EVELINE LECLERCQ AND MIKE KESTEMONT

Advances in Distant Diplomatics

A Stylometric Approach to Medieval Charters

Abstract

The quantitative analysis of writing style (stylometry) is becoming an increasingly common research instrument in philology. When it comes to medieval texts, such a methodology might be able to help us disentangle the multiple authorial strata that can often be discerned in them (issuer, dictator, scribe, etc.). To deliver a proof of concept in 'distant diplomatics', we have turned to a corpus of twelfth-century Latin charters from the Cambrai episcopal chancery. We subjected this collection to an (unsupervised) stylometric modelling procedure, based on lexical frequency extraction and dimension reduction. In the absence of a sizable 'ground truth' for this material, we zoomed in on a specific case study, namely the oeuvre of the previously identified dictator-scribe known as 'RogF/JeanE'. Our results offer additional support for the attribution of a diplomatic oeuvre to this individual and even allow us to enlarge it with additional documents. Our analysis moreover yielded the serendipitous discovery of another, previously unnoticed, oeuvre, which we tentatively attribute to a scribe-dictator 'JeanB'. We conclude that the large-scale stylometric analysis is a promising methodology for digital diplomatics. More efforts, however, will have to be invested in establishing gold standards for this method to realize its full potential.[1]

1 Introduction: stylometry, authorship and diplomatics

In recent decades, and due to the influence of Digital Humanities, the quantitative study of writing style (or stylometry) has established itself as a relatively mainstream approach in literary studies. Certainly for modern literature, there now exist powerful modelling techniques that can shed new light on the authorship of anonymous or disputed documents on the basis of their writing style. Most of these methods are drawn from subfields of computer science, such as machine learning and information retrieval, that are concerned with similar issues re-

garding the provenance and authentication of documents. In computational studies for present-day texts, however, the concept of authorship is rarely explicitly defined, let alone critically questioned. For a given body of written text, the task of authorship identification comes down to the demarcation of the set of individuals that are directly responsible for its verbal (linguistic and stylistic) properties. Typically, although not exclusively, the author is understood to be a single person in these cases. While this operationalization might often suffice for modern texts, the apparent ease with which the concept of authorship can be applied in present-day computer science cannot easily be transposed to the realm of medieval text production.

For medieval texts, authorship is in many instances a much more layered phenomenon than for more recent literature. Medieval autographs, for instance, rarely survive, and one must account for the series of subsequent scribes that were involved in the transmission of a text, who all had their contribution to its final appearance – not to mention the impact of the modern text editor, who has inevitably also left traces in the edited text. In this contribution, we focus on medieval charters, where the phenomenon of authorship is even more complex and fragmented than in other medieval text varieties. A multitude of authorial roles must be distinguished, at least if one would like to do justice to the complex textual production process from which these texts resulted.[2] Indeed, at least three authorial roles must be distinguished when modelling the production of a medieval charter, and commonly, it was not a sole individual who fulfilled all these roles.

Firstly, the authority (*i.e.* the bishop, pope, count, local lord, abbot, …) responsible for the production of the charter can take on the role of author (the person expressing his/her will in the charter[3]) or that of 'issuer' (the person in whose name the charter is produced).[4] In this paper, for example, the bishop (the authority) frequently occurs as the 'issuer' of an episcopal charter. Apart from this author(ity), it is furthermore necessary to distinguish between the 'dictator' – *i.e.* the person who composes and perhaps dictates the (gist of the) text – and the 'scribe', *i.e.* the person putting the act into its final form of the original charter. This paper will henceforth make strict use of the terms 'dictator' and 'scribe' so as to avoid confusion between the two roles.

Identifying (or even just distinguishing between) these individuals is an essential task in diplomatics, the scholarly field that is concerned with the study of medieval charters, because this helps to sit-

2. An excellent discussion of this problem can be found in Tock, "Auteur ou impétrant?" 215–48. Here, we define charters as short legal documents, describing, for example, the accomplishment of a legal act or the existence of a legal fact, or thirdly a fact which gains its validity from the fixed form in which it has been written down (Cárcel Ortí 21)

3. "L'auteur d'un acte juridique est la personne qui manifeste la volonté exprimée dans cet acte." (Cárcel Ortí 23).

