

## Article

# Abbreviations in Medieval Latin Handwriting

**Björn Gottfried, Marius Wegner, Marianna Spano, and Mathias Lawo | Bremen – Berlin**

## Abstract

Abbreviations were used extensively in medieval Latin manuscripts. One reason for this was to allow economical use of parchment or other kinds of writing materials, which were relatively expensive during the Middle Ages. This article elaborates on the employment of abbreviations in medieval Latin handwriting and how they can be extracted from digitized documents with the support of an image-processing software system. The main objective of extraction is to characterize different writing hands. According to our research hypothesis, abbreviations reveal a great deal about specific scribes. As a first step, the stability of this criterion is analyzed. As it turns out, a similar amount of abbreviations and a similar distribution regarding their positions within individual words can be found for the same copyist.

## 1. Introduction

The objectives of the present research are to analyze handwritings and establish the correspondence between the document image and transcription. This allows texts, and particularly abbreviations, to be found and compared in the original documents. Above all, abbreviations require a level of analysis which goes further than the word level. There are several systems being developed to support the transcription process,<sup>1</sup> but current efforts are either text line-based<sup>2</sup> or word-based.<sup>3</sup> A correspondence between the document image and transcription at the level of glyph images and characters would be desirable, however. This would enable one-to-one mapping between all glyph images and their corresponding characters in the transcription. As soon as a word includes abbreviation characters, one-to-many mapping is required. This entails a number of problems regarding the extraction of abbreviations and how they are represented. The current issue, however, concerns the question as to what

abbreviations tell us about the copyist (or copyists) of a text. Are there any tendencies regarding the use of abbreviations by specific writing hands?

The next section discusses the general background to the employment of abbreviations in medieval Latin handwritings. Problems arising from automatic detection of glyphs and abbreviations in documents are then presented as well as possible methods for dealing with them. The resulting transcription can be exported for further processing. There are specific conditions which apply to abbreviations in this context, and it is necessary to explain how abbreviations can be found in original document images after they have been transcribed. Finally, a number of documents are analyzed in an evaluation aimed at showing whether abbreviations alone provide a useful criterion for characterizing individual scribes.

## 2. Abbreviations

There are two general types of abbreviations to be distinguished in Latin texts: abbreviations consisting of letters or combinations of them, and so-called conventional signs, which cannot be connected to letters but may have their origins in ancient tachygraphy (such as  $\rho$  for *com-/con-*,  $\rho$  for *-us*,  $\gamma$  and  $\&$  for *et*,  $\div$  for *est*, and so on). Literal abbreviations can be divided into ‘suspensions’ (*Suspensionskürzungen* in German), which are words interrupted after a certain number of letters ( $\geq 1$ ) regardless of the syntactical connection of the word, and ‘contractions’ (*Kontraktionskürzungen* in German) consisting of at least the initial letter of a word and its ending.

One-letter suspensions are the oldest type.<sup>4</sup> They appear in a number of public documents dating back to the Roman republic (509–27 BCE), and the majority of them are familiar to us from the uniquely transmitted work of the grammarian Marcus Valerius Probus (fl. 1st century CE), a well-known example being SPQR for *Senatus Populusque Romanus*.

<sup>1</sup> Serrano 2013, Wüthrich et al. 2009, and Romero 2007.

<sup>2</sup> Romero 2007.

<sup>3</sup> Serrano 2013.

<sup>4</sup> Bischoff 2009.

Depending on the context, however, it might also stand for another case such as the accusative *Senatum Populumque Romanum*. This ambiguity was already recognized as a problem in antiquity, and the use of this type of abbreviation in legal texts was prohibited in 438 CE.

