LASSE MÅRTENSSON, EKTA VATS
& ANDERS HAST

The Significance of Script Proportions in the Medieval Swedish Charter Script

1 Introduction

Palaeographic features can roughly be divided into two categories. The first of these categories relates to the morphology (form) of the individual script signs (i.e. letters). From this perspective, the shape of the individual script signs is analyzed regarding the chronological development, geographical distribution etc. Many handbooks in palaeography take this perspective, for instance Bernard Bischoff (1990; from a European perspective), Sam Jansson (1943; for Swedish), Didrik Arup Seip (1954; for West Norse), Hreinn Benediktsson (1965; for West Norse), Börje

The research behind this article has been carried out within the project Nya ögon på Sveriges medeltida skrivare (Riksbankens jubileumsfond, NHS14-2068:1).

Mårtensson, L., professor in Scandinavian languages, Uppsala University, Vats, E., researcher in Computerised Image Processing, Uppsala University, & Hast, A., professor in Computerised Image Processing, Uppsala University. "The Significance of Script Proportions in the Medieval Swedish Charter Script". *ANF* 136, pp. 135–162.

Abstract: This article focuses on the extraction, measuring and evaluation of script proportions in the medieval Swedish charter script. The methods for the extraction and the measuring are taken from the field of digital image processing, and the material consists of a selection of charters constituting ground truth regarding the identity of the scribe. In these charters, the scribes explicitly state that he (only males in the material) has written with his own hand. The script proportions in this context mean the relative difference between the script elements of medium height, e.g. minims, and high elements such as ascenders and low elements such as descenders. The results show that the scribes are fairly consequent regarding the script proportions. The charters produced by the same scribe in most cases hold a similar value, even though there are incidental outliers. The earliest scribes of the corpus, active in the earlier part of the 14th century, differ from the later scribes in that they have longer ascenders than the rest. This can be interpreted from a chronological perspective, in that the script may have become more compressed during the later part of the 14th century. Scribes could also vary the script proportions for codicological reasons, if the space demanded that the lines must be placed close to each other. There is also an element of incidental variation, meaning that the scribes actually were not altogether consistent in this respect.

Keywords: Digital palaeography, script proportions, medieval charters.

DOI: https://doi.org/10.63420/anf.v136i.27873

Westlund (2002; for Swedish) and Albert Derolez (2003; from a European perspective). Also in practically oriented investigations in palaeography, the focus is often on the morphology of the separate script signs. In his extensive monograph over the medieval scribes of Sweden, Per-Axel Wiktorsson (2015; see also 2006a and 2006b), identifies the scribes on the basis of the morphology of seven script signs. Nils Dverstorp (2010) also focuses on the scribes' use of certain variants of script signs, and Patrik Åström (2010) puts the Swedish script into a European context on the basis of the framework built up by Derolez and his predecessors. In this article, we refer to such features collectively as *sign-dependent features*.

The other category of features are not related to the separate script signs, but instead they manifest themselves in the script in its entirety, characterizing the script as a whole. Among researchers working on Nordic material, Hedda Gunneng (1992: 24–25) stressed the importance of such features when analyzing late medieval manuscripts. Concluding that the shape of individual script signs was difficult to deal with in quickly executed script, she directed her attention towards aspects that manifest themselves in features appearing in several script signs, such as the difference in proportions between script components of different length (ascenders, minims, descenders; see further below) or the way in which the graphs are attached to each other. In the present article, such features will collectively be referred to as *sign-independent features*.

The research group behind the present article has taken a great interest in investigating sign-independent palaeographic features generally, and in the present study, one such specific feature has been extracted, measured and analyzed. The empirical material comprises a set of medieval Swedish charters. Sign-independent features are often difficult to measure and quantify, at least on a large corpus and with a high degree of exactness. In order to overcome these obstacles, we have made use of methods from digital image processing, described further below. We have focused on the proportions between script components of different height, such as high components (e.g. ascenders), middle height components (e.g. minims) and low components (e.g. descenders).[1]

[1] A pilot study to this investigation was presented at the conference Digital Humanities in the Nordic Countries (Mårtensson, Hast & Vats 2018). That investigation was carried out in order to try this methodology. The work in the present article is based on a refined dataset and an improved and more precise method.

2 Aim

The aim of this article can be formulated in the following points:

1. To present a methodology for extracting the proportions between script elements of middle height, such as minims and loops on the baseline, and high and low script elements, such as ascenders and descenders respectively, in the medieval Swedish charter script. The methodology is carried out on a limited dataset, selected for the purpose of evaluating the extracted features from the perspective of scribal individuality.
2. The feature of script proportions will then be evaluated and analyzed regarding the relevance from an individual perspective, i.e. to what degree scribes vary or are consistent in this aspect, and to what degree the scribes vary from each other. This will lead to preliminary conclusions as to what significance this feature carries for instance in investigations in scribal attribution.

Two issues are especially addressed in this article. Firstly, we are making use of a computer-aided methodology. These methods allow for a very exact measuring process on a feature that is difficult to measure manually.

The second issue that has been given special attention is the evaluation of the extracted data. One major obstacle when analyzing individual variation in the medieval script is that a large amount of the medieval handwritten material is anonymous. In many cases when a certain scribe is assumed to have written a specific piece of writing, this assumption is based on an attribution, in its turn based on a given set of criteria (palaeography, orthography, language). To use previous attributions for evaluation of individually distinctive script features is unsatisfactory, as there is usually a degree of uncertainty in those. In the present investigation, we have created a corpus, a set of charters, where the scribes are known from external factors. This corpus thus constitutes ground truth regarding the identity of the scribes (see further section 5).

It should be noted that only charter script has been analyzed in the present study. Furthermore, the charters in the corpus do not cover the entire medieval period. The earliest charter dates from 1318 and the latest from 1458, which means that the corpus covers 140 years. The script type represented in the corpus is *Cursiva*. However, the aspect of script type will not be addressed further.

3 Previous Research

As stated above, there has been an interest in sign-independent script proportions for the purpose of scribal attribution, see e.g. Gunneng (1992) mentioned above. This aspect has been in focus in other investigations as well, by e.g. Peter Stokes (2014, on English script; see e.g. p. 79). For the Gothic script, Derolez (2003, e.g. p. 58) puts focus on script proportions, and states that the body-height (cf. *core* in the present study) of the script signs became larger in relation to the ascenders and the descenders in the Pregothic script as compared to the predecessor, the Carolingian minuscule. (Regarding the terms *core*, *ascender* and *descender*, see section 4.1.) However, to our knowledge, the method proposed in this investigation has not been used for the present purpose, and certainly not on the medieval Swedish charter script.

