of each character towards which they are working when they write. In order to discretize the *continuum* of handwritten forms and establish script types and their classification criteria, the palaeographical method compares handwriting samples, describes, and analyzes the variations of letter forms. Palaeographers often resort to the idea of script types as *ideal prototypes* [6], with the delineation of artificial alphabets, i.e., sets of abstracted letter forms. Several typologies have been proposed and refined over the years. Most recently A. Derolez proposed a

taxonomy for Gothic book scripts based on letter morphology [16], where some letters with distinctive visual elements serve as the basis for classification.

In this paper, we introduce a methodology that leverages deep learning for the analysis of morphological script types. More precisely, we learn aligned character prototypes from documents and present different methods for qualitative and quantitative analysis. This enables us to confront different documents to existing typologies, potentially adding nuance or complementing them. Indeed, existing typologies present persistent methodological issues [16,53,58], mainly the ambiguity arising from relying on a "global impression" to discern scripts, inconsistencies in nomenclature across scholarly traditions, and difficulties in describing minute morphological differences using natural language [19,56]. These challenges underscore the potential benefits of methods such as ours, which could enhance palaeographical analysis through a systematic and objective approach, and facilitate the integration of quantitative measures with qualitative observations.

This is in line with the position of pioneers like Léon Gilissen [20,21] and others [46,42,52,66,15,37,59,39], who experimented with statistical measurements and the modeling of *measurable elements* of script. Such measurements of scripts pose significant challenges, such as defining a set of descriptors or discriminative handwriting features and ensuring comparable objects and magnitudes [55]. Contrary to classification tasks such as writer and geographical attribution, which are formulated as discriminative learning problems, script type analysis cannot be reduced to a classification problem [60,26,61] and adequate modeling of variations is crucial [57]. Indeed simply matching external samples to pre-defined script types does not help better understanding and questioning the classification criteria.

We thus propose a method for evidence-based paleography focusing on interpretability rather than script classification. Our key idea is to remain close to classical morphological approaches for defining taxonomies and introduce tools to model and analyze letter shapes automatically. We build on the Learnable Typewriter approach [51] and adapt it so that it can learn comparable character prototypes, which requires designing appropriate finetuning strategy and filtering. We then introduce visualizations and graphical tools, as well as an interpretable variability measure. To demonstrate how such tools can be leveraged for palaeographic analysis, we select a corpus in *Textualis Formata* and present a case study on the morphological analysis of its two subtypes, Northern and Southern *Textualis*.

Contribution. In summary, our main contributions are:

- the adaptation of a deep instance segmentation method for palaeographical script type analysis
- a methodology for homologous comparison of characters, including visualization, graphical, and quantitative tools
- a case study demonstrating how these tools can complement the classic taxonomy of A. Derolez [16] for the analysis of Northern and Southern *Textualis*.

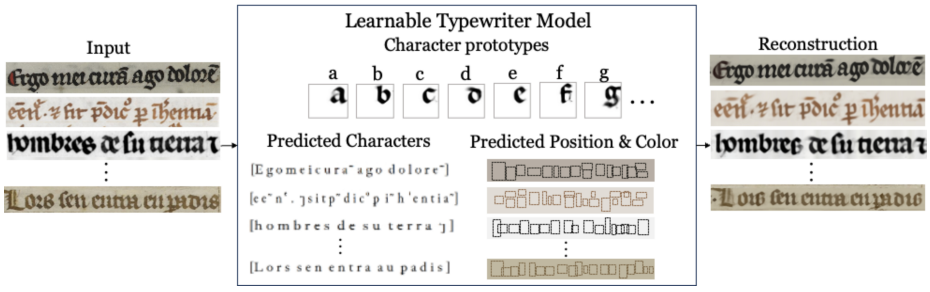


Fig. 1: **The Learnable Typewriter Model** learns to reconstruct text lines using a set of learned character prototypes. We demonstrate how the character prototypes can be used for palaeographic analysis.

2 Related work

We first give an overview of works that develop quantitative methods for palaeographic analysis. We then present "prototype-based" approaches to document analysis, which, although not specifically developed for paleographic analysis, are the basis of our approach.

Quantitative methods in palaeography. In the past two decades, many automatic methods have been developed for writer or script classification [54,40], using texture-based features [38,49,67,34,27,17,25], grapheme-based features [48,50,18,29] and deep learning classification approaches [9,7,10,11,65,64,28]. Some papers, such as [31], make a particular effort to build interpretable features or to visualize deep features responsible for the classification, but their interpretation remains limited.

Another branch of studies aims at producing interpretable outputs for palaeographic analysis of letter structure. The Information System for Graphological Identification [36], extracts the average shape of specific characters via curve and contour detection, standardizing orientation and size, for automatic hand comparison and writer identification. The Graphem project [39] focused specifically on script type features. [14] explores visually interpretable stroke analysis, by extracting connected components to create a strokes code book, and then grouping the strokes through graph coloring for categorization of elementary stroke shapes. Closest to us, focusing on entire letter form variations for script type analysis is the System of Palaeographical Inspection [1,8]. It generates an average character prototype by computing the centroid of semi-automatically segmented occurrences. The prototypes are used both for hierarchical clustering of similar hands and classification of external samples.

However, the results of these approaches can hardly be compared with minute traditional palaeographic analysis and do not provide complete automation. Instead, our idea is to build on methods that directly learn prototypical characters from documents and use them for actual paleographic analysis.

Prototype-based approaches in document analysis. Early methods for document analysis [30,32,68,3,4] use variants of character template matching for analyzing documents. While their main goal is often to perform optical character recognition (OCR), such methods typically also produce finer outputs, such as character segmentation, and learn a character template or prototype. Similar approaches have thus been used for typographical analysis of early prints [47,24,33]. This type of approach has recently been revisited with deep learning tools by the Learnable Typewriter approach [51]. We build on this method and describe it in the next section.

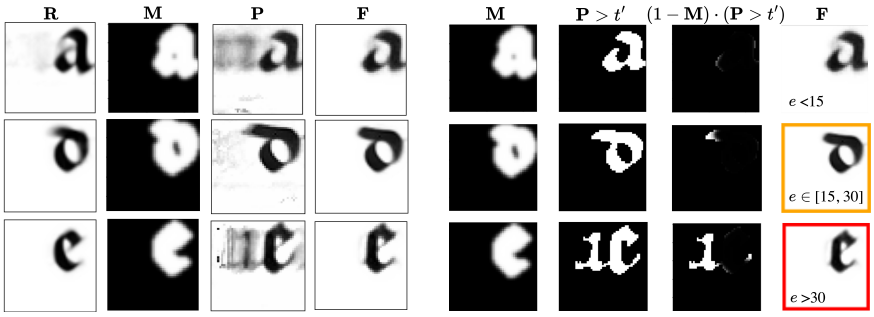
3 Approach

3.1 Learning comparable character prototypes

The Learnable Typewriter. We build our approach on the Learnable Typewriter model [51], visualized in Figure 1. This deep learning model learns to reconstruct text lines by compositing a set of *character prototypes* on a simple background. Given as input the image of a line, it predicts the color of the background, the characters used in the line, and for each character, its position and color. The character prototypes are also learned by the model, and each instance of a character is reconstructed with the exact same prototype. The model can be trained, as in our experiments, using a set of text line images with their transcriptions.