The use of contractions arose from the *nomina sacra* ('holy names') in Christian religious texts describing, for instance, the Holy Trinity as *d̄s p̄r* (*deus pater* = God the father), Jesus Christ as *Īhs X̄ps*, and the Holy Spirit as *s̄p̄s s̄c̄s* (*spiritus sanctus*), where all abbreviations are marked by a simple stroke above the central letter or whole combination of letters. It is possible to trace the Greek origin of the middle example, since *h*, *X*, and *p* most certainly stand for Greek *eta*, *chi*, and *rho*. We can therefore assume that many of the issues summarized in this paper for text written in Latin characters can also be applied to other letter systems. Contractions also pose certain problems for readers and editors of ancient manuscripts: given that the most common handbook of abbreviations by Adriano Cappelli<sup>5</sup> comprises not more than around 15,000 abbreviations, it is quite clear that the majority of inflected forms – especially inflected verb forms – are missing. For Cappelli and his human readers, this was no trouble at all, since they had the philological understanding to combine (correctly as well as freely) the abbreviated roots and endings of a word into a known form. Completely different conditions apply to computers and machine-readable resolutions of abbreviations, of course. Thus, the *Abbreviationes* database by Olaf Pluta<sup>6</sup> currently (as of September 2014) 'comprises over 70,000 entries containing a total of 80,098 references to manuscripts', but even that does not contain every single form. The imperfect tense of the Latin verb *habere* (= to have), for example, comprises 24 different forms: six in the active voice (three singular, and three plural) and six in the passive voice, each of them in the indicative and subjunctive moods. Only three of the twenty-four forms can be found in Cappelli:<sup>7</sup> the third-person singular (indicative *h̄ebat*, *h̄ēbt* = *habebat*, subjunctive *hab̄s&*, *hēs̄t*, *h̄ēr̄s̄t*, *h̄²*; = *haberet*) and the first-person subjunctive *h̄rēm* = *haberem*. As can be seen, there are different ways of abbreviating the same form, and there is no reason why the omitted forms should

not be abbreviated in a similar way. It is therefore hardly surprising that the more extensive *Abbreviationes* database includes eight different forms, namely *habebant*, *habebat*, *habebatur*, *haberem* (three types of abbreviation), *haberent*, *haberes*, *haberet* (abbreviated in nine ways), and *haberetur*. On the other hand, the same abbreviation can be used for completely different words. Thus, *m̄r̄m* might stand for *matrimonium*, *martyrum*, *monstrum* or – not mentioned by Cappelli<sup>8</sup> – also for *magistrum*, *matrum*, *melioem* if you compare it with *m̄r̄i*, for example.

As can be seen from the examples cited for abbreviating *habebat* and *haberet*, it is not only the sequence of the letters that matters (as in Cappelli or the search function in *Abbreviationes*), but also the shape of the abbreviation and its position (an issue considered by the otherwise paleographically normalized entries in *Abbreviationes*). This is evident when we look at letters – there is undoubtedly a difference between reading *ḡ* (= *igitur*) and *ḡ* (= *ergo*) – or combinations with conventional signs written above the line: *p̄* usually stands for *per*, *p̄̇* for *pri*, *p̄<sup>o</sup>* for *post*, *p̄* for *pr(a)e*, and *p* for *pro* (cf. fig. 2), no matter whether it is used as an isolated word (if possible), as a prefix, or as a substitute for the corresponding combination of letters within the word. Where, however, the most common and clearly ambiguous abbreviation signs in manuscripts are concerned, i.e. straight or curved lines, the question may be posed as to whether not only the *position* of the signs above certain letters might be specific to specific writers, but also the length, shape, and direction of the signs. This is one of the fields where computer science produces results which are far more prolific, easily comparable, and reliable than the results obtained by human analysis. Another area of investigation might be the ratio of abbreviations in a larger or smaller context: not only their quality (which abbreviations are used, are there different types of abbreviation for the same word, and so on), but also their quantity (are there more abbreviations at the beginning of the text than at the end or is there a constant proportion, does the ratio of abbreviated and long forms of the same word vary within the text, and so on).

### 3. Extracting glyphs

In order to analyze the use of abbreviations, it is necessary to obtain a mapping between glyphs within the original document image and in the transcription. A single glyph maps

<sup>5</sup> Cappelli 1929.

<sup>6</sup> *Abbreviationes* 2014.