Another important aspect of this investigation is the development over time within one scribal hand regarding this specific script feature, as we are investigating documents written by the same scribe at different points in time. Investigations on scribes' change over time have of course been carried out before, e.g. by Dverstorp (2005), Stefán Karlsson (1964; on Haukr Erlendsson) and others. Focus has usually been on the shape of script signs, i.e. how the scribes have changed from one form of a certain script sign to another. Less attention has been given to sign-independent features, and here less is known about the change versus consistency.

From a methodological perspective, the present study is inspired by the work of Shilpa Pandey and Gaurav Harit (2017). Their investigation focused on separating the core of handwritten script from the ascenders and descenders, and their material was quite diversified handwritten script types. They used only modern handwriting, whereas we extend this approach to medieval script. Even though the medieval Gothic script, produced in Sweden, certainly can be heterogeneous and carried out on different levels of execution, it is still more homogenous than the script analyzed by Pandey and Harit. On the other hand, the medieval charters investigated by us present a greater challenge in the fact that the documents themselves are worn by time, and can be damaged in different ways.

4 Method

In the section on the methodology, both the palaeographic and the computational perspectives will be addressed, but focus will be on the former. Regarding the computational components, a brief description will be given in section 4.2, and additional information is placed in the Appendix.

4.1 Palaeographic Perspective

By script proportions in this context, we mean the proportions between components of the script being of various height, for instance ascenders versus minims versus descenders, as stated above. Thus, the script proportions are partly represented in the relation between graphs, not only in the graphs themselves. In some cases, the graphs themselves manifest such proportions, i.e. when they consist of a high/low component and a component of medium height. Examples of such graphs are 'b', 'h' and 'k', and 'g', 'p' and 'q'. Others do not normally contain such differentiation, i.e. 'i', 'm', 'n' and 'l'. In the latter cases, the proportions are only manifested when the graphs are combined.

Alphabetic minuscules can, from a vertical perspective, be divided into three fields, one upper, one middle and one lower, and the fields are demarcated by four lines. This three-levelled division, often appearing in graphonomic contexts (e.g. Allén 1965: 80–81 and numerous studies after him), often functions as a basis for classifying script signs. Allén (1965: 83) classifies minuscules as scripts signs contained within the middle field, extending into the lower field by a descender ('stapel), and into the higher section by an ascender (also 'stapel') and/or a diacritic sign. For the present investigation, however, the three fields are a basis of description rather than of classification. Components can reach into the upper or the lower fields to a various degree, and it is not only a matter of either/or, as is often the case when classifying script signs into different graph-types etc. The middle field, demarcating the core of the script signs, is the one that contains script elements of medium size, such as minims and loops in 'o', and in combination with ascenders or descenders in 'b', 'd', 'p' and 'q'. In figure 1, the fields are exemplified with the sequence 'frændom ok maghom' ('relatives and relatives by marriage', in the dative).

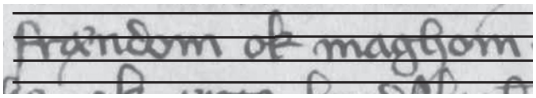

Figure 1. Illustration of the three fields, demarcated by four lines.

In the figure, the fields are illustrated from an ideal perspective, and the lines demarcating the fields are completely straight. In handwritten script, the lines are not straight if they are to follow the contours of the script. Furthermore, when taking a descriptive perspective, it is clear that the script signs are not either of middle height or high or low, but an aspect of degree must be added. In the figure, 'f', 'æ', 'd', 'k' and 'h' all extend into the upper field, but they do it to a varying degree. The 'k' reaches highest, and then 'f', 'd' and 'h' come close to each other, and 'æ', finally, only marginally extends over the core. Also the script signs within the core vary. As can be seen e.g. in the sequence 'maghom', the minims constituting 'm' are slightly lower than 'a'. Lines actually following the script contours of the core would thus be uneven, and the degree of unevenness would be dependent on the level of execution of script.

The method used in the present investigation identifies the core of the script, i.e. the components that are contained within the middle field. Once the core has been identified, the components reaching outside of the core, above and below, are registered. This is illustrated in figure 2.

In the figure, the core is marked by a section of darker blue. Components reaching above the core are marked with yellow, and those reaching below the core are in green. It should be noted that there are components outside the core that are in light blue. These are components that have not been measured, and this is mainly due to two reasons. The first of these is when a descender reaches down and touches an ascender from the line below. In such cases, the algorithm cannot separate the two components, and as a result, they are left unmeasured, in order to reduce the erroneous values. The second reason is if there are components above the core that are not attached to this, such as accents or other diacritic signs. These are not measured either. In some cases, components above the core are connected with the core with a very thin stroke, and they are also sometimes left unmeasured. It should also be noted that no distinction is made between the different types of components reaching above the core; ascenders, the top part of high 's', the top part of 'æ' and so forth have all been marked.

Figure 2. Image of the core as identified by the computer through Gaussian filtering (section 4.2) and connected components reaching outside the core.

The charters begin with an initial up to the left, usually having the height of a number of lines. These initials are so large that the algorithm does not identify them as being a part of the core, and thus they are not accounted for in the measuring process. Consequently, they do not affect the result.

As stated, both the ascenders and other high elements and the descenders and other low elements are measured. One could assume, however, that the consistency in length in the high elements is higher than in the low elements. Most strokes are performed downwards, and it is likely that the beginning of the movement is more controlled than the end of the movement. In the evaluation, focus will therefore be on the high elements, hereafter collectively referred to as ascenders.

From a methodological perspective, one has the choice of aiming the measuring process on individual words or on a charter in its entirety. In the present investigation, we have chosen the approach of measuring the charters in their entirety. The two approaches present different advantages and different problems. When measuring a charter in its entirety, different components reaching outside the core and into the upper or the lower fields are measured collectively, and measures of e.g. the regular ascenders and the components of 'f' and high 's', reaching marginally into the upper area, are given a joint value. The same goes for the components reaching below the core. The advantage is that the measuring values account for the entire written record, and not a selection of letter sequences. Approaches focusing on individual words have the advantage of each measuring being fully comparable with the other. The disadvantage is that the amount of data constituting the basis of the value for each document becomes much smaller, and the incidental variation in each word being measured will play a greater role. As we are working with charters, each document con-

tains a limited set of words, and also highly frequent words would only render a limited set of measuring points in each document. Furthermore, as the charters are in both Swedish and Latin, we would need separate words for each language form, which in itself would reduce the comparability. Thus, both approaches have their strengths and weaknesses, but we have judged the approach of a collective value for an entire charter as the most suitable for our material.

4.2 Computational Perspective

In this section, a brief description of the computational aspects of this investigation is given. A more detailed description is given in the Appendix. The general base of the methodology used in the investigation is taken from the broad field of Handwritten Text Recognition (HTR), which collectively refers to computational methods for the analysis of handwritten material. As stated in the previous section, the proposed method focuses on the identification and measuring of script components of various heights.