Each character prototype is a grayscale image and can be thought of as the average shape of all occurrences of a character in the training data, standardized for color, size, and position. Therefore, training a Learnable Typewriter model on a particular corpus, such as one corresponding to a specific script or handwriting style, will yield the average shape of each character without the need for manual selection of specific character samples, annotation of character positions, or binarization.

Finetuning character prototypes. We propose to compare different documents and different scripts by comparing the character prototypes learned on various corpora. However, directly comparing prototypes learned by different models is not possible, because they are not aligned. Our solution is first to learn a *reference model* using a reference corpus - in our case study, a set of documents in *Textualis Formata* - then finetune the model to reconstruct selected documents, keeping all network parameters frozen except those that only impact the prototypes. Since the positioning, scaling, and coloring of the prototypes are shared, the prototypes will remain aligned and can be directly compared, such as by computing their difference. We define a single reference corpus to obtain reference prototypes, and then finetune them on multiple specific corpora, that may or may not be part of the reference corpus.



(a) **Filtering strategy** (b) **Failure case identification**

Fig. 2: **Prototype filtering and failure case identification.** We use a mask M defined from the reference prototype **R** to remove artifacts from finetuned prototypes **P**, yieling a filtered prototype **F**. We compute an error e associated to the filtering to automatically identify potential failure cases.

Character prototype filtering. While the reference prototypes generally are of high quality, we observed various artifacts in the finetuned prototypes, particularly for less common characters, when trained on a single document. While this does not hinder the qualitative analysis of the prototypes' shapes, it does complicate the quantitative comparison between prototypes. To alleviate this issue, we propose to filter the finetuned prototypes using the reference ones, as visualized in Figure 2.

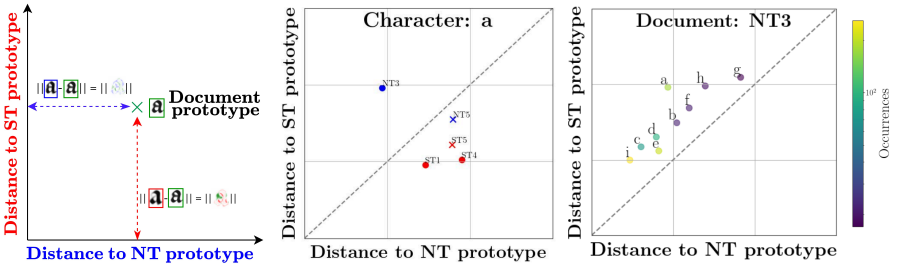
Let us consider a specific character, the associated reference prototype **R** and the associated finetuned prototype **P** for a given finetuning. Using the reference prototype, we define a reference mask **M** as

$$\mathbf{M} = \mathbf{G} * D(\mathbf{R} > t), \tag{1}$$

where **G** is a Gaussian filter, $*$ denotes a convolution, D is a dilation operation and $\mathbf{R} > t$ is the binary mask associated to pixels for which **R** is greater than a threshold t. In our experiments, we use a Gaussian **G** of standard deviation 2, a dilation D of 2 pixels, and a threshold t =0.8. Intuitively, this mask defines in a soft way, for each character, pixels that are close to the reference prototype.

Using this mask, we define a filtered prototype $\mathbf{F} = \mathbf{M} \cdot \mathbf{P}$, where $\cdot$ is the pixel-wise multiplication, which we use for all of our analyses.

Automatic identification of failure cases. While the filtering process described above generally improves the visual quality of the prototypes without changing the appearance of the characters themselves, there are instances where either the appearance is slightly altered or the finetuned prototype is of very low quality. We want to identify such cases automatically, to avoid misinterpretations. To do



(a) Graph interpretation (b) Character graph (c) Document graph

Fig. 3: **Comparison graphs.** The markers correspond to different document prototypes and their coordinates to their distance to the Northern and Southern *Textualis* prototypes. See text for details.

so, for a given character associated with a reference mask $\mathbf{M}$ and a finetuned prototype $\mathbf{P}$ we compute the error e defined by:

$$e = \|(1 - \mathbf{M}) \cdot (\mathbf{P} > t')\|, \tag{2}$$

where $\| \ \|$ is the norm of an image, $\cdot$ is the pixel-wise multiplication, and t' is a scalar threshold, set to 0.65 in our experiments. Intuitively, this error can be interpreted as the number of pixels that are present in the finetuned prototype $\mathbf{P}$ (i.e., have values higher than t') but are filtered out by the mask M. This value e enables us to identify: (i) finetuned prototypes whose shape is significantly different from the reference one and are thus modified by the filtering process, such as the ‹d› in Figure 2, and (ii) finetuned prototypes of low quality that might not be easily interpretable, such as the ‹e› in Figure 2. In our results, we highlight prototypes where $e > 15$ in orange and prototypes where $e > 30$ in red.

3.2 Character prototype comparison for palaeographic analysis

Visual comparison. We can visually highlight the morphological differences between two prototypes by subtracting one from the other. To make this difference easier to understand, we use a colormap that represents zeros as white, and positive and negative values as two distinct colors, typically red and blue. This method reveals pixel-wise differences, facilitating an initial qualitative examination of the morphological disparities (see Table 2 and Figure 7).

Character and document comparison graphs. To quantitatively analyze character prototypes, we introduce an adapted comparison graph, illustrated in Figure 3. In this graph, each point represents a specific document character prototype, with its coordinates defined as its distance in pixel space to two selected prototypes. Since we study *Textualis Formata*, we use the prototypes of Northern

and Southern *Textualis* (NT and ST) (see Section 4.1 for details), other prototypes could be selected for different analysis. The distance to the axes can be interpreted as visualized in Figure 3a.

We employ two complementary types of graphs for our analysis. Firstly, *character graphs* (see Figure 3b) concentrate on a single character across all documents. In these graphs, dots represent the documents selected to train the reference, Northern and Southern *Textualis* models, while crosses denote the remaining documents. Blue, resp. red, markers signify Northern, resp. Southern *Textualis* documents. The identifier for each document is written near its marker. This type of graph allows to easily identify outlier documents for a specific character, such as NT5 for the ‹a› character, which is closer to the ST prototype. Secondly, *document graphs* (see Figure 3c) focus on a specific document, where each dot corresponds to a different character, labeled near the marker. The color of the dots corresponds to the frequency of occurrence, with darker dots representing less frequent characters. This type of graph facilitates the identification, for a given document, of the characters most typical of a subtype, such as the ‹a› for NT3.

Quantifying character variability. According to the literature, the Northern *Textualis* class allows for more morphological variation across documents than the Southern *Textualis*. While the visualizations and graphs can qualitatively support this idea, we further aim to quantify the characters' variability. Thus, we report the standard deviations of character prototypes within one subtype, σ_{NT} for Northern *Textualis* and σ_{ST} for Southern *Textualis*. These standard deviations can be thought of as the average number of pixels that change across two character prototypes of the same subtype.