<sup>7</sup> Cappelli 1929, pp. 157ab, 159a, 160a, 164b, 165a.

<sup>8</sup> Cappelli 1929, p. 226a.

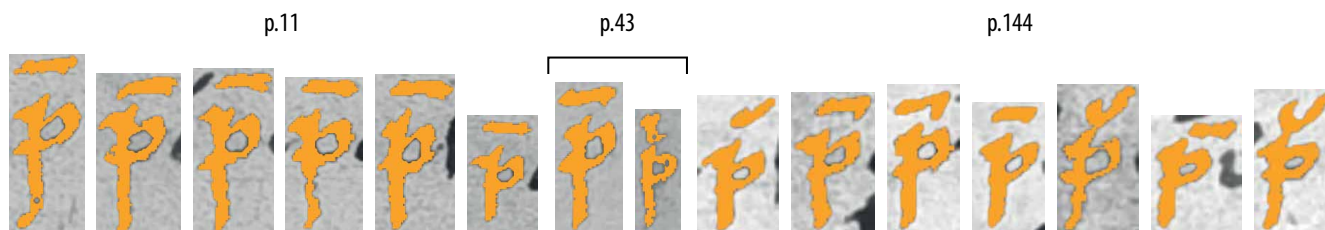


Fig. 1: The abbreviation glyphs of the string 'pre' are shown for three document pages (fol. 11r, fol. 43r, and fol. 144r) distributed over a manuscript of around four hundred pages. Each page is indicated by one of the three boxes. The order of the glyphs corresponds to the order of their appearance in the book. It turns out that there are clear differences in the shape of the glyphs depending on how late in the book they appear. The abbreviation bars in particular differ from case to case. In the last box, the bar is sometimes straight and sometimes curved, whereas the first occurrences look more uniform (box one).

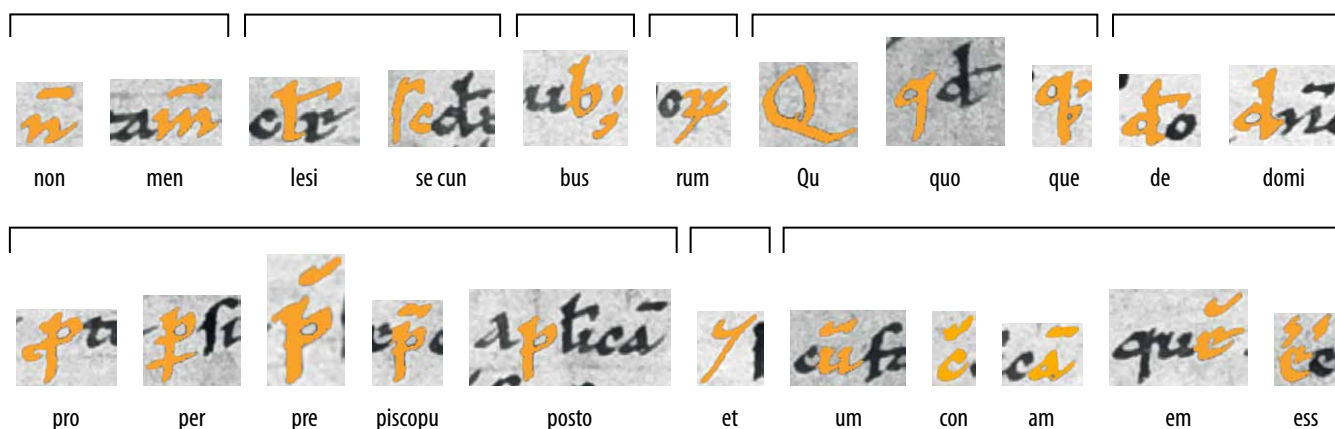


Fig. 2: The most frequent abbreviations employed by Hugh of Flavigny, grouped with respect to visually similar characteristics of the glyph images. These abbreviations are taken from fol. 144r of Philipps 1870.

to a whole string in the case of abbreviations. For this purpose, the software system *Diptychon* employs \$ signs which are used in order to mark the beginning and the end of a string in the transcription. In this way, each chain of characters that is enclosed by \$ signs represents the completed string of an abbreviation. An example is provided in fig. 1, which shows eleven different instances of the character *p* with a bar written above it. This abbreviation expands to *\$pre\$*.