Before the measuring process is started, the images are pre-processed in two steps in order to facilitate the computational work, and in order to increase the accuracy. Both processes serve to clean the script surface in the images of disturbances in the form of damages and similar. Firstly, a so-called shading correction is performed (Hast and Marchetti 2014) on the charter images to be investigated. This means that variation in brightness on the script surface is reduced, e.g. if the image for some reason is shaded or if the parchment or paper varies in brightness. Secondly, damages, blots etc. in the script surface are removed using a so-called two band-pass filtering approach (Vats, Hast and Singh 2017), one of high frequency and one of low frequency. A high frequency band-pass filter separates the text from the background, and a low frequency band-pass filter is used for the removal of damages. The background removal is performed in such a way that the gray-level information crucial for the feature extraction is not affected. This reduces the risk of other features of the written records (i.e. features not belonging to the script) affecting the measured values.

Once the pre-processing is ready, the algorithm automatically estimates the size of the script proportions by applying a so-called Gaussian filter (also known as a filtering kernel) over the whole image. This process generates an easily measurable template surrounding the core region (i.e. a core template), which is used to infer the average size of the core, the

ascenders and the descenders. The size of the ascenders and descenders is computed by measuring the height of the region outside the defined core template. In a graphical form, a Gaussian kernel will appear as a bell-shaped curve. Furthermore, the number of lines is computed by measuring the resulting core template in the image. The distance between the lines is computed in a similar way as for the core template.

5 Material

The present investigation focuses on the measuring and evaluation of the script proportions from an individual perspective, but it is not an investigation in scribal attribution in itself. We want to evaluate the individual relevance of this specific script feature, and this can in the future be used for the purpose of scribal attribution. However, for the evaluation, we need ground truth on scribal identity, where the identity can be affirmed by external evidence. Furthermore, we wanted to investigate the constancy versus variability of the script proportions in a scribal hand over time, as represented in documents produced with time intervals in-between.

For this purpose, we have collected charters in the SDHK collection (Svenskt Diplomatariums huvudkartotek) containing a so-called notarial text, in which it is explicitly stated who has held the pen and produced the script. A notarial text is a declaration, placed below the main text of the charter, of the scribe's name, diocese, etc., and an affirmation that he has taken part in the proceedings in question and that he has written with his own hand, or that he has given the task to another. The notarial texts are also accompanied by a notarial sign, unique for this individual, often consisting of a variation of a cross (Nielsen 1967: 364). The notarial texts are always in Latin, but the main texts are in both Latin and Swedish in our material. See the specification of this under table 1.

Examples of a notarial text and a notarial sign are given in figure 3 and in figure 4 respectively. These examples are taken from the charter SDHK 26456, by the scribe Anders Petersson.

In the notarial text, the scribe presents himself (only males in our material) with *Ego N.N.*, and then he states that he has written with his own hand, *propria manu scripsi/scriptum* or similar. In some cases, the scribe in question has only produced the notarial text, and another scribe has produced the main text. This is formulated in different ways, but two

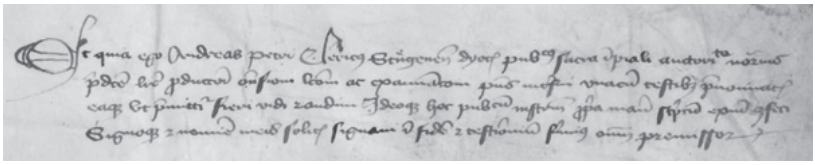

Figure 3. An example of a notarial text, taken from SDHK 26456, produced by Anders Petersson.

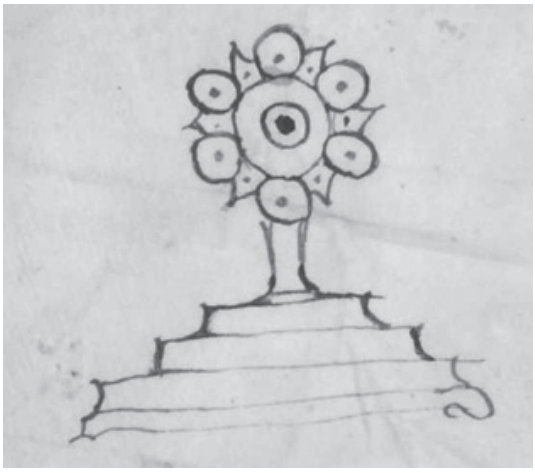

Figure 4. An example of a notarial sign, taken from SDHK 26456, produced by Anders Petersson.

examples are *manu alterius scriptum* or *per alium scribi feci*, and sometimes the name of the other scribe is given. Such charters have not been used in our investigation. Thus, we have searched for charters according to the following criteria:

- The charters must contain a notarial text, stating that the scribe of the notarial text has also produced the main text.
- The scribes must have produced two charters as a minimum filling the criteria of point 1 above.

The search for self-attributed charters filling the criteria given above has given a total of 34 charters.[2] Eight scribes are represented, and in table 1 the scribes and the charters are listed chronologically. The numbers given to them by Per-Axel Wiktorsson (2015) are included. The charters span

[2] We want to thank Claes Gejrot at Diplomatarium Suecanum for help in this work.

from 1318 (earliest charters by Arnvid Johansson) to 1458 (last charter by Anders Petersson).

We decided to measure only the selected corpus of charters, and not to include a large amount of charters apart from them. The reason for this is that the measuring process also incorporates an element of trial and evaluation, and the evaluation is only possible with a limited selection.

Table 1. The charters with notarial text, stating that the scribe has produced the entire charter.

Arnvid Johansson (Wiktorsson's nr 36)		
SDHK	Date	Place
2863	1318 1/4	Julita
2903	1318 1/12	Uppsala
2904	1318 1/12	Uppsala
3560	1328 18/9	Uppsala
Joar Nilsson (Wiktorsson's nr 222)		
SDHK	Date	Place
5148	1345 30/3	Uppsala
5150	1345 1/4	Uppsala
5174	1345 9/5	Barkeryd
5175	1345 9/5	Barkeryd
Lars Mattsson (Wiktorsson's nr 409)		
SDHK	Date	Place
8919	1366 22/7	Uppsala
8941	1366 14/8	Uppsala
8994	1366 19/11	Stockholm
9328	1368 18/8	Karby
9525	1369 14/9	Uppsala
Olof Nilsson (Wiktorsson's nr 575)		
SDHK	Date	Place
8947	1366 28/8	Uppsala
9346	1368 29/9	Uppsala
9391	1369 5/1	Uppsala
11058	1377 9/7	Uppsala