4 Experiments

Research question and analysis framework. To analyze the results of our approach, we adopt the taxonomy formalized by A. Derolez [16], which provides a framework based on morphological criteria. We select a corpus in *Textualis* script type, specifically in its canonized calligraphic form *Formata*, due to its more distinguishable morphological elements compared to more rapidly executed forms. Despite the morphology-based categorization where *Textualis Formata* represents a coherent group, Derolez makes "an important distinction between two fundamentally different species", Northern and Southern *Textualis*, based on their geographical distribution and a set of minute morphological differences. However, these differences vary according to factors such as date, geographical origin, or language, and often intersect, blurring this fundamental distinction in some cases. Our goal is to confront the results of our approach with Derolez's criteria and observations.

Table 1: **Dataset Description.** The 'Doc.' column refers to the names we use for the different documents in this paper, NT and ST stand for Northern and Southern *Textualis*, and the 'Ref.' column reports which documents are used to train the reference and subtype models.

Doc.	Ref.	Shelfmark	Language	Century	Date	Origin	Folio(s)	Lines
NT1	✓	Paris, BnF, Français 403 [62]	French	13^{th}	1226-1250	England	4r	76
NT2	✓	Paris, BnF, Français 12400 [62]	French	14^{th}	1305-1310	Eastern France	92r	54
NT3	✓	Arras, BM Ms. 861 (315) [12]	Latin	14^{th}	-	-	56r	65
NT4	✓	Paris, BnF, Français 1728 [44]	French	14^{th}	1372	-	3r	59
NT5		Paris, BnF, Français 20120 [62]	French	13^{th}	1240-1250	Paris or Orleans	7r	81
NT6		Paris, BnF, Français 619 [35]	French	14^{th}	1375-1400	-	1v	65
NT7		Berlin, SB, Hdschr. 25 [12]	Latin	15^{th}	1451-1500	Flanders	22r-22v	25
ST1	✓	Paris, BnF, Français 9082 [62]	French	13^{th}	1295	Rome	171r	56
ST2	✓	Paris, BnF, Espagnol 65 [5]	Navarrese	14^{th}	1301-1310	-	5v,6r	118
ST3	✓	Paris, BnF, Italien 590 [2]	Italian	14^{th}	1370-1410	Italy	18r	68
ST4	✓	Madrid, FLG, mss 289 (Hand A) [22]	Castilian	15^{th}	1480	Seville	245v	86
ST5		Paris, BnF, Français 187 [62]	French	14^{th}	1350-1386	Milan or Genova	18r	16
ST6		Paris, BnF, Latin 7720 [23]	Latin	15^{th}	1390-1410	Florence	102v	74
ST7		Madrid, FLG, mss 289 (Hand B) [22]	Castilian	15^{th}	1480	Seville	274v	52
Total:								**892**

4.1 Dataset and Experiment Details

Data Selection. Our data was build from two open-access repositories, ECMEN [63,62] and CATMuS [45]. We selected the documents to ensure the variability of the corpus in terms of geographic, linguistic, and chronological distribution, resulting in seven documents for each subscript as listed in Table 1 (labelled NT 1-7 and ST 1-7). We verified and normalized the transcriptions to fit a graphemic approach [13,12]. Despite our efforts to use diverse and representative documents, we acknowledge that biases may exist within our dataset.

Character set choice. From the extended set of characters in medieval manuscripts - including upper and lower case letters, ligatures, punctuation, and abbreviation signs - we follow a standard approach for morphological analysis and focus on the lowercase alphabetic characters where morphology is crystalized through frequent usage. From these characters, we show results on the ones common to all documents, thus excluding ‹j,k,v,x,y,z›.

Trained models. For our analysis, we trained multiple models to obtain character prototypes at different levels of granularity: (i) a script type model for *Textualis*, (ii) script subtype models for Northern and Southern *Textualis*, and (iii) document level models for each document in our dataset. We use the *Textualis* script type model as reference model, and finetune all other models from it as explained in Section 3.1. to validate that our reference and subtype models can be effective to analyze documents they were not trained on, we limited their training to NT 1-4 and ST 1-4, indicated in the 'Ref' column in Table 1.

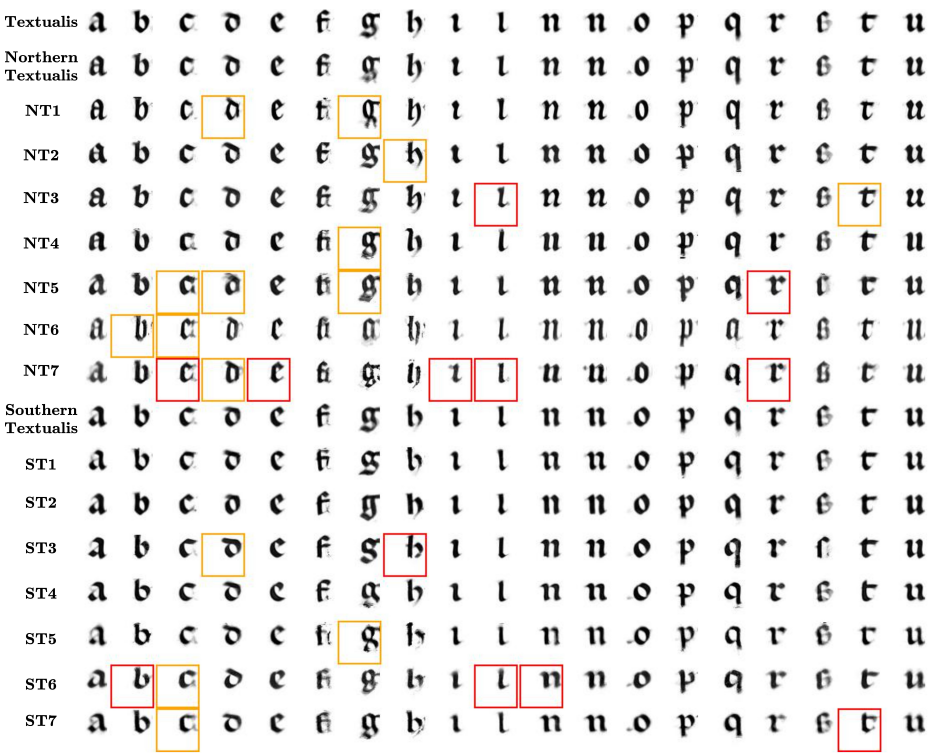
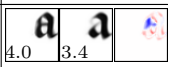
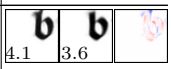
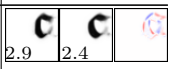
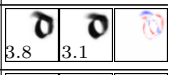
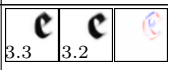
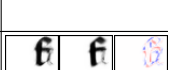
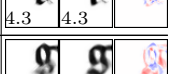
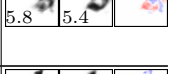
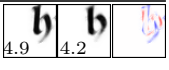
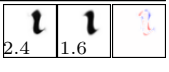
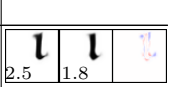
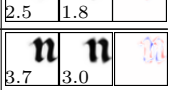
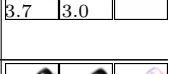
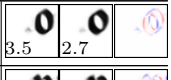
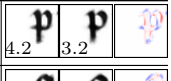


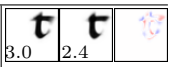
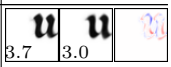
Fig. 4: **Our filtered prototypes** on type, sub-type and document level. The highlighted prototypes are the ones for which filtering had a significant impact (see Section 3.1 and 4.1 for details).