The *Diptychon* software system tries to separate adjacent glyphs automatically. A number of difficult cases might result in glyphs being separated incorrectly inasmuch as *Diptychon* does not make any assumptions about the underlying handwriting. Interactive methods are available in cases such as these in order to correct the proposed separations.<sup>9</sup> Abbreviations present an additional difficulty, since they frequently consist of two or even more disconnected components, such as the *p* with a bar above it, as shown in fig. 1. In these cases, the user can let the system know which disconnected regions pertain to the same logical unit.

Each logical unit is mapped to a string in the transcription. This mapping is also done automatically as follows: the rows of the document image are detected and, for each text row, an input field for the transcription is generated on the user interface. The line by line correspondence between document image and transcription breaks down the correspondence problem to single rows, and the correspondence is established for each single row based on the linear order of the glyphs in the image. Apart from some specific difficulties such as supplements between the lines which the user has to correct manually, this approach enables seamless mapping between glyphs and transcription. By employing \$ signs, it especially includes the handling of abbreviation glyphs and the corresponding strings in the transcription.

Some of the most frequent abbreviations used by 11th-century author Hugh of Flavigny in his chronicle are shown in fig. 2. They are grouped with respect to the visual characteristics of the glyph images, showing that similar or even identical glyphs might represent different abbreviations, for example *pre* and *piscopu*. In these cases, the context of the abbreviations is required in order to translate them

<sup>9</sup> Gottfried, Wegner, and Lawo 2013; Gottfried, Wegner, and Lawo 2014.

correctly. A horizontal bar is frequently written above a glyph, indicating an abbreviation. This is not always the case, however; a high degree of expertise is required if paleographers are to recognize abbreviations correctly.

#### 4. Exporting abbreviations

While the user interface of the *Diptychon* system requires the user to employ \$ signs in order to mark abbreviations, these signs are not desired when exporting the transcription to a file. For the critical edition of a manuscript text, however, it can be helpful to indicate where abbreviations are found in the original document. Moreover, the edited text should make it clear which characters mark the beginning of an abbreviation and are hence shown as glyphs, in contrast to the abbreviated letters that do not appear in the original document. All the abbreviated letters are therefore enclosed in parentheses in the export file, and the characters shown as glyphs are placed before the parentheses. This enables the editor to subsequently analyze the abbreviations purely on the basis of the exported transcription. For instance



is transcribed in *Diptychon* as *\$con\$scili\$um\$* and is exported as *c(on)ciliu(m)*. The isolated abbreviation



becomes *\$et\$* in the transcription and is exported as *e(t)*, and



is transcribed as *\$francoy\$rum\$* and exported as *Francor(um)*.

There are also examples of glyphs which are not the characters that start the abbreviation, such as *qđ* in fig. 2, which stands for *(quo)d*. There are certain abbreviation characters which seem to have been invented solely for the purpose of abbreviating strings. Most other abbreviations are only distinguished by an extra sign extending a common character, such as a vertical bar above the abbreviation glyph, or a point or semicolon to the right of the glyph.

#### 5. Searching for abbreviations

Once the transcription of a document is available, it is possible to search for all occurrences of a string and in particular for occurrences of abbreviations. While the underlying search methods make use of the transcription which is stored in a symbolic form, each character or abbreviation string is linked to the corresponding glyph within the original document