Lars Finvidsson (Wiktorsson's nr 393)		
SDHK	Date	Place
15960	1402 10/11	Without place [Vadstena]
16001	1403 14/1	Without place
17654	1411 5/6	Skänninge
Peter Tidikesson (Wiktorsson's nr 653)		
SDHK	Date	Place
21861	1432 10/11	The sacristy in the church of Enköpings-Näs
21866	1432 18/11	Uppsala
22154	1434 20/5	Uppsala
24733	1445 30/4	Uppsala
Peter Svensson (Wiktorsson's nr 651)		
SDHK	Date	Place
21982	1433 4/8	Vadstena
23055	1438 3/9	Linköping
23881	1441 17/7	Kalmar
26344	1453 23/10	Vadstena
Anders Petersson (Wiktorsson's nr 36)		
SDHK	Date	Place
26318	1453 27/8	Strängnäs
26456	1454 25/2	Strängnäs
26457	1454 25/2	[Strängnäs]
26464	1454 4/3	Strängnäs
27171	1458 15/2	Strängnäs

Remarks:

- According to the SDHK database, SDHK 23055 (Peter Svensson) contains both Swedish and Latin. This is correct, but to this category (i.e. charters containing both Swedish and Latin) can also be added SDHK 21861 and 22154 by Peter Tidikesson, SDHK 23055 by Peter Svensson, SDHK 26464, 26457 and 27171 by Anders Petersson. They are all classified as Latin in the SDHK database.
- Arnvid Johansson has also produced the charter SDHK 2880, containing a notarial text stating that he has written in his own hand, both the main text and the notarial text. However, this charter is only preserved in a copy.

- The charters SDHK 8989, 8992, 10584, 10559 and 40536 also have a notarial text stating that Olof Nilsson has written in his own hand, but they are only preserved in copies.
- The charter SDHK16645 also has a notarial text stating that Lars Finvidsson has written in his own hand, but this is only preserved in a copy.

6 Results and Discussion

In the present section, the results of the investigation will be illustrated, discussed and evaluated. Firstly, a plot will be presented with the measurements extracted from the charters. In the plot, the charters by the different scribes have been given different colours and/or different forms.

The dimensions in the plot are the proportions between core and ascenders (horizontal) and between ascenders and descenders (vertical). This means that the ascenders are in focus in the evaluation, as stated above. The position of 1 in the horizontal dimension marks the point

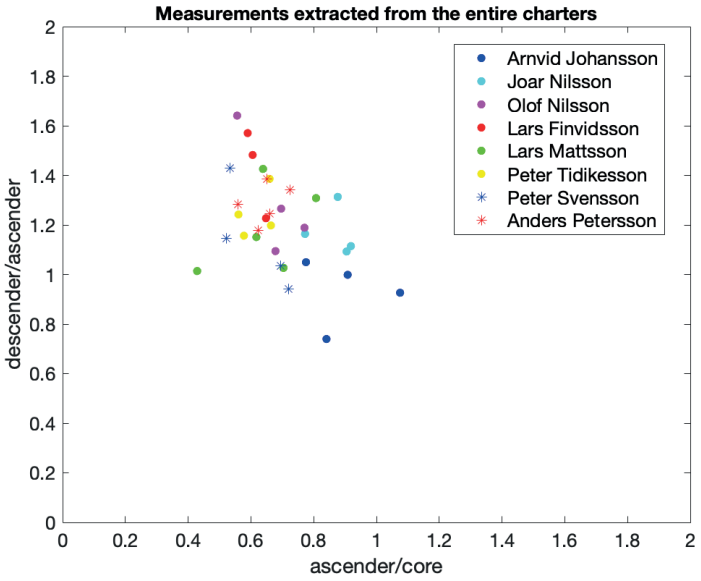

Figure 5. Plot over the measured values extracted from the entire charters.

where the approximate core size is the same as the size of the ascender. Only one charter has a higher value than 1 (by Arnvid Johansson), and this means that the core in all other cases is larger than the ascenders. The position 1 in the vertical dimension marks the point where the descenders are of the same length as the ascenders, and a higher value indicates that the descenders are longer than the ascenders. The majority of the charters has a value above 1, meaning that the descenders in most cases are larger than the ascenders. It must be stressed, though, that these values are average values based on the charters in their entirety. Incidental examples of ascenders/descenders can certainly be found that breaks this pattern.

As can be seen, all the charters constitute one cluster from an overall perspective, with a few outliers. One conclusion that can be drawn from this is that this measure alone does not suffice for separating scribal hands from each other. However, the charters produced by the same scribe are generally regularly distributed within this cluster, suggesting a degree of consistency regarding this feature within the hands.

The charters by Joar Nilsson, Peter Tidikesson and Anders Petersson are fairly well collected according to the scribe, and the charters by Olof Nilsson are also positioned close to each other apart from one outlier. The charters by Arnvid Johansson are actually rather spread out, but they still hold a particular space in the plot, namely down to the right. The charters by Anders Petersson are also positioned close to each other. It should be noted, though, that one charter, 26464, is severely damaged, but it appears that the damaged parts have been excluded in the measuring process, and that only the preserved script has been measured. The charters by Lars Finvidsson have a similar value in the horizontal dimension but they vary in the vertical dimension.

The charters by Lars Mattsson and Peter Svensson are more scattered. Seen only from the horizontal dimension, the charters by Peter Svensson have a fairly similar value, and other scribes vary equally much in this dimension, see e.g. Arnvid Johansson. However, in combination with different values in the vertical dimension, they give a scattered impression. Lars Mattsson's charters are spread out in both dimensions, and they are positioned in the middle of the cluster, forming no clear pattern.

In the following, the tendencies observable in the results will be discussed under separate headings.

6.1 Chronological Variation

It should be noted that the entire corpus comprises a continuum in the plot, and no clusters are clearly separated from each other. Still, two scribes have their main distribution to the right in the horizontal dimension (the length of the ascenders related to the core), namely Arnvid Johansson and Joar Nilsson. According to these values, these two scribes differ from the others in that they produce notably long ascenders in relation to the core. They are also the earliest scribes in the corpus, having written their charters 1318–1328 (Arnvid Johansson) and 1345 (all four charters; Joar Nilsson) respectively. In figure 6, the charters of Arnvid Johansson and Joar Nilsson are coloured in red and green respectively, and the other charters are in dark blue.

