In Search of the Scribe

Letter Spotting as a Tool for Identifying Scribes in Large Handwritten Text Corpora

Lasse Mårtensson[a], Ekta Vats[b], Anders Hast[b] & Alicia Fornés[c]
[a]Stockholm University, [b]Uppsala University, [c]Universitat Autònoma de Barcelona

In this article, a form of the letter spotting-method is used on a large set of handwritten documents in order to identify those that contain script of similar execution. The point of departure for the investigation is the mediaeval Swedish manuscript Codex Holmiensis D 3. The main scribe of this manuscript has yet not been identified in other documents. The current attempt aims at localizing other documents that display a large degree of similarity in the characteristics of the script, as possible candidates for being produced by the same scribal hand. In the letter spotting process, a set of 'g':s, 'h':s and 'k':s have been selected as templates, and a search has been made for close matches among the mediaeval Swedish charters. The search resulted in a number of charters that displayed great similarities with the manuscript D 3. The method thus proofed to be a very efficient sorting tool, localizing similar script samples.

Keywords: digital palaeography, mediaeval charters, mediaeval manuscripts, scribal attribution, word spotting, writer identification

The increasing interest in digitization of historical manuscript collections for their preservation has led to exploration of automated methods for information retrieval. However, automatic recognition of handwritten historical documents is a challenging task from a technical perspective.

Mårtensson, Lasse et al. "In Search of the Scribe: Letter Spotting as a Tool for Identifying Scribes in Large Handwritten Text Corpora". *HUMAN IT* 14.2 (2019): 95–120.

The causes to the difficulties are several, for instance the variability of the script signs produced by hand, damages on the script surface (paper or parchment), ink bleed-through, faded ink, and even variation due to lighting variation during the document scanning.

The aim of this article is to use letter spotting for identifying similar scribal hands in a large corpus of handwritten documents. We have used mediaeval Swedish material for our investigation (see further below). The reasons for searching for scribes in mediaeval handwritten material are several. One of the most central ones is that of placing an undated and anonymous manuscript (which they most often are), you can place it into a chronological and a geographical context if you can identify the same scribal hand/hands in charters. In the charters, the place of provenance is given and they are furthermore dated. Since a scribe can be expected to have been active for approximately 20–40 years (Åström 2003, 57; Wiktorsson 2004, 236), the dating of the charters in which the same scribal hand is to be found narrows down the possible time of origin for the manuscript. As stated below, the points of departure for dating late mediaeval Swedish manuscripts are few, and therefore the mentioned procedure is of the highest importance for this material.

The present investigation belongs to a set of methods that is often referred to as *word spotting* (see for instance Llados *et al*. 2012; Wahlberg, Mårtensson & Brun 2014; Hast & Fornés 2016; Giotis *et al*. 2017). The term *word spotting* refers to the fact that the process results in word forms being identified and localized in handwritten documents. The most common purpose for this is when you have a lot of unedited material and a researcher needs to look for words for some reason, for instance for lexicographical purposes, or simply for finding relevant text passages in unedited material. In the current investigation, however, the usage of this method is another, namely to identify scribes in large handwritten text corpora. To the best of our knowledge, this usage of word spotting is new, and the present investigation must be followed by further research and evaluation. As the current process focuses on individual letters, the

term *letter spotting* will be used throughout the text. There exist a large set of digital methods for scribal attribution (also used on mediaeval Swedish material, see, e.g., Wahlberg, Mårtensson & Brun 2014; Mårtensson, Wahlberg & Brun 2015; see also He *et al*. 2016), but the present investigation has rather the focus on identifying similar script samples in a large collection of handwritten documents that display similarities regarding the morphology of the script signs.

The material for this investigation is the mediaeval Swedish manuscript, *Codex Holmiensis D 3* (The Royal Library, Stockholm), and the entire mediaeval Swedish charter corpus, to the extent that this has been photographed. The manuscript is anonymous, but dated to 1488–90 on the basis of the water marks. It was produced by two scribes, of which the first wrote pages 3–13 and the second wrote pages 14–641 (Backman 2017, 59, 67–68). Henceforth in this article, the second scribe will be referred to as the main scribe. Recently, the manuscript has been in focus in an investigation by Agnieszka Backman (2017), with an emphasis on the codicological and material aspects of the manuscript.

In the present investigation, it is the main scribe that is in focus. The manuscript comprises a varied content, for instance *Erikskrönikan*, the so-called *Eufemiavisorna* (Herr Ivan Lejonriddaren, Flores och Blanzeflor and Hertig Fredrik av Normandie), the narrative of Charles the great in Old Swedish, but also some religious works, such as Tungulus, a story about a soul's journey through the realms of the dead. The reason for our interest in this specific manuscript stems from the fact that one of the authors of the present article, Lasse Mårtensson, is conducting a philological investigation of the Tungulus/Tundalus-text in the Old Swedish tradition. It would be of great interest for that purpose to see if the scribe of this manuscript has produced charters, as this would render a possibility to place the manuscript into a larger context. As stated, the scribe of this manuscript is anonymous, and he/she is not identified in Per-Axel Wiktorsson's work *Skrivare i det medeltida Sverige* (2015, 1–4).