4. "L'auteur d'un acte écrit est la personne au nom de qui cet acte est intitulé." (Cárcel Ortí 24).

uate charters in space and time. Lexical (*i.e.* word-level or phrase-level) analyses, amongst others, have been previously applied to assess which individuals might have transcribed or composed specific charters; especially the use of fixed or formulaic language has frequently drawn attention in this respect. The term *dictamen* is commonly used to capture aspects of the discourse or writing style in the field of diplomatics.[5] This type of fixed-form writing heavily depended on local preferences and habits in the composition of documents, such as standard formulas shared by the dictators of a particular writing office. Such intertextual phenomena are crucial to understand how texts participated in the discursive culture of the Middle Ages (Culler 1380–1396). Somewhat surprisingly, computational techniques so far have been applied only infrequently in diplomatics, especially on a larger scale, although these could offer a useful complement to the kind of manual, and typically very labour-intensive approaches that already exist in the field.[6] Nevertheless, important advances have already been made regarding the electronic availability of source collections, as multiple digital editions and online databases of charters have emerged online in the last few decades such as the *Diplomata Belgica, Chartae Galliae* and *monasterium.net,* as well as a plethora of tools for transcribing and analysing historical texts such as the projects *Groningen Intelligent Writer Identification System,*[7] *Transkribus,*[8] *Escriptorium,*[9] *Arkindex*[10] and *Digipal.*[11] When it comes to stylometric techniques, one could furthermore mention Michael Gervers' DEEDS-project, used primarily to date undated medieval English charters (Gervers and Margolin 2007).[12] More recently, Nicolas Perreaux has applied unsupervised statistical techniques to a corpus of 250,000 charters to discover general trends in the diplomatic scriptural practices of France and the regionalization of these texts' vocabulary (Perreaux 2015 and 2016). This work is an important precursor to ours.

Thus, in an analogy to 'distant reading' (as opposed to 'close reading') in literary studies, this paper sets out to assess the feasibility of a stylometric approach in the context of 'distant diplomatics', *i.e.* the application of computational techniques to analyze charter corpora that are larger than one could analyze manually using conventional methods. Perreaux has already demonstrated that a panoramic perspective allows the detection of patterns and trends that would otherwise remain invisible. Whereas his research focuses on studying the dissemination of the vocabulary used in charters and its reflection of medieval society, our focus will lie with a more specific appli-

5. The definition used by Cárcel Ortí describes dictamen as "the internal arrangement of an act, its style, which is supposed to meet the rules of the chancery and the precepts of rhetoric" (Cárcel Ortí 82).

6. The latest advancement toward a digital approach has been the method of E. De Paermentier, which in a nutshell carefully compares selected formulas in a charter corpus, drawn from extensive databases. (De Paermentier, 48–67). As such, the method of E. De Paermentier could be viewed as a manual version of a stylometric analysis.

7. Web. Accessed 20th May 2021.

8. Web. Accessed 20th May 2021.

9. Web. Accessed 20th May 2021.

10. Web. Accessed 20th May 2021.

11. Web. Accessed 20th May 2021.

12. Currently, the part of the project which analyses the text of the charters has been discontinued; a database of dated charters allowing for comparison is still available and being expanded.

cation, namely stylistic similarities in the presence of the layered authorial structure presented above. The main challenge here is that we must mostly work in the absence of a 'ground truth', *i.e.* most of the time we do not have a small starting corpus for which we know the actual dictators and against which we can compare a wider set of documents. That will inevitably call into question, if we find stylistic similarities at all, at which authorial layer these resemblances must be interpreted.

From the beginning, we should note a few major methodological challenges. Firstly, charters contain fairly short texts, with a typical length ranging from 200 to 400 words, and even less once preprocessing techniques have been applied (see figure 1). This is problematic for many stylometric techniques (Luyckx, Daelemans 35–55; Koppel, Schler, and Shlomo 83–94). However, shorter texts such as blogs, e-mails, newspaper articles and tweets have already succesfully been considered in stylometric approaches (Hirst and Feiguina 405–17; Diederich, Kindermann, Leopold, and Paass 109–23; Sanderson and Guenter 482–91). Secondly, charter texts have a very formulaic character: their language and used formulas (dictamen) are largely fixed, depending on the writing office in which they were produced. In the larger discussion concerning the personal voice of the dictator in the dictamen of charters, it is nevertheless assumed that a degree of personal, stylistic freedom was still possible (Dijkhof 50). The application of stylometric methods to charter texts must nevertheless acknowledge this idiosyncratic aspect of the material.