From a horizontal perspective, the charters of Arnvid Nilsson and Joar Nilsson form a section of their own, with only two other charters being positioned in the same section. Of these two charters, Olof Nilsson has produced one and Lars Mattsson the other. Olof Nilsson is also an

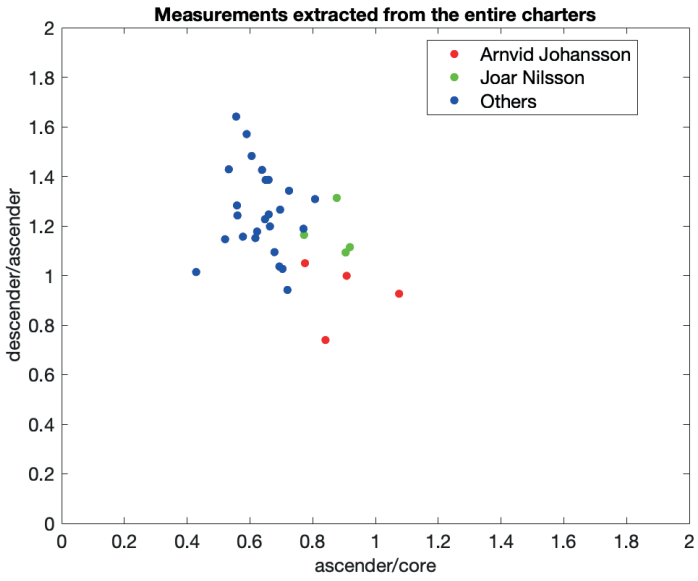

Figure 6. The charters of Arnvid Johansson and Joar Nilsson as compared to the other investigated charters.

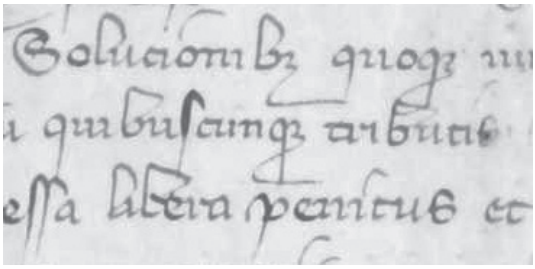

Figure 7. Detail from Arnvid Johansson.

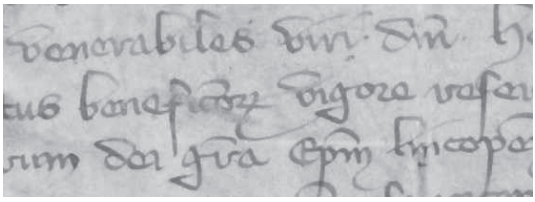

Figure 8. Detail from Joar Nilsson.

early scribe, although not as early as Arnvid Johansson and Joar Nilsson. However, the other charters of Olof Nilsson are not distinguished by high values in the horizontal dimension, so this charter should probably not be explained by chronology. The charters by Lars Mattsson are generally scattered, as already noted.

Upon a check in the charters produced by Arnvid Johansson and Joar Nilsson, this divergence in proportions appears to be correct. In figure 7, a detail from Arnvid Johansson is given, and in figure 8 is a detail from Joar Nilsson.

With the bare eye, the difference in proportions between the script produced by Arnvid Johansson and Joar Nilsson as compared to the other scribes stands out as correct. This difference is not surprising. In the pilot investigation mentioned in footnote 1 above, charters from the 12th and the early 13th centuries were measured (not only Swedish ones), and these were strongly divergent from the later charters in terms of the script proportions, e.g. SDHK 2253 (produced in 1308 in Poitiers). A detail from this charter is given in figure 9.

As mentioned above, the compactness of the Gothic script as compared to the predecessor, the Carolingian minuscule, is well known (Derolez 2003: 58), at least as manifested in book script. It is likely that a gradual compressing development also took place within the Gothic script. The results of our investigations point in that direction. However, for a clearer

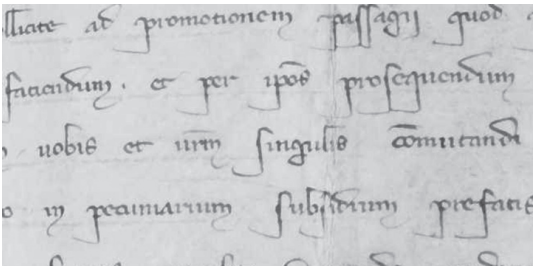

Figure 9. Detail from SDHK 2253.

picture of the chronological development to emerge, another selection of data must be investigated with the present method. Still, the values of Arnvid Johansson and Joar Nilsson indicate that an investigation of chronology would be a fruitful task for future research. If these two scribes are representative for their time period, the results of the investigation would indicate that a change of the script proportions, especially regarding the length of the ascenders, took place around the middle of the 14th century.

Apart from the two scribes from the first half of the 14th century, no clear chronologically conditioned variation can be observed. The latest scribe, Anders Petersson, with charters from 1453–1458, does not diverge markedly from the other scribes.

As stated, a general chronological difference appears to be manifested between the two earliest scribes of the corpus and the rest of the scribes. On an individual level, however, no chronological variation can be observed, e.g. manifested in differences between the earliest and the latest charters by the same scribe (i.e. individual chronology). Several scribes have charters with divergent values, but the deviations cannot in any of the cases be explained by chronology. Thus, this investigation does not support the idea of change over time within a scribal hand regarding the script proportions.

6.2 Codicological Variation

As stated in the previous section, the charters of Arnvid Johansson display significant differences in proportions between core and ascenders. However, within the charters produced by him, there is a marked difference between SDHK 2904 (produced in 1318) and SDHK 2903 and 3560 (produced in 1318 and 1328 respectively). The question is how this

difference can be explained. It appears that there is no general tendency of individual change over time in the investigated material, and it is not likely that these incidental examples should be interpreted in that way. Can another explanation be found, or is this just a case of general variation within one hand?

First, the charters 2904 and 3560 will be contrasted, as they differ in both the horizontal and the vertical dimension. Upon a closer scrutiny, it appears that the two charters differ somewhat with respect to the compactness of the script surface, i.e. the space between the lines. SDHK 2904 gives a more compact impression than SDHK 3560, manifested in the text lines being closer to each other in the latter than in the former.

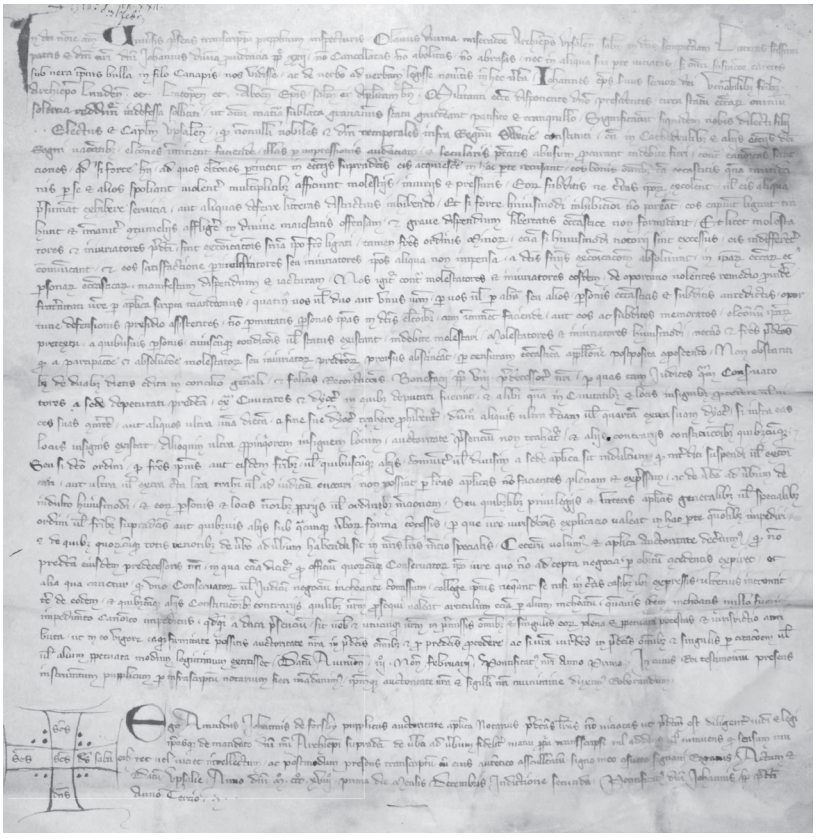

Figure 10. An image of SDHK 2904.