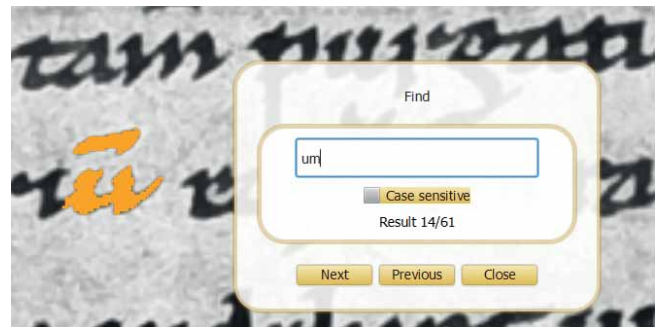


Fig. 3: The search dialog provided by *Diptychon*.

image. The whole region of the latter is accessible, meaning that it can be emphasized by color highlighting. The user gets the impression that the glyphs are found in the original document image. This enables the paleographer to gain an insight into the distribution of abbreviations, their visual appearance, and the contexts in which they are used. Fig. 3 shows the search dialog.

It is possible to search for an arbitrary string which might specifically contain space characters. This allows a search for certain types of abbreviations such as those that start a word. For this purpose, the search string should start with a space character followed by the requested abbreviation. Another option is to search for abbreviations that are found at the end of a word or form an entire word in themselves. In the latter case, the abbreviation has to be enclosed by space characters. Likewise, a broader context of adjacent words can be taken into account. Abbreviations are not used in every case, however, and sometimes all the characters of a word are given. Nevertheless, the search result contains every occurrence of the relevant string, either as part of an abbreviation or given explicitly, in order to show how abbreviations are used in the relevant document.

#### 6. A case study on abbreviation criteria

The long-term strategy is to employ different criteria as a means of characterizing writers. One such criterion concerns the way in which a writer uses abbreviations. Before this criterion can be applied, however, it has to be analyzed to determine its robustness. If a writer uses abbreviations to a similar degree in different documents or in pages of documents written at different times, the employment of abbreviations can be deemed stable and can therefore be regarded as a robust criterion.

As an example, the robustness of abbreviations is analyzed for Hugh of Flavigny, who was a late 11th-century chronicler. A manuscript (probably an autograph) by this writer still