Letter Spotting as a Tool for the Identification of Scribes

The common procedure for most letter spotting methods is that a template of a word form is chosen (a sequence of letters, or, as in our case, individual letters), and then the computer searches for other formations in the material that come close to the template. The idea behind the current usage of letter spotting, i.e., for identifying scribal hands in large masses of handwritten material, goes as follows. When performing letter spotting in handwritten text, one must allow for some variation. If exact similarity is demanded, the process would not render one single hit, as handwritten script always contain variation. However, the degree of variation allowed can be controlled. If a large degree of variation is allowed, the process would render many hits, correct ones but probably also many false hits (i.e., other letter or letter sequences than the ones of the template). If, however, a very small degree of variation is allowed for, fewer hits are given, but the hits will probably come very close to the template. When searching for words, and the aim is to extract as many occurrences as possible of the word in question, the latter alternative would not be very well adapted to the purpose, as many correct hits would be missed because they deviated too much from the template. But if the purpose instead is to identify those very letters or letter sequences that have a very similar form to that of the template, this approach could be efficient. The purpose here is, then, to make a letter spotting search allowing for very small variation, making the computer localize only the occurrences that show very large similarities with the template. This would give hits that share the form of the template on a micro level, and these are possible candidates for being performed by the same scribal hand as the letter(s) of the template.

It must be stressed that this letter spotting method does not give automatic answers to the scribal identity in the documents being investigated. The point is that through this method, a search is conducted through a very large corpus, consisting of many thousands of documents, and find those that display similarities, and which may be possible candidates for

containing the same scribal hand. The process of scribal attribution is another aspect, and has to be addressed with other methods.

The Letters 'g', 'h' and 'k'

In this investigation, searches have been made for separate letters (i.e., not sequences). The letters selected for the purpose are 'g', 'h' and 'k'. These have been chosen because they together provide suitable tools for sorting out charters that come close in time to the manuscripts and as they contain variation that can be expected to carry important information regarding the individual variation of the scribe. All of them consist of several components, and they have a rather complex morphology. Of course, other letters could have come in question, and in his mapping of the mediaeval Swedish scribes, Wiktorsson (2015a, 27) focuses on the form of the letters 'g', 'w-', 'æ', 'ø', 'y', '-n', 'k' and 'h'. Especially 'ø' would have been very interesting, as the 'ø' by the main scribe of D 3 has a very characteristic shape. There are, however, some technical problems concerning the search for this type of 'ø', as the diacritic sign of this type of 'ø' is easily confused with 'o' with other script elements from the line above descending close to the loop.

The letter 'g' has in the past been used for the purpose of scribal attribution (see, e.g., Wiktorsson 2015, 27), and it can also rather easily be extracted for the present purpose. Furthermore, there is a chronological component regarding 'g' that in our case makes it suitable for identifying charters from a specific time period. There is a 'g'-type that has its origin at the end of the Middle Ages in the Swedish script and that is used in D 3. This 'g' is characterized by a stroke drawn very long below the baseline, turning then sharply upwards (see figure 1). Thus, as the type of 'g' used in the investigated manuscript manifests chronological characteristics, a large number of charters from other time periods than that of the manuscript should be sorted out.

Figure 1. A 'g' from the manuscript D 3

The search for 'h' is motivated from another perspective. The 'h' does not, as 'g', display clear chronological variation. The 'h' used in D 3 is the one that is normally used in the Cursive script of the Late Middle Ages in Sweden (and elsewhere), with a loop on the top of the ascender and the right vertical stroke descending below the baseline (see figure 2). This letter form gives a number of possibilities for individual variation, such as the direction and the shape of the right stroke descending below the baseline and also the height proportions between high and low script elements. This letter is therefore a good tool for sorting scribes on an individual level.

Figure 2. An 'h' from the manuscript D 3

The third letter chosen, 'k', displays possibilities for both chronological and individual variation. Regarding the former, the type of 'k' with the leg extending vertically belongs to the late Middle Ages. For the individual variation, it should be noted that 'k' consists of several components and therefore gives several possibilities for individual variation. A 'k' from the

manuscript D 3 is given in figure 3. It should be noted that the baseline is on the level of the leg, and the ascender thus extends below the baseline.

Figure 3. A 'k' from the manuscript D 3

Other letters that are of interest are for instance 'æ' and 'ø', and especially the latter letter has a very characteristic form in the manuscript D 3. However, it is very difficult to perform letter spotting on these letters, as they are executed in D 3. The 'æ' is identical to 'a', apart from the fact that it has a dot up to the right of the body of the letter (see figure 4). It is therefore easily confused with 'a' when there is another letter beginning on the same spot as the dot. A search for 'æ' would thus render many hits on 'a'. The 'ø' i D 3 consists of a 'ø' with one diacritical sign above the body of the letter and one below it (see figure 5). The 'ø' is therefore easily confused with 'o' when other letters from the line above are descending over the 'o'. However, in the evaluation, the 'æ' and the 'ø' will be investigated in the charters that come close to the manuscript D 3 in the letter spotting process.

Figure 4. An 'æ' from the manuscript D 3

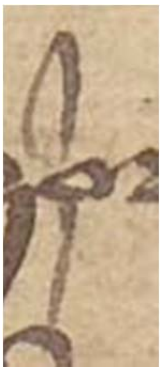

Figure 5. An 'ø' from the manuscript D 3

One could of course also consider searching for words for our purpose, but a problem is that the charters are rather short text, and in order for a regular pattern of a scribe to become clear, you need a sufficient number of occurrences, and in the charters this can in some cases be a problem. Also if one looks for common words such as 'and', you still have to build the judgement on a rather small set of occurrences. Thus, the incidental variation could distort the result. For separate letters, the result can be built on many occurrences.

Technical Aspects of Letter Spotting Methodology

This section describes the letter spotting methodology from a technical point of view. A simple segmentation-free handwritten letter spotting algorithm is proposed, that begins with pre-processing of document images for background noise removal. Thereafter, key-points are detected for the letters in the document and further represented using feature descriptors. Finally, matching is performed for writer identification. The following text explores the letter spotting algorithm in a step-by-step manner.