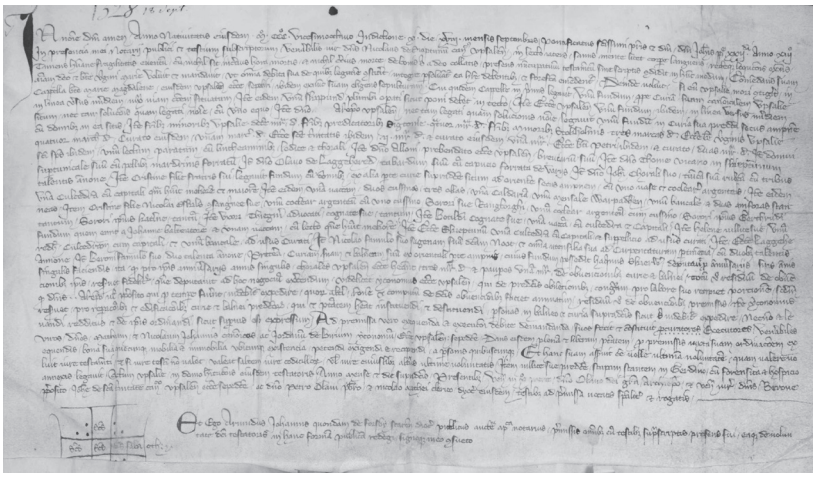

Figure 11. An image of SDHK 3560.

If this impression is correct, the formation of the script surface could explain the difference between the two charters. As stated in section 4.2, a side effect of the method used in this investigation was that the space between the lines was measured. This means that we have the possibility to quantify the space between the lines. This measurement is based on the number of pixels in the images, which means the quality of the images may affect the values to some extent. Still, this should work as a rough measure, as the images in this case are of approximately the same size. It turns out that the average distance between the lines is higher in SDHK 2904 than in 3560, with 73 versus 51 pixels. This supports the idea of higher compactness of the script surface in 3560 than in 2904. The two charters are shown in figure 10 and figure 11.

The space for the ascenders and the descenders thus appears to be larger in SDHK 2904, and probably Arnvid Johansson has used the space for creating longer ascenders. This indicates that the formal layout of the script surface in terms of space between the lines affects the script proportions within a scribal hand. The interpretation of charter 2903 is less clear. The value is indeed lower in SDHK 2903 (61 pixels) than in 2904, suggesting a more compact script surface in the former charter, but not to the same extent as in 3560. The compactness of the script surface is thus only one factor among several affecting the script proportions. This aspect can probably, at least partly, explain the position of charter 2904

far to the right in the plot, but it is not the main factor determining the script proportions.

6.3 Incidental Variation

Another explanation must be taken into account regarding the variation between the charters produced by the same hand, namely the incidental variation. With the bare eye it is possible to discern differences in proportions in sequences of script signs in one and the same document, and the results of the present investigation also point in that direction. The consistency with which at least the investigated scribes upheld the script proportions appears not to have been complete, and one can assume that this is true also in general. One important conclusion to be drawn from this from a methodological perspective is that the measured value must be based on a larger text mass, in order to reduce the effect of the incidental variation.

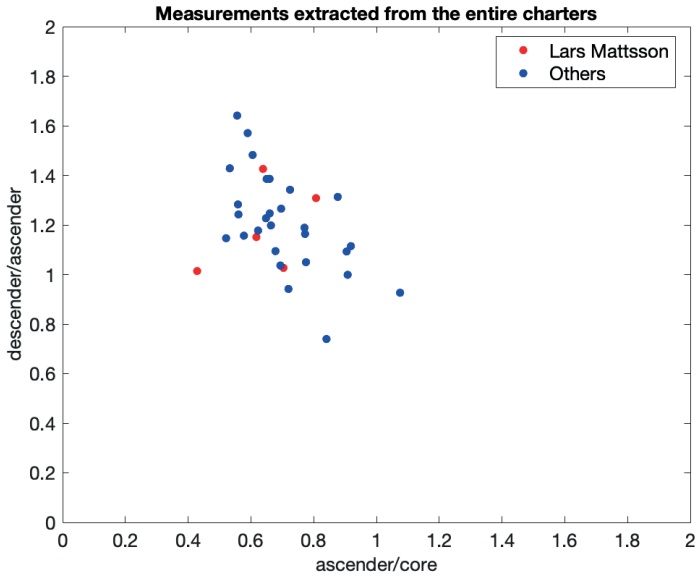

Figure 12. The Charters of Lars Mattsson as compared to the other investigated charters.

One scribe that is distinguished by variation is Lars Mattsson. He has produced five charters, and they are all distributed irregularly within the plot. This is shown in figure 12.

Lars Mattson produced his five charters between 1366–1369, and he is a rather early scribe in the corpus. Still, four of his charters are spread out among the others in the middle of the cluster, with one outlier, SDHK 9525, positioned outside the cluster. This charter has a damage, but it appears that the used method at least in some cases has managed to deal with damages. The charters are relatively similar from a codicological perspective, except for SDHK 8999, which is slightly deviant from the rest. It contains a larger number of lines than the other charters, and furthermore lines with a larger number of words than the other charters. It should be noted though that the codicologically deviant charter is not placed as an outlier in the plot.

Lars Mattsson is distinctly different from the two earliest scribes, Arnvid Johansson and Joar Nilsson. The results from Lars Mattsson's charters suggest that the incidental variation can be large regarding the script proportions. It is not altogether easy to narrow down the exact reasons for this variation. Is it due to the scribe's actually varying the script proportions, or is it the characteristics of the charters that may vary, for instance in that a charter may contain more high elements than another? This investigation does not have the data to answer this question, but it should be noted that most scribes' charters are more closely gathered than Lars Mattsson's. However, as the investigated corpus is rather small, it is difficult to draw conclusions about the individual regularity in this aspect. It might be that the variability of Lars Mattsson is the normal distribution, whereas the regularity of the other scribes may be deviant.