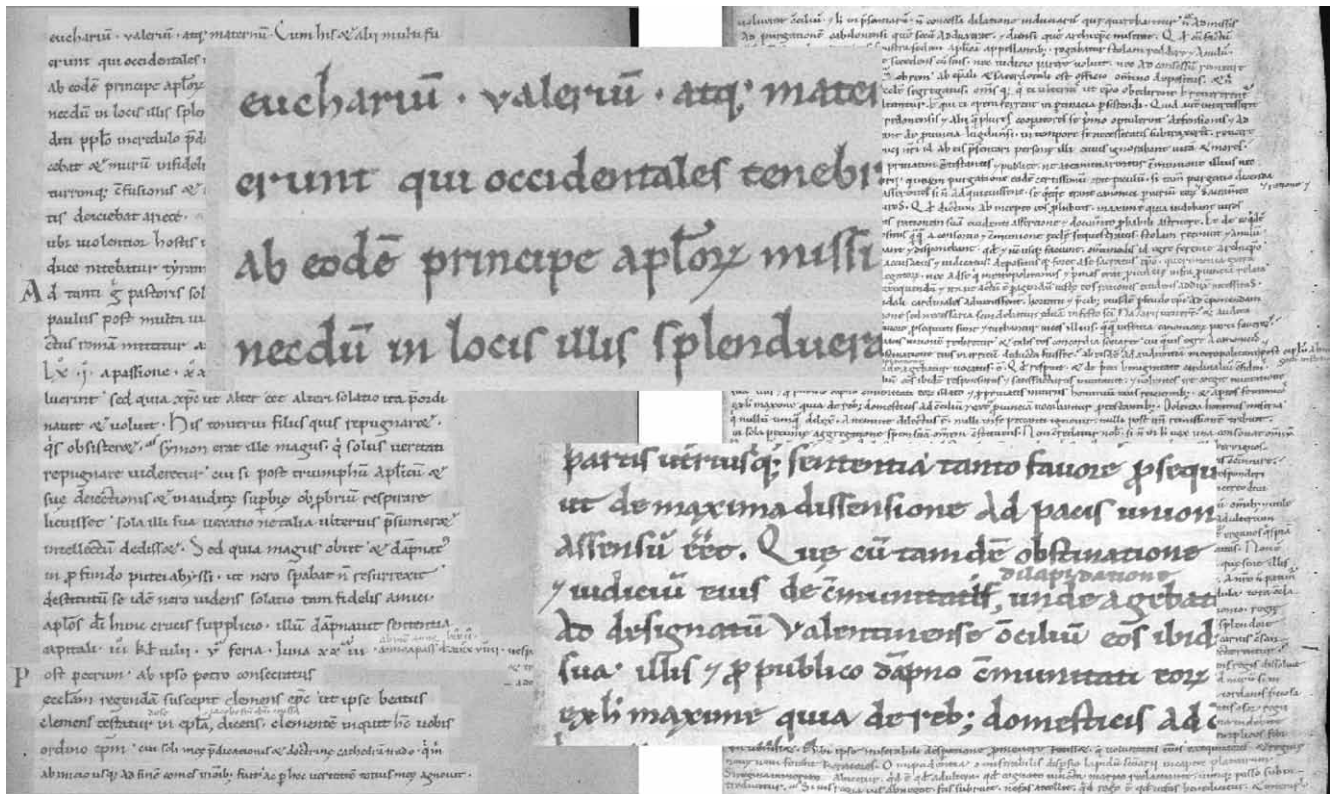


Fig. 4: Hugh of Flavigny, fol. 11r (left) and fol. 144r (right).

exists in the Berlin State Library. It comprises approximately 400 pages in two volumes. This is extensive enough to allow a comparison between the use of abbreviations at an early stage in the codex and how they are used in another passage which comes later in the same book. Fig. 2 shows two pages – the one on the left is from the initial part of the first volume, and the page on the right is from the end of the work. The first page looks relatively unobtrusive, whereas the latter page is much more complex, since the writer is trying to exploit the full range of the parchment. To be precise, folio 11r comprises 1,497 glyphs in 38 text lines, and folio 144r consists of 4,872 glyphs in 60 lines.

While 46 different abbreviations are used on fol. 11r, there are 62 different abbreviations on fol. 144r. This results from the fact that the text on fol. 144r is 3.25 times longer in terms of the number of glyphs than the text on fol. 11r. In other words, 7.5% of the glyphs on fol. 11r are abbreviations, while there is a similar amount of abbreviations on fol. 144r (7%).

Four types of abbreviations can be distinguished:

- a) at the beginning of a word
- b) at the end of a word
- c) at an intermediate position in the word

d) discrete from other glyphs (abbreviations that form a whole word in their own right).

The following holds for fol. 11r: (a) 26%, (b) 38%, (c) 16%, and (d) 21%. This shows that the frequencies are quite similar to fol. 144r: (a) 28%, (b) 37%, (c) 13%, and (d) 22%. Basically, a writer might employ abbreviations in each of the four different categories to an arbitrary degree. It can therefore be concluded that both pages together, which are placed more than two hundred and fifty pages apart, provide a first indication that Hugh of Flavigny used abbreviations consistently.

Both pages have 23 abbreviations in common, which are shown in fig. 5 together with their frequency distribution. This means that 50% of the abbreviations on fol. 11r are common to both pages, and 37% of those on page 144. Of the common abbreviations, five out of the first third have a similar ranking concerning their frequencies, namely *um*, *et*, *em*, *am*, and *per*.