The central focus here is the analysis of dictamen with the aid of stylometric techniques applied to a single case study. This case study concerns the twelfth-century bishops' charters of Cambrai, chosen for two reasons. Firstly, this particular writing office was rather well-organized and established during a time period when chanceries, dictamen and the *artes dictandi* in general were still in the process of development, thus offering a large corpus of preserved sources readily available in online databases. At the moment of writing, 505 charters dated between 1100 and 1201 are available (with their transcriptions) in the *Diplomata Belgica*[13] and 140 in the *Chartae Galliae*.[14] 100 out of 645 charters were duplicates (included in both databases), which left us with a corpus of 545 individual texts to encode. Secondly, the bishops' charters have been amply studied, allowing for a deeper analysis instead of still having to do the ground work (Van Mingroot, *De bisschoppelijke kanselarij*, 693–701; Van Mingroot, *Les chartes*, 1–15; Barré, "Chancellerie *épiscopale*," 129–46; Leclercq,

13. This online database aims at containing all preserved charters relating to the current territory of Belgium dated before 1250. De Hemptinne, *et al.* In this paper, charters from this database will be referred to as 'DiBe' numbers.

14. This online database collects charters, dated up until the end of the thirteenth century, relating to the current territory of France. In this paper, charters from this database will be referred to as '*Chartae Galliae* Charter n°' numbers.

"L'élaboration des chartes *épiscopales*"). An explorative analysis like ours, we hope, could mean a step forward in the wider application of statistical methods in diplomatics.

The structure of this paper is as follows: first, we will describe the materials and detail the methodology, with ample attention to the limitations of our undertaking. Next, we zoom in on a case study (RogF/JeanE) and discuss the result from a stylometric perspective. In a follow-up section, with a qualitative analysis, the paper interprets the stylometric results with a conventional dictamen approach. Importantly, we resort to text reuse detection as an aid in assessing the similarities between documents. Finally, we comment on the possibilities of such an approach for dictator research, and its place in the future of the field of diplomatics.

2 Method and limitations

Below, we report on a stylometric analysis that we have performed on a representative corpus of medieval charter material; that is to say: a large enough group of charters, that is adequate in scope, but which doesn't overwhelm the scholar either. The main limitation of this material, as mentioned, is that most of these charters lack trustworthy labels, or a 'ground truth' in machine learning terms, regarding the dictators (and often also scribes) that were involved in their production. This means that we have to resort to an unsupervised analysis of the material which cannot be evaluated using traditional quantitative measures, such as attribution accuracy. We shall therefore critically discuss a visualization (in the form of a scatter plot) in terms of the global and local clusters which emerge. Through a close reading of some of the original charters in these local clusters, we aim to determine how these groupings relate to the state of the art in the non-quantitative scholarship in the field. To add focus to our discussion, we shall reserve special attention for a specific group of charters which were in all likelihood produced by a single individual introduced in greater detail further below: "RogF/JeanE."

In diplomatics, individuals who where active in the production of charters often cannot be identified with a specific name. As such, they are attributed using an identifier based on when they were active in a specific group, *c.q.* the bishop's chancery. The term 'chancery' itself could be seen as problematic: the definition of the term var-

ies according to the period and region under scrutiny. In this paper, we use the term as *wissenschaftliche Hilfskonstruktion* designing "the grouping of the people in charge of the production of charters and other official documents, in a more or less regular capacity, in the service of a public authority."[15] The label, or 'emergency name', RogF/JeanE alludes to the presence of this dictator during the periods of office of Roger of Wavrin (1179–91) and Jean II of Antoing (1192–96) as bishops of Cambrai. The capital letter presents the chronological 'order' in which they appeared; RogF/JeanE appears as the sixth chancery collaborator under Roger of Wavrin and as the fifth under Jean II of Antoing. This chronological order is of course dependent on the source material, *i.e.* the preserved charters.