6.4 General Remarks

It should be stressed that the evaluation of the feature in question has been carried out on ground truth regarding scribal identity. Furthermore, the correctness of the measurements has been controlled, indicating that the computational method has worked well. The fact that the charters produced by the same hands usually have similar values, and thus are positioned close to each other in the plot, suggests that this feature carries relevance from an individual perspective. This feature in itself cannot, at least not according to the methodology used here, separate scribal hands from each other, but the feature is to a large extent regular within a hand. However, in many cases the results of the measurements are difficult to

observe with the bare eye. In some cases, the results are visible, such as the length of the ascenders in the script of Arnvid Johansson and Joar Nilsson respectively and the greater compactness of the script surface in SDHK 3560 than in 2904, but sometimes it is more difficult to catch the difference with the eye, for instance the diversity of Lars Mattsson's charters. To some extent, the computer seems to register similarities and differences in the script of the charters that the human eye cannot easily see. In the computational measurements, there is a very large number of measure points, and to illustrate this with a few chosen examples from the charters becomes haphazard, especially as the incidental variation appears to have been considerable.

7 Conclusion

The purpose here was to investigate the individual relevance of this script feature in charter script, evaluated on ground truth regarding scribal identity. As a consequence, the dataset is limited, and it allows only for tentative conclusions. The advantage of this study, however, is that the feature is evaluated only on the charters that with certainty were produced by a specific scribe. The sparse material with explicit information about the scribe is a problem when dealing with charters, and actually medieval script generally. The charters constituting the material of the present investigation are thus an important corpus, even though they are few.

The results of the current investigation suggest that the script proportions actually carry relevance from an individual perspective. The charters produced by the separate scribes come close to each other in the measured values, even though several scribes have produced at least one charter that is deviant from the rest. One important conclusion to be drawn from this investigation is that the significance of this method lies in the quantity of the examples, not in the separate values for each document. Separate examples, e.g. one separate document, can be misleading, whereas the pattern given by the mass of examples gives correct tendencies. Thus, to use the current method for identifying a scribe in one document, consisting of one page, would not be possible, and as shown, charters produced by different hands can display very similar values. But the measure gained from this investigation can be used as one feature of several to characterize the script of a given document.

The investigated dataset has not been chosen in order to investigate the significance of the script proportions from a chronological perspective. Still, the results indicate a general change in the script proportions around the middle of the 14th century, towards a more compressed script, with relatively shorter ascenders as opposed to the core. This is not surprising, considering the script used in e.g. charters from the early 13th century, where this difference is much more marked than in documents from the later part of the Middle Ages. Still, another dataset would have to be investigated with the current method in order for safer conclusions regarding the chronological development to be drawn.

It seems that variation in the script proportions within one hand can vary due to codicological reasons, i.e. when the scribe needs to compress the text and place the text lines close to each other. There is one example pointing to this, and it is indeed not surprising, as a compressed text surface does not allow for high ascenders or low descenders. Still, this is one aspect that must be accounted for when analyzing the results of an investigation of the current feature.

The method works generally better with larger amounts of text. This is of course expected, as the effect of the incidental variation becomes more neutralized in a longer text than in a short one. When observing a piece of handwritten medieval text, it is clear that the variation can be considerable, especially if the script has been executed quickly. This variation was used by e.g. Gunneng (1992), as mentioned above, as an argument against the criteria morphology when analyzing late medieval script. It seems that the incidental variation in the proportions of the script is also considerable.

One final issue will be addressed, which is of relevance for future digitization projects on handwritten corpora. For investigations like the present one, the quality of the images is of utmost importance. In order to meet as many research purposes as possible with the digitization efforts, the process should be carried out in collaboration with expertise within digital image processing, in order to secure the usability of the material for such research tasks. Indeed, the possibilities presented by the digital image processing used on handwritten material reach much further than the field of the present investigation, i.e. computer aided palaeography. Through methods from the digital image processing, unedited material can be searched, e.g. by means of Word Spotting (see e.g. Wahlberg et al. 2014), and the format and the layout of the written records can be measured and mapped with great precision and on large quantities of material (see e.g. Binmakhashen and Mahmoud 2019), just to mention a few examples. The possibilities are great, but in order to accomplish

such research tasks, the quality of the images must meet the demands of the methods.

Bibliography

Allén, S. (1965): *Grafematisk analys som grundval för textedering med särskild hänsyn till Johan Ekeblads brev till brodern Claes Ekeblad 1639–1655*, Almqvist & Wiksell, Göteborg.

Binmakhashen, G. M., & S. A. Mahmoud (2019): "Document Layout Analysis: A Comprehensive Survey". *ACM Computing Surveys*, Vol. 52, No. 6. Available at: https://doi.org/10.1145/3355610.

Bischoff, B. (1990*): Latin palaeography. Antiquity and the Middle Ages*, transl. by Daíbhí Ó Cróinín & David Ganz, Cambridge University Press, Cambridge.

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

Dverstorp, N. (2005): "Från g1 till g2. Om paleografisk förändring hos medeltida skrivare, tillämpat på handskrifterna Cod. Holm. A 108 och Cod. Holm. B 185". C. Falk & L.-O. Delsing (eds.), *Studier i svensk språkhistoria 8*, Lunds universitet, Lund, pp. 93–101.

— (2010): *Skrivaren och skriften. Om skrift- och handskriftproduktion i Vadstena kloster*, Universitetet i Oslo, Oslo.

Gunneng, H. (1992): "Ductus och skrivarhänder". J. Carlquist (red.), *Föreläsningar i medeltidsfilologi*, Stockholms universitet, Stockholm.

Hast, A. and A. Marchetti (2014): "Improved Illumination Correction that Preserves Medium Sized Objects". *Machine GRAPHICS and VISION*, vol. 23, no.1/2, pp. 3–20.

H. Benediktsson (1965): *Early Icelandic script. As illustrated in vernacular texts from the twelfth and thirteenth centuries*, The Manuscript Institute of Iceland, Reykjavík.

Jansson, S. (1943): "Svensk paleografi". J. Brøndum-Nielsen (red.), *Palæografi*, Bonnier, Stockholm, pp. 82–134.

Mårtensson, L., A. Hast & E. Vats (2018): "Extracting Script Features from a Large Corpus of Handwritten Documents". Presentation given at the conference Digital Humanities in the Nordic Countries, 3rd Conference, 7–9 March 2018, Helsinki.

Nielsen, H. (1967): "Notar". *Kulturhistorisk leksikon for nordisk middelalder* 12, Rosenkilde og Bagger, København, pp. 363–365.

Seip, D. A. (1954): *Palæografi B. Norge og Island*, Bonnier, Stockholm.