In order to confirm these results, a third page was analyzed which was positioned right in the middle of the other two pages. The distribution of abbreviations differs slightly in comparison to the other two examples, but the tendency is the same (the values for fol. 11r and fol. 144r are given in

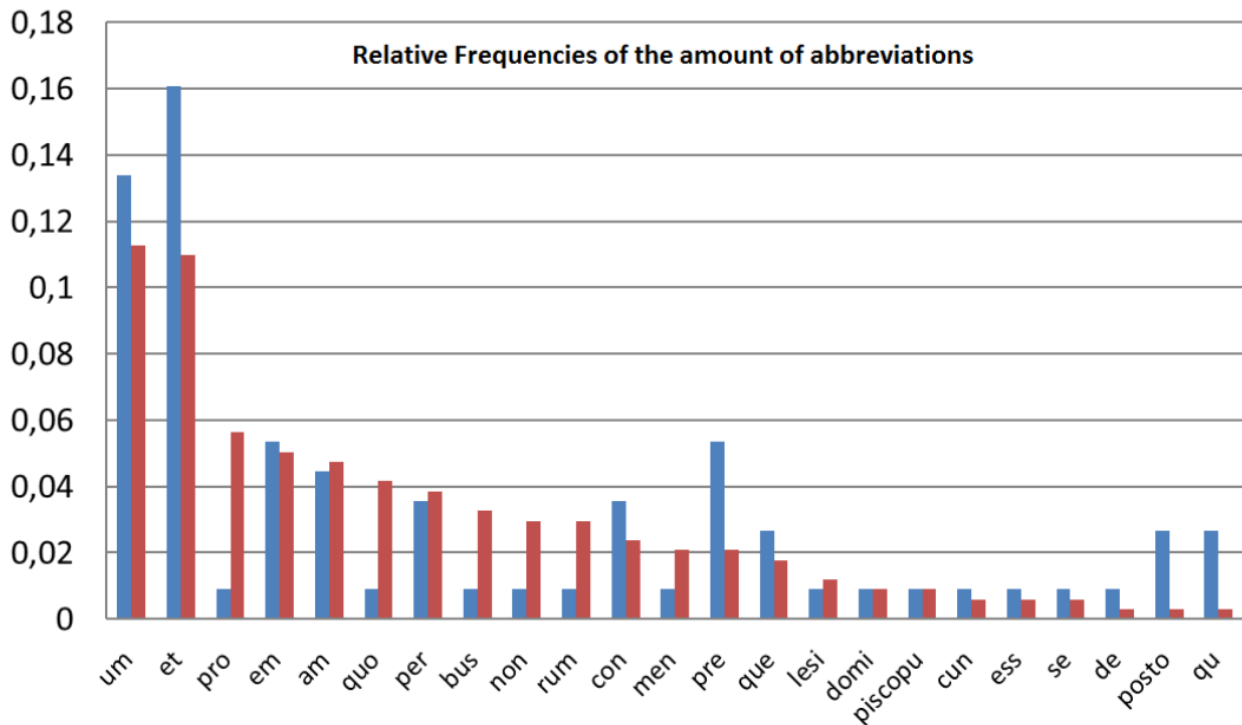


Fig. 5: There are 23 common abbreviations used on fol. 11r (blue) and fol. 144r (red). Their relative frequencies are similar as regards some of the most commonly used abbreviations and somewhat different when it comes to other abbreviations.

parentheses): The following holds for fol. 43r: (a) 13% (26%, 28%), (b) 52% (38%, 37%), (c) 17% (16%, 13%), and (d) 17% (21%, 22%). In other words, most abbreviations are found at the end of words in the third document sample – the same as on the other two pages. A similar amount of abbreviations appear in intermediate positions, but there are slightly fewer at the beginning of words as well as isolated abbreviations on page 43. There are a total 9.8% of abbreviations on this page with 2,232 glyphs (7.5% on fol. 11r, and 7% on fol. 144r). Having analyzed the robustness of the abbreviation criterion in this way, future work will have to reveal how stable the use of abbreviations is for other writers as well.

We could also ask how many abbreviations are found within documents written by other writers. In order to shed light on this question, we examined another document: a charter by Emperor Charles IV issued in the year 1361.<sup>10</sup> This contains 1,128 glyphs with an abbreviations ratio of just less than 5%, thus showing a less frequent use of abbreviations. 29% of the abbreviations are found at the beginning of words, while 38% are found at the end. This appears to be similar to the use of abbreviations by Hugh of Flavigny. There is, however, a clear difference concerning isolated abbreviations, namely

that there are only around 5% in the imperial document, but an average of 20% in the chronicle by Hugh of Flavigny. Another difference concerns abbreviations that are situated in the middle of a word: there are around 29% in the imperial document, and approximately half as many in the other case. Although this demonstrates a clear difference in the employment of abbreviations, it is only regarded as a first indicator for our assumption that abbreviations may have been used differently by different writers.