Pre-processing

Before a letter spotting method for writer identification can be applied, the input document image needs to be pre-processed. To begin with, the textual region of the document image must be separated from the non-text or irrelevant parts. For instance, when document images are being scanned in libraries, a standard protocol is followed to add an additional ruler and colour pattern to the images in order to render the colour information and real scale for reconstruction. For high-level document image analysis, these additional objects are typically regarded as noise, and are required to be removed in the pre-processing steps. Similarly, the artefacts in the background that contain non-textual information are generally irrelevant for writer identification.

Figure 6. Original coloured input document image from the SDHK database

In the first step, the coloured input document image is converted to a grayscale image, and additional objects such as rulers and colour bars are

removed, such that the textual region of interest is available for further processing. Figure 6 presents the original document image from the SDHK database (https://sok.riksarkivet.se/sdhk), and figure 7 represents the grayscale image obtained after initial pre-processing. However, at this stage the document image contains background noise in the form of stains, wrinkles and contrast variation.

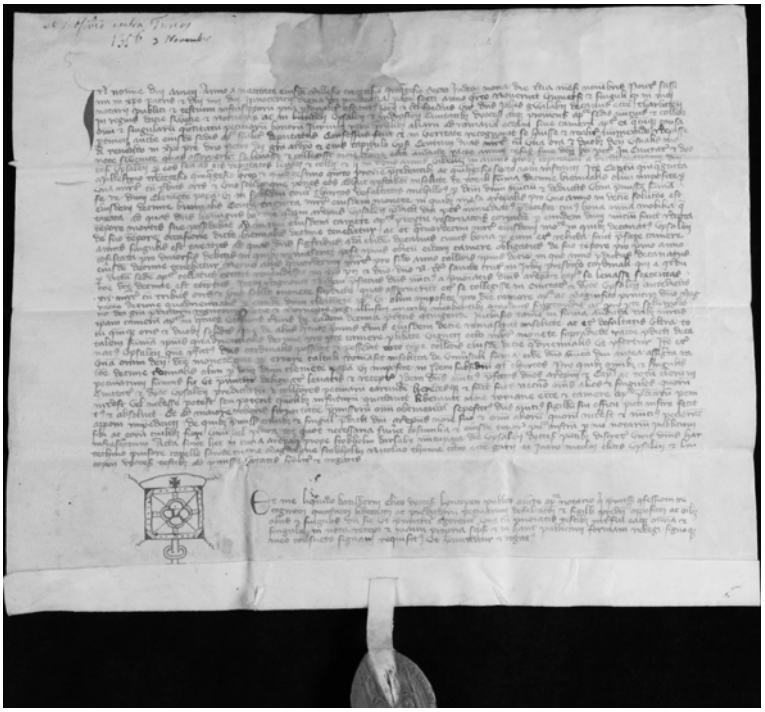

Figure 7. Pre-processed grayscale image with non-textual objects removed

Therefore, in the second step, the document image is passed through two band-pass filters (Vats, Hast & Singh 2017) for further background noise removal. Figure 8 represents the binarized image with foreground text preserved and noisy background removed to a considerable extent.

Figure 8. Binarized document image with noisy background removed to a fair extent

Key-point Detection and Feature Representation

Key-point detection refers to finding interest points in an image that contain crucial information. The selection of key-point detectors has a great impact on the performance of letter spotting algorithms. After a set of key-points has been detected, a suitable representation of their values has to be defined to allow matching between a query letter and the document image. This work employs a combination of key-point detectors and a fast feature descriptor from our previous work (Hast & Fornés 2016; Hast & Vats 2017) to perform handwritten letter spotting. The key-points are computed for the whole document page and the query letters. The reader

is referred to the above mentioned two works for a detailed description of the key-point detection and feature representation methods.

Matching Process for Letter Spotting

The goal of letter spotting system is to retrieve all instances of user query letters in a set of document images. The proposed letter spotting (or letter matching) method is regarded as segmentation-free as it can be directly applied to whole document page. This section discusses the overall key-point based matching algorithm for segmentation-free letter spotting.

To begin with, a nearest neighbour search is performed within the subgroups of the detected key-points. An optimal sliding window is selected to perform the key-point based letter matching. When a letter is found, the matching points are removed from the set of points so that the same letter is not found again in the subsequent sliding window searches.

The resultant correspondences from the key-point matching between the query letter and the sliding window contain certain outliers (or false positives) that must be removed for reliable results. Therefore, a simple preconditioner (Hast & Kylberg 2015) is used that creates a cluster of corresponding matches and finds potential inliers. Preconditioner based matching efficiently captures complex variations in handwritten letters.

The matching algorithm computes the similarity between the feature representation (feature vectors) of sample query letters (say, ‘g’, ‘h’ and ‘k’), and the document image at page level. In total, eight query examples of each letter are used as a template for searching a query letter in the document page. Figure 9 presents an interesting visualization of the query letters ‘g’, ‘h’ and ‘k’ in three-dimensional search space based on our experimental analysis. The found instances of query letter ‘g’ are represented in ‘red’, and for letters ‘h’ and ‘k’ in ‘green’ and ‘blue’, respectively. Both the position as well as the size of the coloured spheres, and the colour represents the similarity between the query letters and the

corresponding letters found in the document. Since each axis corresponds to the similarity measure used, it means that spheres close to the origin have very low similarity. Some documents might have varying degree of similar 'k' but dissimilar 'g' and 'h', and these correspond to spheres along the 'k' axis. The same reasoning applies to each letter and the corresponding axis. Hence, the documents we search are as far away from the origin as possible, and are therefore found in the upper corner opposite to the origin.