4.2 General results

Prototype quality. Figure 4 shows the character prototypes generated by our approach across type, sub-type, and document levels. We note several limits in these prototypes. Firstly, the prototype for ‹m› is not well modeled (𝔫). This issue stems from the network being trained with a CTC loss allowing to use the same prototype twice to model the same letter. As a result, we exclude ‹m› from our analysis. Secondly, the two allographs ‹ſ/s› are represented by a single prototype resulting in an averaged representation of the two (𝔰), necessitating cautious examination. Thirdly, we meticulously scrutinized all prototypes highlighted in orange and red, where filtering significantly impacted the outcomes. In almost all instances, the filtering was meaningful and eliminated irrelevant artifacts from the prototypes. Nevertheless, it is worth mentioning that: (i) the lengthy shaft of the ST3 ‹d›(𝔡) and the hairline extension of the limb of ‹h› in NT2 (𝔥) are slightly severed; (ii) in general, documents that are not part of the reference set, are less accurately modeled and more impacted by the filtering,

Table 2: Comparison between Derolez' criteria for Northern and Southern *Textualis* and our subtype prototypes.

‹Ch.›	Derolez'criteria	NT σ_{NT} | ST σ_{ST} | diff.
‹a›	**NT**: Closed form with variations like "box-‹a›" **ST**: Open form or slightly closed with hairline	 4.0 3.4
‹b›	**NT**: Sloped or forked ascender tops **ST**: (i) Flat ascender tops, (ii) round lobe	 4.1 3.6
‹c›	**NT**: Angular or broken lobe curves **ST**: Semi-circular lobe	 2.9 2.4
‹d›	**NT**: (i) Lengthened and (ii) concave shaft **ST**: (i) Shorter shaft and (ii) almost horizontal, (iii) round bowl	 3.8 3.1
‹e›	**NT**: (i) Diagonal direction of the hairline and (ii) angular or broken lobe curves **ST**: (i) Horizontal or no hairline, (ii) semi-circular lobe form	 3.3 3.2
‹f›	**NT**: Incurvation of the shaft foot to the right **ST**: Flat foot	 4.3 4.3
‹g›	**NT:** Tendency for the closed, "8-shaped" form **ST**: Tendency for the open, "Rücken -g" form **Note**: Various intermediate forms and difficult to classify	 5.8 5.4
‹h›	**NT**: (i) Incurvation of the shaft foot to the right, (ii) extended or dislocated limb and (iii) sloped or forked tops **ST**: (i) Flat ascender foot, (ii) circular limb on the baseline and (iii) flat ascender tops	 4.9 4.2
‹i›	**NT**: (i) Accentuated (diamond-shaped or forked) headline and (ii) extended hairline for the foot **ST**: (i) Approach stroke for the headline and (ii) flat end for the foot	 2.4 1.6
‹l›	**NT**: Sloped or forked ascender tops **ST**: Flat tops	2.5 1.8
‹n›	**NT**: Accentuated (diamond-shaped or hairlines) for the headline and (ii) same for feet **ST**: (i) Approach stroke hairline for the headline and (ii) flat feet	 3.7 3.0
‹o›	**NT**: Broken / more vertically elongated curves **ST**: Circular arc forms	 3.5 2.7
‹p›	**NT**: (i) Artificial spurs on the left and (i) decorated descender feet **ST**: No spurs and (ii) flat descender feet	 4.2 3.2
‹q›	**NT**: (i) Lengthy and (ii) decorated descenders **ST**: (i) Short and (ii) flat descenders	 4.8 4.0
‹r›	**NT**: (i) Hairline endstroke for shaft foot and (ii) angular horizontal stroke **ST**: (i) Flat shaft foot and (ii) straight horizontal stroke	 2.8 2.4
‹s›	**NT**: Incurvation of the shaft foot to the right for ‹ſ› and (ii) closed and angular curves for ‹s› **ST**: (i) Flat shaft foot for ‹ſ› and (ii) open semi-circular curves for ‹s›	 2.7 2.4

‹t›	**NT**: Vertical pendant hairline of the headstroke **ST**: No ornaments **Note**: Different levels of shaft projection above headline and length of horizontal stroke	 3.0 2.4
‹u›	**NT**: Accentuated (diamond-shaped or sloped) headline **ST**: Flat or left approach stroke for headline	 3.7 3.0

especially the forked ascender tops of ‹h› and ‹b›. This information loss, while limited, should be considered during the analysis.

Palaeographical relevance of the subtype prototypes. To showcase how our prototypes can be related to classical palaeographic analysis, we systematically compare in Table 2 Derolez's general morphological criteria to our Northern and Southern *Textualis* prototypes, highlighting their variations by visualizing their difference. We find that Derolez's observations closely align with the variations that our prototypes enable us to visualize. Additionally, we report our variability scores σ_{NT} and σ_{ST} for each letter, which were consistently higher for Northern *Textualis*, which is consistent with Derolez's claim that this script subtype generally exhibits higher intra-class variation.

4.3 Character graph analysis.

In this section, we provide examples of how our character graphs, together with our prototypes, can support a detailed palaeographic analysis of the variations of a specific character (examples presented in Figure 5).

Discriminative characters. We first analyze the results for four discriminative characters, ‹a,o,p,h›. The letter **‹a›** is often considered as a distinguishing criterion between script types, so much so that W. Oeser [41] distinguished seven categories within the Northern *Textualis* script subtype mainly based on allographs of ‹a›. Most striking in our **‹a›** character graph is that the prototypes for NT5 ( ) and NT7 ( ) are actually closer to the ST prototypes. This is consistent with the observation that open ‹a› forms are standard for ST. The dispersion of the characters on the graph also provides insight into the variability of ‹a› in this subtype. The group associated to NT1-4 corresponds to the closed "box-a" form in NT2 ( ) and NT4 () and the double-bow variant in NT1 () and NT3 (). NT6 presents a more vertically elongated form and stands out (). While there is morphological variations across ST documents, with round shapes (ST1 ; ST2 ; ST5 ; ST6), or with more angular inner bows (ST3 ; ST4 ; ST7), the consistent use of an open form, or only closed with a hairline, distinguishes them from the NT subtype, and all ST documents prototypes are closer to the ST prototype.