15. Definition by Eveline Leclercq as part of a larger discussion on the problematic term 'chancery'. Leclercq 4-7.

Preprocessing

The composition of this corpus has been guided by the following principles. It contains the texts of all charters available in the databases *Diplomata Belgica* and *Chartae Galliae* for which the Cambrai bishops Roger of Wavrin (1179–91) and Jean II (1192–96) were the responsible authorities, including the charters of the twenty-three beneficiaries implied in the bishop's charters during this period. All charters issued *by* and *for* these beneficiaries are included, as considering only the charters issued *by* the beneficiaries would present too limited a corpus. In total, 108 other authorities are present, such as the Pope, the Count of Flanders and local lords. Of the 591 charters mentioned in the consulted databases, 545 are included with transcriptions (92.2%); though not exhaustive, this corpus is large enough to allow for preliminary conclusions while being managable enough to keep an overview. Next, we have encoded this corpus of 545 charters, by marking the presence of conventional sections as follows: *invocatio* (the protocol: invocation, intitulatio, greeting), *dispositio* (preamble, the corpus of the text – *narratio* and *dispositio* – as well as the announcement of the witnesses), *testes* (the witness list) and *postscriptum* (date, subscriptions and eventual appreciation) using an ad hoc encoding scheme. The analysis reported below is largely restricted to the *dispositio*, as it can be argued that the other sections are (a) too formulaic to reveal any personal, stylistic traits or (b) too rich in case-specific proper nouns to yield stylistic features that would scale well to other documents.

Additionally, we have recorded charter-specific metadata in a

separate spreadsheet.[16] This metadata includes for each charter (when available):

Identifier: the file name, composed of the charter reference number in the *Diplomata Belgica*, preceded by the name of the subcorpus. The episcopal charters carry the identifier "Ca_," whereas the beneficiaries are mentioned as a whole (for example "Affligem_1677").

Text: the reference to the text edition in the *Diplomata Belgica* (DiBe) or *Chartae Galliae* (ChaGal), followed by their reference number in the respective database.

Date: the date of the charter, specifically based upon the tag "datation interprétée" in the *Diplomata Belgica*.

Authority: the diplomatic author of the charter, designated by a name or initials. The initials are derived from these available in the *Diplomata Belgica*: "CO" (Count), "DUC" (Duke), "ROI" (King), "EMP" (Emperor), "SE" (lord), "EP" (Bishop), "AEP" (Archbishop), "Pope," "LEGAT" (papal legate), "Other" (all other available titles, mostly of private persons). In case of multiple charters issued by the same private person, these last initials are replaced by a specific name, as well as the specific title such as "bourgeois" or "châtelain." Finally, charters could be issued by clerical institutions or their members; these charters carry the name of the institution as the *authority* tag.

Specific name: the proper name of the implied authority, for example, "Roger of Wavrin."

Issuer: the person in whose name the charter is produced.

Beneficiary: the beneficiary of the charter.

Subject: the subject treated in the charter text.

Authenticity: whether the charter is considered a forgery, of doubtful provenance, or not suspicious.

*Tradition***:** the textual tradition, specifically "copie" (a copy, of which sometimes the type has been specified in the databas-es: "copie (cartulaire)," "copie (source diplomatique)," "copie (manuscrit)" or "copie (d'érudit)"), "inconnue" (unknown), "original" or "vidimus."

Scribe: the individual responsible for the transcription of the original charter. In case the charter has only been preserved as a copy, the placeholder provided is "n/a." Often, this person is unknown, in which case the label "ncid" (no

current identification) has been used. When the scribe has been identified, the initials are adopted from the preliminary palaeographic research.[17]

Dictator: the person who composed the text. In case of doubt, the initials are followed by a question mark. Most often, these persons remain unknown. Whenever possible, the place of redaction has been determined as follows: "chancery," "beneficiary," "mixed" (in case of a dictamen showing characteristics of the chancery as well as the beneficiary). "Uncertain" has been adopted for the undetermined cases.

Colour code authority: a unique code, allowing for regrouping the charters by authority.

Colour code subcorpus: a unique code, allowing for regrouping the charters by subcorpus.

Because this textual corpus contains many references to named entities (*e.g.* proper nouns), which are less usable for a stylistic analysis, we have decided to 'mask' these named entities, inspired by the practice proposed by Stamatatos (Stamatatos 461–73). Because the corpus is drawn from critical editions that consistently capitalize most named entities, this process could be automated. The only problematic cases that remain are capitalized words at the start of sentences. Therefore, unless the token also appeared in a non-capitalized form somewhere in the corpus, we masked out all capitalized words by replacing them with a series of asterisks of the same length as the original token ('Rogerus' would, for instance, become '*******'). Finally, the entire *dispositio* was lowercased and punctuation was removed. Below, we define words as space-free character strings in the preprocessed material. Figure 1 below summarizes some charter-level lexical statistics about the corpus: with a boxplot, we show the distribution of the word length of the charters, the number of unique words in them, as well as the number of unique words after masking (including the .25, .50 and .75 quartiles to give an idea of the distribution). Cumulatively, the charters in this dataset amount to 107,635 words, of which 18,540 are unique; the latter number drops to 12,254 after masking.