The similarity between letters is computed on the basis of confidence scores, and in the case of writer identification it is calculated by taking the mean of all the hits, where hits are defined as the number of times a certain letter in the document was found by the eight different query letters. The found letter with the highest confidence score is returned as an output. In this way, the results obtained from letter spotting can be utilized for writer identification. Since, each writer inhibits a unique way of writing different letters in a handwritten document, the results from letter spotting of sample letters ('g', 'h', 'k') can be used to deduce the writer identity. This is possible because a letter handwritten by a scribe can be similar but not exactly the same as a letter written by another scribe. Typically, handwritten text represents high variability in writing styles, not only among different authors, but also for the documents written by a specific author. Also, handwriting is unconstrained, where the letters may be skewed, slanted, or written in cursive manner, etc. The framework testing of the writer identification method based on the proposed letter spotting algorithm is progressing with positive results, and will be further empirically evaluated as future work.

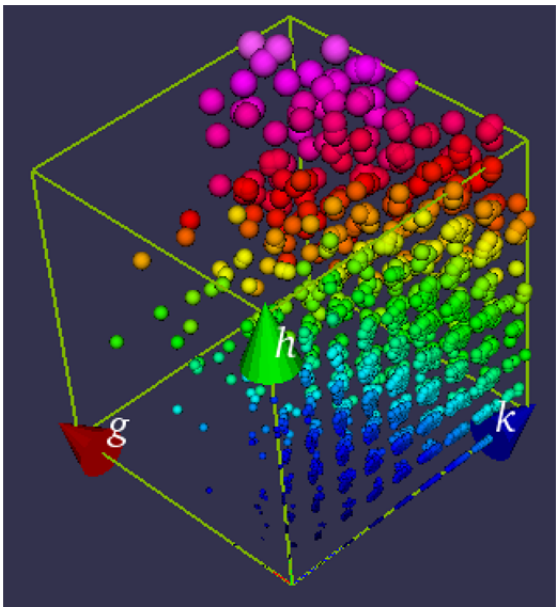

Figure 9. Visualization of the query letters 'g', 'h' and 'k' in three-dimensional search space. Each axis corresponds to the similarity measure used, and it means that spheres close to the origin have very low similarity. The purple coloured spheres represent the documents having highest similarity to the scribe in question

It is worth mentioning that by using a feature descriptor (Hast & Vats 2017), the letter spotting problem is reduced to a much faster search problem. The matching algorithm is simple and computationally inexpensive, and the use of preconditioner speeds up the matching process to a great extent.

Palaeographical Evaluation

In this section, a philological evaluation will be made of five of the best hits given from the letter spotting process, those charters having the highest mean value. It should be noted that there are a number of hits that are good in the sense that they have a high mean value, but that on a

closer scrutiny displayed several differences from the manuscript D 3. These have been excluded here. Furthermore, in the evaluation made here, external evidence is not taken into account in this evaluation (e.g., place of production and similar aspects). Here, only script features are taken into account. All the charters presented here had a high mean value in the letter spotting process. The mean value is calculated from the similarity of 'g', 'h' and 'k' taken together (as compared to the form of the corresponding letters in D 3), so this is a measure of the similarity of these three letters taken together.

This evaluation is divided into two parts, one that concerns the general impression of the script and one that focuses on the shape and ductus of individual letters. For the first purpose, details from the five charters are shown in figures 11–15, and in figure 10 a detail from the source of the comparison is given, the main hand of the manuscript D 3.