The letter **‹o›** is also particularly discriminative. The treatment of its (generally) two mirroring arcs, using broken or semi-circular strokes, often has visual echoes in letters with lobes and arcs like ‹b, c, e, p, q›, which contributes to the visual evaluation of a hand or script type as wide/round or narrow/angular.

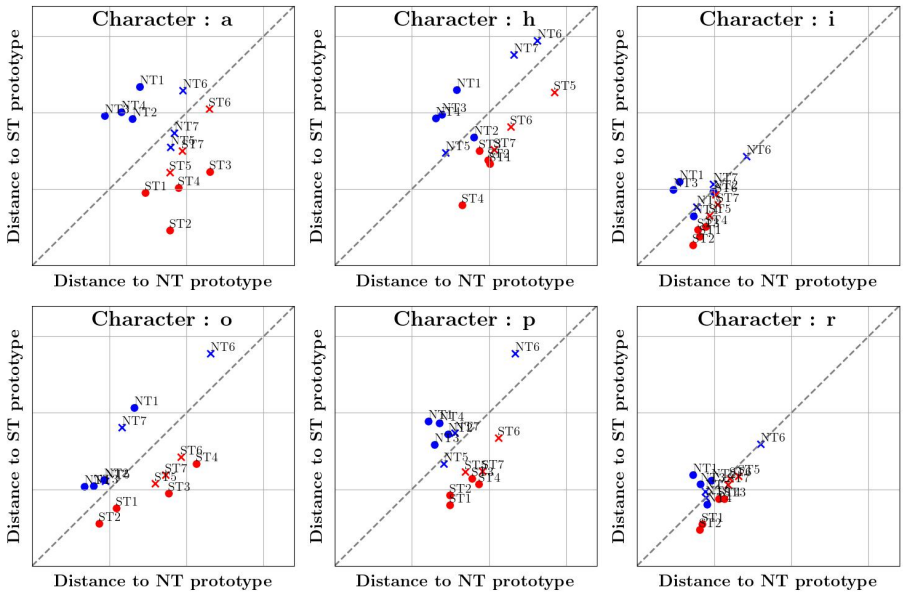


Fig. 5: **Character comparison graphs**.

This is confirmed by the fact that all the points of documents identified as NT are above the diagonal and all the ones corresponding to ST below, meaning that the prototypes for each document are closer to its subtype prototype than the other one. For NT, the forms of NT2-5 (𝒐𝒐𝒐𝒐) are particularly well reconstructed, better than NT1 (𝒐) and NT7 (𝒐) which are slightly more narrow and vertically elongated. Again, the form of NT6 stands out (𝒐), it consists of double broken strokes resulting in a narrow, quadrangle shape. Similarly for ST, ST1 (𝒐) and ST2 (𝒐) are particularly close to the ST prototype, being less wide than ST3-7 (𝒐𝒐𝒐𝒐𝒐). This is consistent with Derolez' assertion that angularity/roundness separates the two subtypes, while intra-class variation is associated with different degrees of narrowness/breadth.

Regarding the letter ‹**h**›, it highlights one of the limitations of our approach. The extended limb, characteristic of NT (cf. Table 2) is clearly present in all associated documents and prototypes. However, because its position varies in a document, the limb appears dimmed in the NT prototype (𝒉). Moreover, its position varies significantly across different documents, which can result in a greater distance between a document prototype and the NT prototype. This explains why the NT2 prototype is actually more similar to the ST prototype, while NT5 is as close to both. Note that the shift of the limb in NT2 (not curved to the left but rather extended straight down) compared to the NT prototype is so significant that our filtering partially erases it (as can be confirmed by examining the masked region, similar to Figure 2b), which was flagged by our

automatic failure identification. This emphasizes the necessity to confront our graphs with the visual appearance of the prototypes and the documents for interpretation.

For the letter **‹p›**, we note that the consistent presence of artificial spurs at the baseline level and in general of decorations are characteristic of NT (). The two subtypes appear well separated, except for the deviation of NT5 (), which we will further analyze in Section 4.4.

Non-discriminative characters. We now examine **‹i›** and **‹r›**. In our graphs, the points corresponding to these letters in all documents are close to the diagonal, i.e., they are as close to both the NT and ST prototypes. The fact that they are almost all close to the origin indicates that they present little variation. For ‹i›, NT1 and NT3 stand out as more typical of NT, and they indeed show clear diamond-shaped headlines ( ). The NT4 prototype ( ), on the other hand, is actually closer to the ST prototypes, and it does not present any headline decoration typical of NT. For both characters, the NT6 prototypes are much further than the rest from both the NT and ST prototypes and correspond to much narrower and elongated forms ( ).

4.4 Document graph analysis.

In this section, we discuss and interpret document graphs for the examples presented in Figure 6. Additionally, we visualize the differences between the document prototypes and subtype prototypes in Figure 7, leveraging them to better understand the graphs.

Class-representative cases. We start by examining two documents that are very typical of their subtype. For **NT3**, a 14$^{\text{th}}$ century manuscript in Latin, the graph clearly shows that all the character prototypes are closer to the NT prototypes than the ST prototypes. More in details, NT3 presents a closed, double bow ‹a› ( ), which progressively dominated over other variants from the end of the 13$^{\text{th}}$ c. [16], ascenders are consistently sloped on the left side (, hairlines directed to the right at the baselines (), and there are clear diamond-shaped headlines ( ). All characters thus fully correspond to the expected NT forms. For **ST2**, copied in the first decade of 14$^{\text{th}}$ c. in the Iberian peninsula, the character prototypes are, on the contrary, all closer and conforming to the ST prototypes: compact letters with very short flat-top ascenders ( ), flat-feet descenders ( ) and strokes so bold, hairlines almost become invisible ( ).

Ambiguous cases. A particular interest of our document graph method is the identification and analysis of documents that partially diverge from their assigned subtype. **NT5** stands out in the graphs, as seven character prototypes are closer to ST than to NT prototypes, with a particular difference for ‹a,g,q› ( ). Copied in the 13$^{\text{th}}$ c., in the Paris/Orleans area, NT5's forms are significantly smaller in size, an example of "pearl-script", generally used for Parisian pocket-Bibles. Even though its level of execution is still *Formata*, due to their

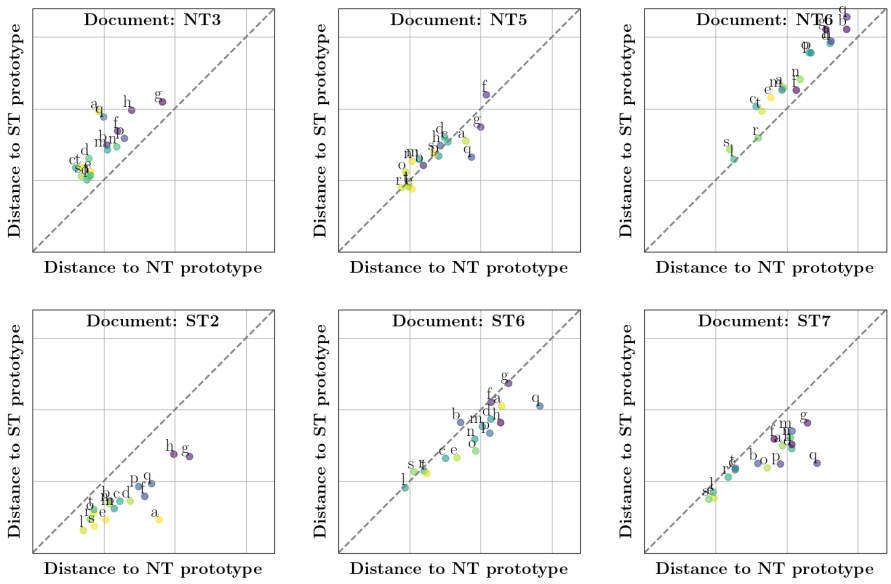


Fig. 6: **Document comparison graphs**.