### 7. Summary

There is a striking use of abbreviations in medieval Latin manuscripts. It is therefore of interest to examine exactly how abbreviations were used by different writing hands, and whether they can be isolated as a distinguishing feature. The *Diptychon* software system we use has been developed for the precise extraction of glyph images and the investigation of abbreviations in medieval documents. Initial indications that our research criterion is stable for single scribes, but that there might be clear differences with regard to different writers have been found in a dataset containing 9,729 glyph images. These images were extracted from the documents of just two scribes. Evidently, results will have to be compared for many more different writers in future work in order to substantiate our hypothesis.

<sup>10</sup> Bayerisches Staatsarchiv Nürnberg, RU Nürnberg 1086.

## ACKNOWLEDGEMENTS

We gratefully acknowledge the funding support provided by the German Research Foundation (DFG) under Grant Numbers GO 2023/4-1/LA 3066/4-2 and LA 3007/1-1 under the project name *Diptychon*. We are thankful to Jan-Hendrik Worch for supporting the software development.

## REFERENCES

- Abbreviationes*<sup>TM</sup>, ed. by Olaf Pluta, (<http://www.ruhr-uni-bochum.de/philosophy/projects/abbreviationes/index.html>).
- Bayerisches Staatsarchiv Nürnberg, *RU Nürnberg 1086*.
- Bischoff, B. (2009), *Paläographie des römischen Altertums und des abendländischen Mittelalters*, 4. Aufl. (Berlin: Erich Schmidt; Grundlagen der Germanistik 24), 202–223.
- Cappelli, A. (1929), *Lexicon Abbreviaturarum. Dizionario di Abbreviature Latine ed Italiane*, 4th edition (Milan: Hoepli; several reprints).
- Gottfried, B., Wegner, M., and Lawo, M. (2013), ‘Diptychon: A transcription assistant system for the separation of glyphs in medieval manuscript texts’, in E. Angelopoulou et al., *20th SAOT Workshop on Automatic Pattern Recognition and Historical Document Analysis, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany. June 14–15, 2013* (The Digital Library of the Göttingen Academy of Sciences and Humanities).
- , ———, and ——— (2014), ‘Towards the interactive transcription of handwritings: anytime anywhere document analysis’, *International Journal on Document Analysis and Recognition (IJ DAR)*, (DOI 10.1007/s10032-014-0234-7), 1-15
- Lawo, M. (2010), *Studien zu Hugo von Flavigny* (Hannover: Verlag Hahnsche Buchhandlung; Monumenta Germaniae Historica. Schriften, 61).
- Romero, V., Toselli, A. H., Rodriguez, L., Vidal, E. (2007), ‘Computer-assisted transcription for ancient text images’, in M. S. Kamel and A. C. Campilho (eds.), *Image analysis and recognition (ICIAR). 4th international conference, (Montreal, Canada), Lecture Notes in Computer Science*, vol. 4633: 1182–1193.
- Serrano, N., Gimnez, A., Civera, J., Sanchis, A., Juan, A. (2013), ‘Interactive handwriting recognition with limited user effort’, *International Journal on Document Analysis and Recognition*, 1–13.
- Staatsbibliothek zu Berlin, Preußischer Kulturbesitz, Ms. Phill. 1870, fol. 11r and fol. 144r.
- Wüthrich, M., Liwicki, M., Fischer, A., Indermühle, E., Bunke, H., Viehhauser, G., Stolz, M. (2009), ‘Language model integration for the recognition of handwritten medieval documents’, in *10th International Conference on Document Analysis and Recognition (ICDAR)* (Piscataway, NJ: IEEE Computer Society), 211–215.