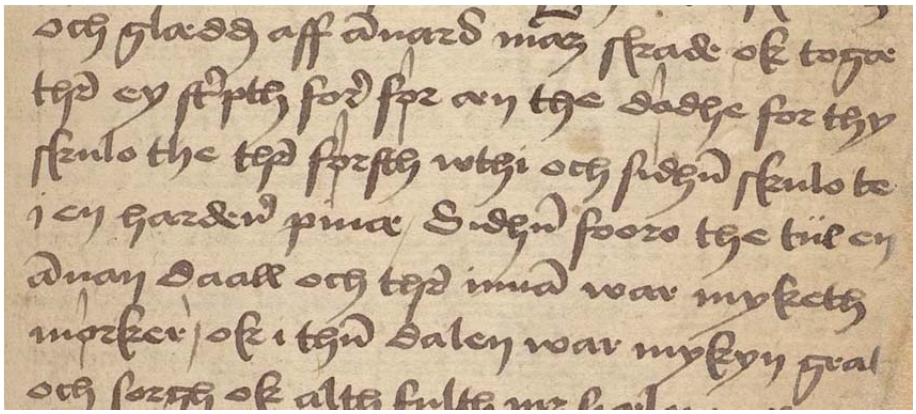

Figure 10. Detail from Cod. Holm. D 3

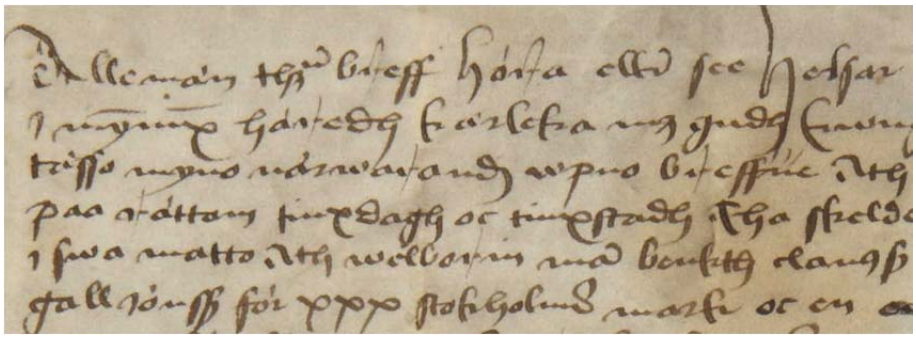

Figure 11. Detail from SDHK 26610

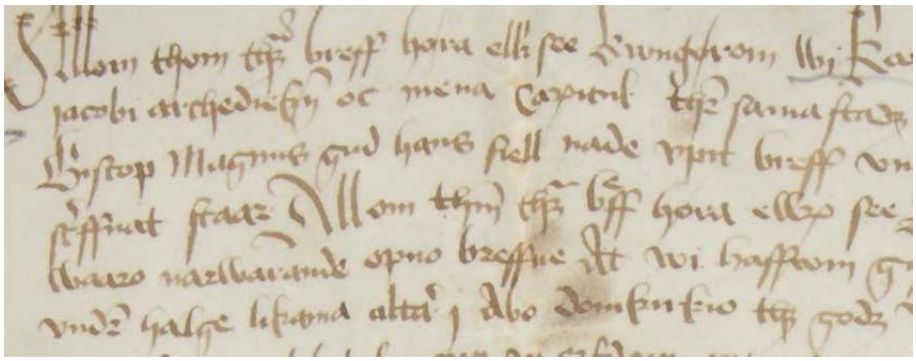

Figure 12. Detail from SDHK 28689

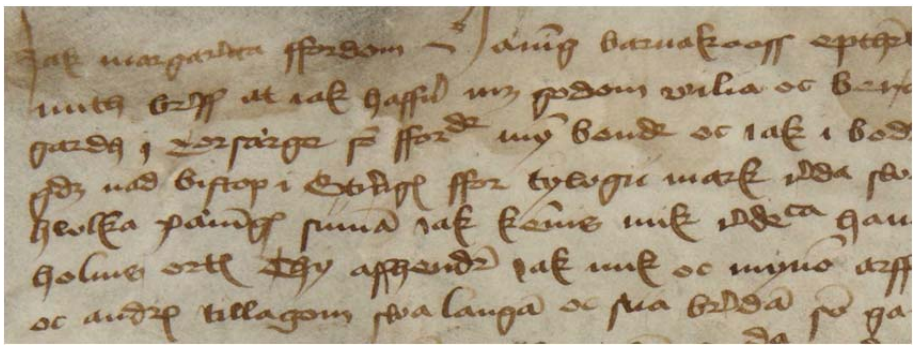

Figure 13. Detail from SDHK 29670

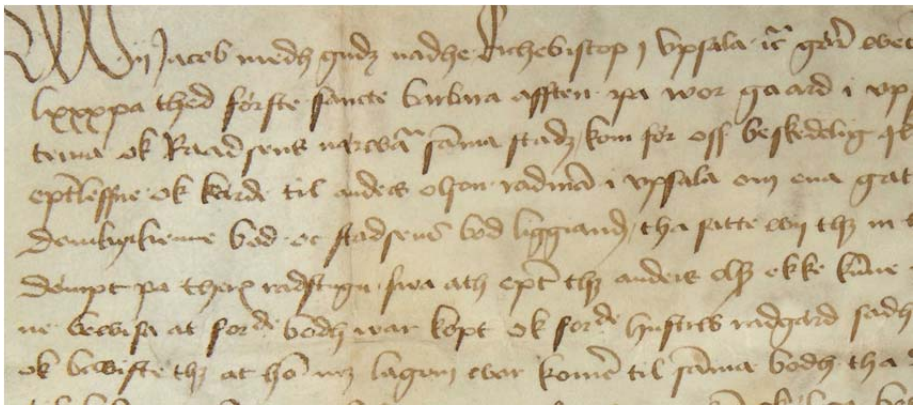

Figure 14. Detail from SDHK 31446

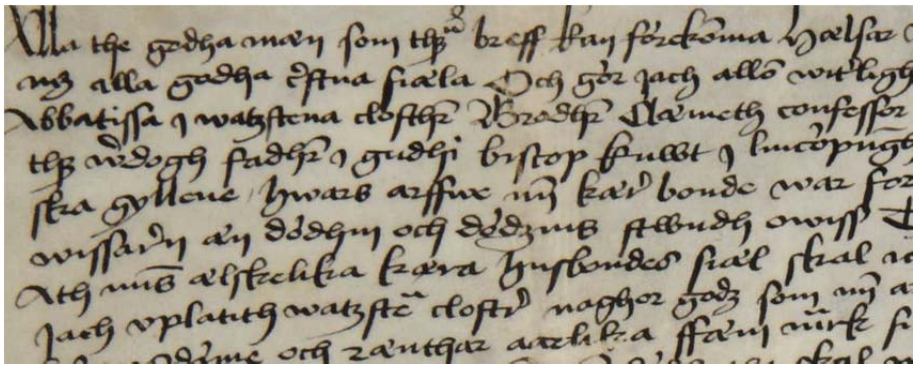

Figure 15. Detail from SDHK 32156

On a first impression, these charters certainly display differences as compared to D 3, although to a varying degree. The charters SDHK 29670 and 32156 look similar to D 3 based on a general impression, whereas SDHK 26610, 28689 and 31446 look more different. However, it is clear that one aspect of the script in these samples separates the charters from the manuscript, and that is the level of execution. The level of execution as a category is a description of the care with which the script has been carried out, and usually three levels are accounted for: Formata, Libraria and Currens (see Derolez 2003, 21). The difference between the

three levels is not sharp, and script falling within the Libraria level, for instance, can vary a great deal. It is clear, though, that the level of execution is higher in D 3 and compared to the script in the charters. The charters also vary a great deal among themselves; the charter SDHK 26610 gives the impression of being very quickly executed, whereas SDHK 28689 appears to have been carried out more carefully. One can expect the level of execution to be higher in a manuscript than in a charter, and of course a scribe was able to vary this aspect. This means that a general impression can be misleading regarding scribal identity. Another aspect that may alter the general impression is the pen (sharpness, hardness etc.). It seems that the charter SDHK 31446 has been written with a blunter pen than the other documents above, as the script is less duo-linear (i.e., displaying less variation in width) there.