size, certain letters are simplified, resulting in forms that are closer to ST, like open ‹a› (𝔞) - a characteristic of early NT samples -, and spurless ‹p› (𝔭). On the contrary, a closer examination of ‹g› reveals it is in reality characteristic of NT, but not well modeled by our prototypes, in part because of the high level of variation, resulting in a blurred prototype, and in part because it was too different from the reference documents, which was actually flagged by our automatic failure case identification. **ST6**, copied in Florence at the end of 14$^{\text{th}}$/beginning of 15$^{\text{th}}$ c., also presents two class diverging characters, ‹b› and ‹s/ſ›. For ‹s/ſ›, even though not directly obvious due to the fusion of the two allographs, ‹ſ›'s foot is curved to the right (𝔣), a characteristic proper to NT. At the same time, ‹b›'s slightly dislocated lobe (𝔟), also present in ‹h› (𝔥) and ‹p› (𝔭), characteristic of later examples of *Textualis* [16], is disrupting the typical circular lobe shape of ST. These dislocated lobes are clear in our Figure 7.

Later Textualis examples. Later specimens of both subtypes present notable differences from their earlier counterparts. **NT6**, copied in the last quarter of the 14$^{\text{th}}$ c. is an example of later Northern *Textualis*, and, while the graph shows that all prototypes are closer to the NT prototypes than the ST prototypes, it also clearly shows that they are very different from both, i.e., they are more on the top right of the graph. This document particularity lies in its strict angular forms, with diamond-shaped minim feet (𝔦𝔫) as well as its exaggerated narrow shapes, with a total absence of round strokes for arcs (𝔡 𝔠 𝔬), typical of 14$^{\text{th}}$-15$^{\text{th}}$ c. (esp. Northern) *Textualis*. **ST7** is an example of late, 15$^{\text{th}}$ c. Iberian

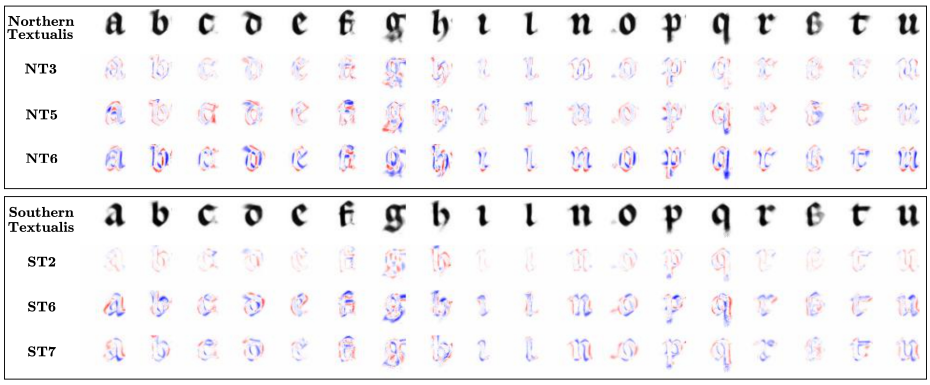


Fig. 7: **Visual comparison.** Subtype and document prototypes and their pixel-wise differences with positive values in blue and negative in red.

Textualis, discussed separately by Derolez due to its tendency towards more angular forms. Particularities of this type lie in the presence of hairlines and angular shapes (a c e) alongside typical rounder ones (d o q) with flat tops and feet (b h n p). However, these characteristics are only discernible while looking at the prototypes and their differences, and not directly in the graph, where the prototypes are closer to ST, and not particularly poorly reconstructed. This can be understood both by the fact that these particularities do not make the prototypes more similar to the NT ones, and by the fact that another Iberian *Textualis* (ST4), from the same manuscript but from a different hand, was used in our reference set, and thus our ST prototype already models some characteristics of this *Textualis* type. This highlights the impact of the prototype training data on our analysis, prototypes utilized as the basis for our comparison graph axes.

5 Conclusion

In this work, we introduced a deep learning-based methodology for interpretable script comparison and analysis. By applying it to the two subtypes of *Textualis Formata* script type defined by A. Derolez —Northern and Southern *Textualis*— we showed how such an approach can complement qualitative document analysis, by quantifying specific elements and summarizing information. We believe our approach contributes to bridging the gap between traditional and learning-based approaches to paleography.

Aknowledgements This study was supported by the CNRS through MITI and the 80|Prime program (CrEMe Caractérisation des écritures médiévales), and by the European Research Council (ERC project DISCOVER, number 101076028). We thank Ségolène Albouy, Raphaël Baena, Sonat Baltacı, Syrine Kalleli, and Elliot Vincent for valuable feedback.