S. Karlsson (1964): "Aldur Hauksbókar", *Fróðskaparrit: Annales Societatis Scientiarum Færoensis* 13, Tórshavn, pp. 114–121. (Also printed in *Stafkrókar*, Rit Stofnunar Árna Magnússonar á Íslandi 49, Reykjavík 2000.)

Stokes, P. (2014): *English Vernacular Minuscule from Æthelred to Cnut circa 990 – circa 1035*, D. S. Brewer, Cambridge.

Wahlberg, F., M. Dahllöf, L. Mårtensson & A. Brun (2014): "Spotting Words in Medieval Manuscripts". *Studia Neophilologica* 86/2014, pp. 1–16. (Also available at: http://www.tandfonline.com/doi/abs/10.1080/00393274.2013.871975#.VL0THGSG95U)

Vats, E., A. Hast & P. Singh (2017): "Automatic Document Image Binarization using Bayesian Optimization". *The 4th International Workshop on Historical Document Imaging and Processing* (*HIP 2017*), in conjunction with *ICDAR 2017*, full paper, pp. 1–6.

Westlund, B. (2002): "The development of Latin script III: in Sweden". Oskar Bandle et al. (eds.), *The Nordic Languages. An International Handbook of the History of the North Germanic Languages* 1, Valter de Gruyter, Berlin/New York, pp. 841–849.

Wiktorsson, P.-A. (2006a): *Skrivare i det medeltida Skara stift*, Skara stiftshistoriska sällskap, Skara.

— (2006b): "Namn, autografer och handstilar. Till identifieringen av skrivare i medeltidens Värmland och Dalsland". L. Peterson, S. Strandberg & H. Williams (red.), *Namn och runor. Uppsalastudier i onomastik och runologi till Lennart Elmevik på 70-årsdagen 2 februari 2006*, Uppsala universitet, Uppsala, pp. 259–266.

— (2015): *Skrivare i det medeltida Sverige* 1–4, Skara stiftshistoriska sällskap, Skara.

Åström, P. (2010): "Myten om Vadstena-stilen". *Arkiv för nordisk filologi* 125, pp. 99–125.

Appendix. A Description of the Technical Process

The technical process is divided into two main stages: pre-processing and automatic size estimation, which will be discussed here according to this division. Some sections in this Appendix are similar to those of the main text section 4.2. One additional reference is added here, and it is placed in the end of this Appendix.

The SDHK charters often contain some disturbances in the script surface such as paper stains, faded ink, variation in brightness in the script surface (hereafter *shading*) and ink bleed-through, which is an obstacle for the task of automatic size estimation. Therefore, pre-processing of charters is performed by applying shading correction (Hast and Marchetti 2014), where the black pixels are filled with the mean of the image. Background noise removal is carried out using a two band-pass filtering approach (Vats, Hast and Singh 2017). For example, figure 13 presents a sample input image of a charter, and figure 14 presents the corresponding pre-processed image, with gray-level information preserved.

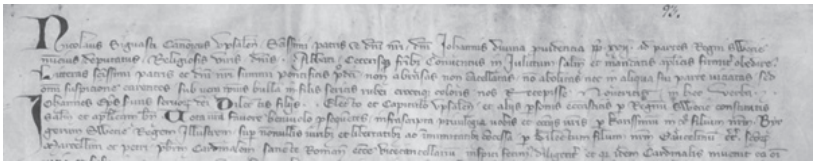

Figure 13. A sample input image of a charter.

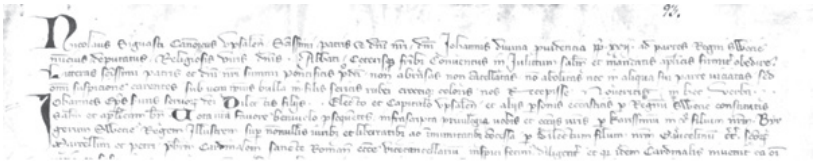

Figure 14. Corresponding pre-processed image after shading correction and two band-pass filtering.

For estimation of the script proportions, the proposed algorithm first extracts contours of letters using an edge detection filtering approach (Hast 2014), and then uses a Gaussian filtering kernel over the whole image. Figure 15 presents a sample output after applying the edge detection filter on a segment of figure 14, highlighting the contours of extracted letters. As a result, an easily measurable template surrounding the core region (i.e. core template) is generated, which is used to compute the average size of the core, and the components reaching above this (high elements, e.g. ascenders) and below this (low elements, e.g. descenders). By accurately estimating the height of the core template, one can compute the size of the ascenders and descenders using the proposed algorithm, and also the distance between the lines.

The proposed algorithm automatically finds the optimal kernel size for obtaining an optimally fitting core. Another issue that the proposed algorithm addresses is that the core template does not always fit well locally, both due to the lines being bent and the fact that the scribe did not write consecutive graphs completely straight. This causes some parts to fall outside the core template. Also, some ascenders and descenders are only partially captured, especially if several ascenders or descenders are close to each other, and this makes the template for the core thicker than it should ideally be. Figure 16 highlights these phenomena, where the ascenders are represented in yellow and the descenders in green. Also, it

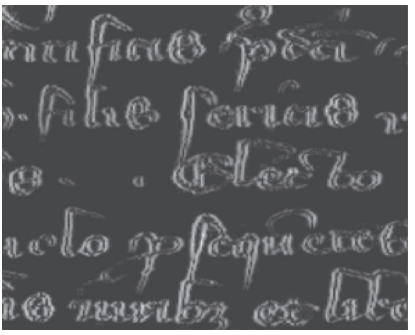

Figure 15. A sample output after applying the edge detection filter on a segment of figure 14.

Figure 16. Representation of the script proportions, where some parts of the characters fall outside the core template, and some ascenders and descenders are partially captured.

Figure 17. Representation of script proportions after post-processing, generating more accurate results.

can be observed that some components outside the core are represented in light blue color. This can be an indication of two things. First, the ascenders and descenders hook into each other, and here the algorithm cannot decide where the descender ends and the ascender starts. Second, the core is substantially different at that location, which reflects that the text is bending more there. Therefore, the ascender or descender is not

perfectly computed in these cases (hence in light blue) as it might be regarded as shorter than it actually is.

Post-processing of the script proportions further refines the results. To do so, we refined the core template by removing the core values larger or smaller than the estimated core size (+/– 2 pixels), and eliminated marginally small ascenders and descenders. This renders the computation of script proportions more accurate, as can be seen in figure 17.

In summary, the use of the methodology made it possible to deal with the often damaged surface of the charters, and with the sometimes slanting script lines, features often present in the written sources of the Middle Ages.

Additional reference

Hast, A. (2014): "Simple filter design for first and second order derivatives by a double filtering approach". *Pattern Recognition Letters*, vol. 42, no.1, June 2014, pp. 65–71.