For the purpose of comparing the occurring letters, the graphs of the five documents, examples of the different letters from the documents are presented in table 1 below. It must be remembered that these examples only serve as examples, and they do not give the full picture of the shape of the letters in question in the documents. When looking at individual occurrences of letters, one must bear in mind that they contain some degree of incidental variation, existing in these examples and not forming a pattern. The examples given in the table are thus only meant to be illustrations, forming a point of departure for the discussion below.

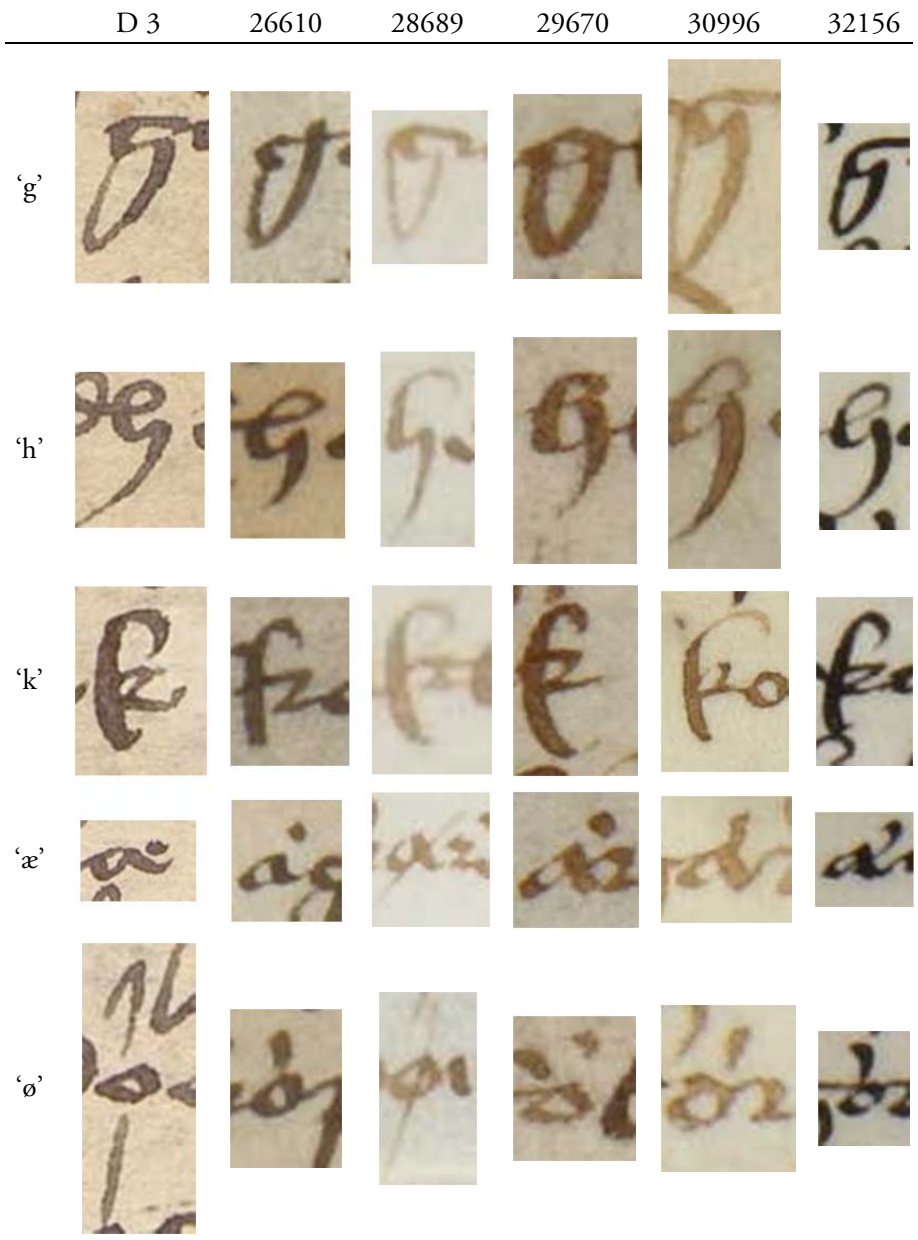

Table 1. Graphs from the manuscript D 3 and the charters

The 'g' is in all three documents mainly of the type where the component below the baseline goes almost straight down, and then turning sharply upwards towards the body of the letter. It should be noticed, however, that other types of 'g':s occur in the documents as well, where the element below the baseline does not form a sharp point, but forms a softer bow. It is clear that the scribes could vary between these forms in one and the same piece of writing. The movements of the pen, the ductus, also appears to be very similar in the 'g':s, even though the 'g' of SDHK 26610 seems to have been performed in a slightly different way. This graph has been carried out in two strokes, whereas the other ones seem to have been performed without the pen being lifted. The current letter spotting method does not take ductus into account, and the shape of letter can certainly come very close even though the ductus differs. This aspect, however, could be an indication of the scribe of SDHK 26610 having a different writing habit regarding 'g' than the rest of the current scribes. It must be remembered though, that the ductus may vary due to the surrounding letters. Some combinations of letters give the scribe the possibility of attaching the letters without the pen being lifted, and this may result in a ductus that is different from when the same letters are in other surroundings. In this case, however, the deviant ductus of 'g' in SDHK 26610 cannot be explained in this way.