References

1. Aiolli, F., Simi, M., Sona, D., Sperduti, A., Starita, A., Zaccagnini, G.: Spi: a system for palaeographic inspections. AIIA Notizie **4**, 34–38 (1999)
2. Alba, R., Rubin, G., Boschetti, F., Fischer, F., Clérice, T., Chagué, A.: HTRomance, Medieval Italian corpus of ground-truth for Handwritten Text Recognition and Layout Segmentation [dataset] (2023). https://doi.org/10.5281/zenodo.8272751, https://github.com/HTRomance-Project/medieval-italian, v1.0.1
3. Baird, H.S.: Model-directed document image analysis. In: Proceedings of the Symposium on Document Image Understanding Technology. vol. 1 (1999)
4. Berg-Kirkpatrick, T., Durrett, G., Klein, D.: Unsupervised transcription of historical documents. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 207–217 (2013)
5. Bordier, J., Gille Levenson, M., Brisville-Fertin, O., Clérice, T., Chagué, A.: HTRomance, Medieval Spain corpus of ground-truth for Handwritten Text Recognition and Layout Segmentation [dataset] (2023), https://github.com/HTRomance-Project/middle-ages-in-spain, v0.0.6
6. Cencetti, G.: Lineamenti di Storia della scrittura latina: dalle lezioni di Paleografia (Bologna a.a. 1953-54). Guerrini Ferri, G., Bologna (1997)
7. Christlein, V., Bernecker, D., Maier, A., Angelopoulou, E.: Offline writer identification using convolutional neural network activation features. In: Pattern Recognition: 37th German Conference, GCPR 2015, Aachen, Germany, October 7-10, 2015, Proceedings 37. pp. 540–552. Springer (2015)
8. Ciula, A.: Digital palaeography: using the digital representation of medieval script to support palaeographic analysis. Digital Medievalist **1** (2005)
9. Cloppet, F., Daher, H., Églin, V., Emptoz, H., Exbrayat, M., Joutel, G., Lebourgeois, F., Martin, L., Moalla, I., Siddiqi, I., Vincent, N.: New Tools for Exploring, Analysing and Categorising Medieval Scripts. Digital Medievalist **7** (Feb 2012). https://doi.org/10.16995/dm.44
10. Cloppet, F., Eglin, V., Helias-Baron, M., Kieu, V.C., Stutzmann, D., Vincent, N.: ICDAR 2017 Competition on the Classification of Medieval Handwritings in Latin Script. In: 14th IAPR International Conference on Document Analysis and Recognition. ICDAR 2017. pp. 1371–1376. CPS, Kyoto (2017). https://doi.org/DOI 10.1109/ICDAR.2017.224, 00000
11. Cloppet, F., Eglin, V., Kieu, V.C., Stutzmann, D., Vincent, N.: ICFHR2016 Competition on the Classification of Medieval Handwritings in Latin Script. Proceedings of International Conference on Frontiers in Handwriting Recognition pp. 590–595 (2016)
12. Clérice, T., Chagué, A., Vlachou-Efstathiou, M.: CREMMA Medii Aevi [dataset] (Oct 2023), https://github.com/HTR-United/CREMMA-Medieval-LAT, v0.1.2
13. Clérice, T., Pinche, A.: Choco-Mufin, a tool for controlling characters used in OCR and HTR projects (Sep 2021). https://doi.org/10.5281/zenodo.5356154, https://github.com/PonteIneptique/choco-mufin
14. Daher, H., Églin, V., Brès, S., Vincent, N.: Étude de la dynamique des écritures médiévales: analyse et classification des formes écrites. Gazette du livre médiéval **56**(1), 21–41 (2011)
15. Davis, L.F.: Towards an automated system of script classification. Manuscripta **42**(3), 193–201 (1998)

16. Derolez, A.: The palaeography of Gothic manuscript books: From the twelfth to the early sixteenth century. Cambridge University Press (2003)
17. Djeddi, C., Meslati, L.S., Siddiqi, I., Ennaji, A., El Abed, H., Gattal, A.: Evaluation of texture features for offline arabic writer identification. In: 2014 11th IAPR international workshop on document analysis systems. pp. 106–110. IEEE (2014)
18. Djeddi, C., Siddiqi, I., Souici-Meslati, L., Ennaji, A.: Codebook for writer characterization: A vocabulary of patterns or a mere representation space? In: 2013 12th International Conference on Document Analysis and Recognition. pp. 423–427. IEEE (2013)
19. Gasparri, F.: Remarques sur la terminologie paléographique. Revue d'Histoire des Textes **13**(1964), 111–114 (1966)
20. Gilissen, L.: L'expertise des écritures médiévales: recherche d'une méthode avec application à un manuscrit du XIe siècle: le lectionnaire de Lobbes, Codex Bruxellensis 18018. Scriptorium/Les Publications de Scriptorium **6** (1973)
21. Gilissen, L.: III. ductus et rapport modulaire. Scriptorium **29**(2), 235–244 (1975)
22. Gille Levenson, M.: Towards a general open dataset and model for late medieval Castilian text recognition (HTR/OCR). Journal of Data Mining and Digital Humanities (2023). https://doi.org/10.46298/jdmdh.10416
23. Glaise, A., Clérice, T., Boschetti, F., Fischer, F., Chagué, A.: HTRomance, Medieval Latin corpus of ground-truth for Handwritten Text Recognition and Layout Segmentation [dataset] (2024), https://github.com/HTRomance-Project/medieval-latin, v0.0.6
24. Goyal, K., Dyer, C., Warren, C., G'Sell, M., Berg-Kirkpatrick, T.: A probabilistic generative model for typographical analysis of early modern printing. arXiv preprint arXiv:2005.01646 (2020)
25. Hannad, Y., Siddiqi, I., El Kettani, M.E.Y.: Writer identification using texture descriptors of handwritten fragments. Expert Systems with Applications **47**, 14–22 (2016)
26. Hassner, T., Rehbein, M., Stokes, P.A., Wolf, L.: Computation and palaeography: potentials and limits. Kodikologie und Paläographie im digitalen Zeitalter **3**, 1–30 (2015)
27. He, S., Schomaker, L.: Delta-n hinge: rotation-invariant features for writer identification. In: 2014 22nd International conference on pattern recognition. pp. 2023–2028. IEEE (2014)
28. He, S., Schomaker, L.: Deep adaptive learning for writer identification based on single handwritten word images. Pattern Recognition **88**, 64–74 (2019)
29. He, S., Wiering, M., Schomaker, L.: Junction detection in handwritten documents and its application to writer identification. Pattern Recognition **48**(12), 4036–4048 (2015)
30. Hochberg, J., Kelly, P., Thomas, T., Kerns, L.: Automatic script identification from document images using cluster-based templates. IEEE Transactions on Pattern Analysis and Machine Intelligence **19**(2), 176–181 (1997)
31. Kestemont, M., Christlein, V., Stutzmann, D.: Artificial paleography: computational approaches to identifying script types in medieval manuscripts. Speculum **92**(S1), S86–S109 (2017)
32. Kopec, G.E., Lomelin, M.: Supervised template estimation for document image decoding. IEEE Transactions on Pattern Analysis and Machine Intelligence **19**(12), 1313–1324 (1997)
33. Kordon, F., Weichselbaumer, N., Herz, R., Mossman, S., Potten, E., Seuret, M., Mayr, M., Christlein, V.: Classification of incunable glyphs and out-of-distribution