The 'h':s of the six documents also show rather large similarities on the eye measure level, all consisting of a loop to the right of the ascender and a prolonged minim to the right. But also here, a deviant ductus can be discerned in one graph, namely in the 'h' of SDHK 29670. This graph seems to have been executed in two strokes, one forming the ascender and the loop attached to this, and the other one consisting of the prolonged minim, i.e., the right element. The latter element must have been drawn separately from the ascender, as it actually crosses the ascender at almost 90 degrees. Furthermore, in the rest of the graphs, the bottom of the ascender is drawn directly to the right element, the prolonged minim,

whereas in the graph of SDHK 29670, the bottom of the ascender is drawn to the following graph. As stated above, the ductus in individual letters could differ due to the surrounding graphs, but the discussed aspect of the ductus of 'h' in SDHK 29670 seems to be recurring, irrespectively of surroundings. A difference regarding the shape can also be seen in the 'h':s of SDHK 28689, namely regarding the shape of the loop to the right of the ascender. In this charter, the loop is not always closed, and indeed does not always form a loop at all, but extends only to the right.

Also the 'k':s of the six documents belong to the same type. They have a horizontal leg on the baseline, and the ascender extends below the baseline, in most cases ending in a hook to the right, although there are instances that deviate from this basic pattern. The example from SDHK 26610 does not have this hook, and this feature is less marked in this charter. Furthermore, the form of the loop on top of the ascender has a slightly different form in the 'k'-graph in SDHK 30996. Here, the loop is not closed (i.e., rather forming a bow), and this is actually the case in most of the 'h':s in this charter. This feature thus forms a pattern in the charter SDHK 30996, and is not a case of incidental variation in the selected graph above.

The letters 'æ' and 'ø' are not accounted for in the current letter spotting process, as it turned out that they were difficult to separate from 'a' and 'o'. When 'æ'- and 'ø'-graphs of the investigated documents are contrasted (see table 1), it becomes clear that none of the charters have throughout the same types as D 3. In all cases, the 'æ' and the 'ø' consist of an 'a' or an 'o' respectively, with a diacritic added. Looking at the mediaeval Swedish script generally, the diacritics of 'æ' and 'ø' can have rather different characteristics. In charter SDHK 28689, the diacritic of 'ø' and 'æ' in both cases consists of a long stroke drawn from above the letter, through or to the right of the body of the letter, to a position below it. These forms are markedly deviant from all the other investigated documents, and are not discussed further below.

The 'æ' in D 3 consists of a one-compartment 'a' with a dot up to the right. The one-compartment 'a' is shared by the 'æ' of all the discussed documents, but the diacritic is not identical. In SDHK 26610 and 29670, the diacritic has the form of a dot, as in D 3, whereas it rather has the form of a comma in SDHK 30996 and 32156.

The 'ø' in D 3 is rather elaborated, consisting of two separated strokes, one above the body of the letter (having the shape of a bow) and one below (a straight stroke). In the charters in the table above, with the exception of SDHK 28689, the diacritic of 'ø' is executed in one stroke, usually in the form of a comma, above the body of the letter. One interpretation is, of course, that this indicates that the scribe that produced the manuscript has not produced any of the charters. One possibility, however, is that the scribe used the more elaborated form primarily in book script, and that the same scribe used a simpler form, with only one diacritic (the one over the 'o'), in charters. As stated above, the level of execution is higher in the manuscript than in the charters. An 'ø' coming fairly close to the one in D 3, but without the stroke below the 'o', can for instance be said to be represented in SDHK 32156.

This evaluation shows that all the charters contain some element of variation as compared to D 3. If these differences are to be explained as variation within one and the same hand, due to variation in the level of execution or similar, or if they are to be interpreted as variation between different hands, will not be addressed in this investigation, but will be discussed further in forthcoming publications. For the purpose of identifying similar script samples, the method has proven efficient. Some of the charters can certainly be counted out as candidates to being produced by the same scribe as D 3, such as SDHK 26610 and 28689, whereas for instance SDHK 32156 has great similarities with D 3 (although having a slightly different form of the 'æ').

Final Remarks

The present letter spotting process has been demonstrated to be a very efficient tool in localizing documents in large collections of handwritten records that display similarities. It is in current attempt not intended as an automatic tool for scribal attribution, but as a way of finding script samples that share the same characteristics regarding the form of the chosen letter forms. However, apart from being an efficient sorting tool, this process also renders a measure of the similarity between the script, as manifested in the chosen letter form, in different documents. An important future step is to evaluate these measures further, to trace these numbers back to the script signs. One of the most important tasks for future research in palaeography is to perform large scale measuring of script features, and to visualize these. The numbers received from the present are likely to contain important information regarding the characteristics of the investigated letters, as the identified script samples (with one example) showed great similarities with the manuscript that was the point of departure (D 3).

A very important aspect of this investigation, not touched upon before in the article, is that this letter process also measures the proportions between high and low script elements (e.g., ascenders vs minims). This measure is of a great value, and this aspect has earlier among researchers in palaeography been pointed out as an aspect that can be of importance for the study of individual variation among the mediaeval scribes (e.g., Gunneng 1992). This will be further examined in a future publication.