detection with joint energy-based models. International Journal on Document Analysis and Recognition (IJDAR) **26**(3), 223–240 (2023)
34. Lebourgeois, F., Moalla, I.: Caractérisation des écritures médiévales par des méthodes statistiques basées sur les cooccurrences. Gazette du livre médiéval **56-57**, 72–100 (2011)
35. Leroy, N., Pinche, A., Camps, J.B., Clérice, T., Chagué, A.: HTRomance, Medieval French corpus of ground-truth for Handwritten Text Recognition and Layout Segmentation [dataset], https://github.com/HTRomance-Project/medieval-french, v0.0.7
36. Mamatsis, A.R., Mamatsi, E., Chalatsis, C., Arabadjis, D., Kampouri, P., Papaodysseus, C.: A novel methodology for writer (hand) identification: establishing rigas feraios wrote two important greek documents discovered in romania. Heritage Science **11**(1), 38 (2023)
37. McGillivray, M.: Statistical analysis of digital paleographic data: what can it tell us? Digital Studies/Le champ numérique **11** (2005)
38. Moalla, I., Lebourgeois, F., Emptoz, H., Alimi, A.: Image analysis for palaeography inspection. In: Second International Conference on Document Image Analysis for Libraries (DIAL'06). pp. 8–pp. IEEE (2006)
39. Muzerelle, D.: À la recherche d'algorithmes experts en écritures médiévales. Gazette du livre médiéval **56**(1), 5–20 (2011). https://doi.org/10.3406/galim.2011.1979
40. Nigam, S., Verma, S., Nagabhushan, P.: Document analysis and recognition: A survey. Authorea Preprints (2021)
41. Oeser, W.: Das «a» als Grundlage für Schriftvarianten in der gotischen Buchschrift. Scriptorium **25**(1), 25–45 (1971)
42. Ornato, E.: Ii. statistique et paléographie: peut-on utiliser le rapport modulaire dans l'expertise des écritures médiévales? Scriptorium **29**(2), 198–234 (1975)
43. Parkes, M.B.: English cursive book hands, 1250-1500. Oxford : Clarendon (1969)
44. Pinche, A.: Cremma Medieval [dataset] (Oct 2023), https://github.com/HTR-United/cremma-medieval
45. Pinche, A., Clérice, T., Chagué, A., Camps, J.B., Vlachou-Efstathiou, M., Levenson, M.G., Brisville-Fertin, O., Boschetti, F., Fischer, F., Gervers, M., et al.: Catmus-medieval: Consistent approaches to transcribing manuscripts (2023), https://univ-lyon3.hal.science/hal-04453952v1
46. Poulle, E.: Paléographie et méthodologie: vers l'analyse scientifique des écritures médiévales. Bibliothèque de l'École des chartes **132**(1), 101–110 (1974)
47. Ramel, J.Y., Sidère, N., Rayar, F.: Interactive layout analysis, content extraction, and transcription of historical printed books using pattern redundancy analysis. Literary and Linguistic Computing **28**(2), 301–314 (2013)
48. Schomaker, L., Bulacu, M.: Automatic writer identification using connected-component contours and edge-based features of uppercase western script. IEEE transactions on pattern analysis and machine intelligence **26**(6), 787–798 (2004)
49. Schomaker, L., Franke, K., Bulacu, M.: Using codebooks of fragmented connected-component contours in forensic and historic writer identification. Pattern Recognition Letters **28**(6), 719–727 (2007)
50. Siddiqi, I., Vincent, N.: Text independent writer recognition using redundant writing patterns with contour-based orientation and curvature features. Pattern Recognition **43**(11), 3853–3865 (2010)
51. Siglidis, I., Gonthier, N., Gaubil, J., Monnier, T., Aubry, M.: The learnable typewriter: A generative approach to text line analysis (2023), https://arxiv.org/abs/2302.01660

52. Sirat, C.: L'examen des écritures: l'œil et la machine: essai de méthodologie. Ed. du Centre National de la Recherche Scientifique (1981)
53. Smith, M.: (review) derolez (albert), the palaeography of gothic manuscript books. from the twelfth to the early sixteenth century, cambridge, 2003. Scriptorium **58**(2), 274–279 (2004)
54. Sommerschield, T., Assael, Y., Pavlopoulos, J., Stefanak, V., Senior, A., Dyer, C., Bodel, J., Prag, J., Androutsopoulos, I., de Freitas, N.: Machine learning for ancient languages: A survey. Computational Linguistics **49**(3), 703–747 (2023)
55. Stansbury, M.: The computer and the classification of script. In: Kodikologie und Paläographie im digitalen Zeitalter - Codicology and Palaeography in the Digital Age. vol. 2, p. 238. BoD, Norderstedt (2009)
56. Stokes, P.A.: Describing handwriting, part i-v. Blog Post (2011), https://digipal.eu/blog/describing-handwriting-part-i/, last accessed on 15/03/2024
57. Stutzmann, D.: Variability as a key factor for understanding medieval scripts: the oriflamms project (anr-12-corp-0010). In: Brookes, S., Rehbein, M., Stokes, P. (eds.) Digital Palaeography. Digital Research in the Arts and Humanities, Routledge, https://halshs.archives-ouvertes.fr/halshs-01778620
58. Stutzmann, D.: Nomenklatur der gotischen Buchschriften: Nennen? Systematisieren? Wie und wozu? (Rezension über: Albert Derolez: The Palaeography of Gothic Manuscript Books. From the Twelfth to the Early Sixteenth Century. Cambridge u.a.: Cambridge University Press 2003.). IASLonline (2005), http://www.iaslonline.de/index.php?vorgang_id=995
59. Stutzmann, D.: Paléographie statistique pour décrire, identifier, dater... Normaliser pour coopérer et aller plus loin ? In: Kodikologie und Paläographie im digitalen Zeitalter 2 - Codicology and Palaeography in the Digital Age 2, pp. 247–277. No. 3 in Schriften des Instituts für Dokumentologie und Editorik, BoD, Norderstedt (2010), https://kups.ub.uni-koeln.de/4353/
60. Stutzmann, D.: Système graphique et normes sociales : pour une analyse électronique des écritures médiévales. In: Medieval Autograph Manuscripts. Proceedings of the XVIIth Colloquium of the Comité International de Paléographie Latine, held in Ljubljana, 7-10 September 2010, pp. 429–434. No. 36 in Bibliologia, Brepols, Turnhout (2013), https://www.brepolsonline.net/doi/10.1484/M.BIB.1.101494
61. Stutzmann, D.: Clustering of medieval scripts through computer image analysis: Towards an evaluation protocol. Digital Medievalist **10** (Jun 2016). https://doi.org/10.16995/dm.61
62. Stutzmann, D.: Ecmen (2017), https://github.com/oriflamms/ECMEN
63. Stutzmann, D.: Les «manuscrits datés», base de données sur l'écriture. In: De Robertis, T., Giovè Marchioli, N. (eds.) Catalogazione, storia della scrittura, storia del libro. I Manoscritti datati d'Italia vent'anni dopo, pp. 155–207. SISMEL - Edizioni del Galluzzo, Firenze (2017)
64. Stutzmann, D., Helias-Baron, M.: ICDAR 2017 Competition on the Classification of Medieval Handwritings in Latin Script - Dataset (Nov 2017), https://zenodo.org/record/5527690
65. Tang, Y., Wu, X.: Text-independent writer identification via cnn features and joint bayesian. In: 2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR). pp. 566–571. IEEE (2016)
66. Tomiello, A.: Dalla littera antiqua alla littera textualis. Gazette du livre médiéval **29**(1), 1–6 (1996)
67. Wolf, L., Dershowitz, N., Potikha, L., German, T., Shweka, R., Choueka, Y.: Automatic paleographic exploration of genizah manuscripts. In: Fischer, F., Fritze, C.,

Vogeler, G. (eds.) Kodikologie und Palaographie im Digitalen Zeitalter 2 - Codicology and Palaeography in the Digital Age 2, Schriften des Instituts für Dokumentologie und Editorik, vol. 3, pp. 157–179. BoD, Norderstedt, Germany (2011)
68. Xu, Y., Nagy, G.: Prototype extraction and adaptive ocr. IEEE Transactions on Pattern Analysis and Machine Intelligence **21**(12), 1280–1296 (1999)