Dr. Lasse Mårtensson is a professor in Scandinavian Languages at the Department of Swedish and Multilingualism at Stockholm University, Sweden. His research interests include Old Norse philology, including palaeography, textual criticism and codicology. He is also one of the editors of the journal Scripta Islandica, with a focus on Old Norse studies.
Contact: lasse.martensson@su.se

Dr. Ekta Vats is a postdoctoral researcher in the Division of Visual Information and Interaction, Department of Information Technology at Uppsala University, Sweden. Her research interests include machine learning, computer vision and fuzzy set theory, with focus on document image analysis and understanding. She is currently an active researcher in the field of historical handwritten text recognition, and investigates algorithms for document binarization, semi-automatic text transcription and word spotting.
Contact: ekta.vats@it.uu.se

Dr. Anders Hast is an associate professor at the Division of Visual Information and Interaction, Department of Information Technology at Uppsala University. His research interests include computer graphics, visualization, computer vision and image processing, with main focus on historical handwritten text recognition.
Webpage: www.cb.uu.se/~aht/
Contact: anders.hast@it.uu.se

Alicia Fornés is a Senior Research Fellow at the Universitat Autònoma de Barcelona (UAB, Barcelona), and an adscribed researcher at the Computer Vision Center (CVC, Barcelona). Her research interests include document image analysis, digital humanities and graphics recognition.
Webpage: www.cvc.uab.es/people/afornes/
Contact: afornes@cvc.uab.es

References

BACKMAN, AGNIESZKA (2017). *Handskriftens materialitet. Studier i den fornsvenska samlingshandskriften Fru Elins bok (Codex Holmiensis D 3)*. Uppsala: Department for Scandinavian Languages, Uppsala University.

CODEX HOLMIENSIS D 3. Preserved at the Royal Library, Stockholm, Sweden. (manuscript)

DEROLEZ, ALBERT (2003). *The Palaeography of Gothic Manuscript Books: From the Twelfth to the Early Sixteenth Century*. Cambridge: Cambridge University Press.

DHALI, MARUF A., ET AL. (2017). "A Digital Palaeographic Approach towards Writer Identification in the Dead Sea Scrolls." *Proceedings of the 6th International Conference on Pattern Recognition Applications and Methods-Volume 1*: ICPRAM.

GIOTIS, ANGELOS P., ET AL. (2017). "A Survey of Document Image Word Spotting Techniques." *Pattern Recognition* 68: 310–332.

GUNNENG, HEDDA (1992). "Ductus och skrivarhänder." *Föreläsningar i medeltidsfilologi*. Ed. Jonas Carlquist. Stockholm: Department for Scandinavian Languages, Stockholm University.

HAST, ANDERS & ALICIA FORNÉS (2016). "A Segmentation-Free Handwritten Word Spotting Approach by Relaxed Feature Matching." *12th IAPR Workshop on Document Analysis Systems (DAS), 2016*, 150–155.

HAST, ANDERS & GUSTAF KYLBERG (2015). "Clustering in 2d as a fast Deterministic Alternative to Ransac." *The Second Workshop on Features and Structures (FEAST), collocated with ICML, 2015*.

HAST, ANDERS & EKTA VATS (2017). "Radial Line Fourier Descriptor for Handwritten Text Representation." *Journal of WSCG* 26.1: 31–40.

HE, SHENG, ET AL. (2016). "A Multiple-Label Guided Clustering Algorithm for Historical Document Dating and Localization." *IEEE Transactions on Image Processing 25, 2016,* 5252–5265

LLADÓS, JOSEPH, ET AL. (2012). "On the Influence of Word Representations for Handwritten Word Spotting in Historical Documents." *International Journal of Pattern Recognition and Artificial Intelligence* 26.5.

MÅRTENSSON, LASSE, FREDRIK WAHLBERG & ANDERS BRUN (2015). "Digital Palaeography and the Old Swedish Script. The Quill Feature Method as a Tool for Scribal Attribution." *Arkiv för nordisk filologi* 130: 79–100.

SVENSKT DIPLOMATARIUMS HUVUDKARTOTEK (SDHK). Sweden, Stockholm, Swedish National Archives (Riksarkivet). <https://sok.riksarkivet.se/SDHK> [2016-11-01].

VATS, EKTA, ANDERS HAST & PRASHANT SINGH (2017). "Automatic Document Image Binarization using Bayesian Optimization." *Proc. 4th International Workshop on Historical Document Imaging and Processing,* 89–94, New York: ACM Press.

WAHLBERG, FREDRIK, LASSE MÅRTENSSON & ANDERS BRUN (2014). "Scribal Attribution using a Novel 3-D Quill-Curvature Feature Histogram." *Proceedings of the 14th International Conference on Frontiers in Handwriting Recognition (ICFHR-2014)*, 732–737.

WAHLBERG, FREDRIK, ET AL. (2014). "Spotting Words in Medieval Manuscripts." *Studia Neophilologica* 86: 171–186.

WIKTORSSON, PER-AXEL (2004). "Några sätt att datera medeltida laghandskrifter." *Arkiv för nordisk filologi* 119: 227–244.

WIKTORSSON, PER-AXEL (2015). *Skrivare i det medeltida Sverige. 1–4*. Skara: Skara stiftshistoriska sällskap.

ÅSTRÖM, PATRIK (2003). *Senmedeltida svenska lagböcker:136 lands- och stadslagshandskrifter: Dateringar och dateringsproblem*. Stockholm: Department for Scandinavian Languages, Stockholm University.