Figure 1. Charter-level lexical statistics: boxplots for the word length and number of unique words (before and after masking) for the 545 charters in the dataset. We include the .25, .50 and .75 quartiles to give a sense of the distribution.
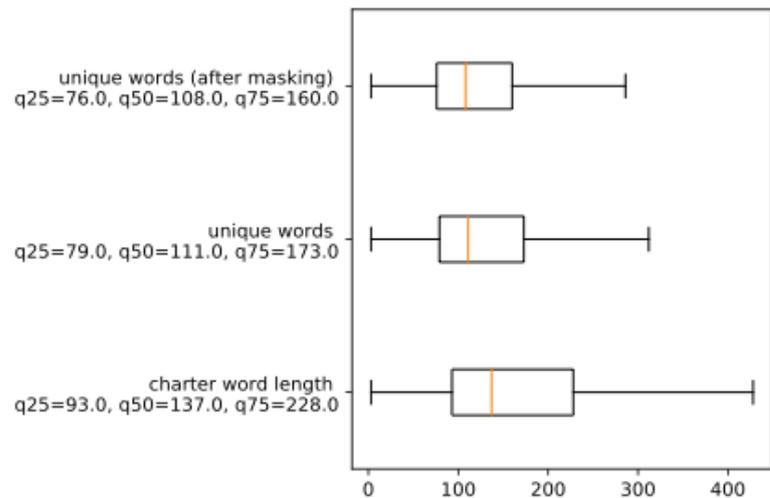
Analysis: stylometry

For the analysis, we rely on a bag-of-words model, which is common practice in digital text analysis (Sebastiani 1–47). We represent the corpus as a frequency table, in which each row represents the *dispositio* of a single charter, and each column a particular word. The entirety of words in this table are referred to as the vocabulary. The cells of this table contain the relative frequency of each term. The rows in such a table are commonly called 'document vectors:' a list of numbers that characterize a document through a list of frequency values. We have applied a common scaling procedure to these numbers that is called TF-IDF normalization (term frequency-inverse document frequency). Simply put, this procedure will assign a relatively higher weight to terms that are relatively less frequent in the corpus, and thus more document-specific. Importantly, we have limited our bag of words model to the 1,000 words that have the highest frequency across the entire corpus. This limitation ensures that we work with a limited set of words that are well enough distributed over the data; *i.e.* not too sparsely distributed (Rybicki and Eder 315–21).

Next, we have created a square distance matrix: using an established distance metric from geometry (the Euclidean distance), we have calculated the distance between each combination of two charters. In the fictional case of a corpus consisting of three documents, this calculation would yield 9 (= 3 x 3) distance values that can be organized in a square matrix of 3 rows and 3 columns. This matrix would be symmetrical (*i.e.* the distance between document A and B would be equal to the distance between document B and A). Additionally, this matrix would have a diagonal of three zeros, because the Euclidean distance of a document to itself would always be zero. Previous

research has shown that such distance metrics are well suited to capture the stylistic differences in writing style between document vectors. If two documents have a similar stylistic profile, the geometric distance between them will be relatively small – and vice versa, in the case of two documents that have a highly dissimilar writing style. Due to a lack of ground truth data, we have not been able to optimize this system. We, therefore, resort to parameters that seem well established in the literature.

Note that in the present case, we would end up with a 548 x 548 distance matrix, in which each row can again be considered a document vector, because it characterizes a single charter through a list of distance values, *i.e.* it reflects how distant a document is with respect to all other documents. We then represent this distance matrix using a popular visualization method called t-SNE, or *t-Distributed Stochastic Neighbor Embedding* (Van der Maaten and Hinton). This method visualizes the entire corpus as a two-dimensional scatterplot, in which each charter corresponds to a geometric point. Importantly, this technique is well-apt at preserving *local* clusters of similar documents in the resulting plot: charters with a similar stylistic profile can be expected to end up in the same region in the plot. Our discussion below will confront the clusters which emerge from this charter map with insights derived from traditional scholarship on this material.

Analysis: Text Reuse Detection

To support the discussion of the local clusters in the charter map, we have also applied a text reuse detection algorithm: early on we noted that many of the lexical similarities between charters seemed to correspond to the use of specific formulaic chunks. Text reuse detection, which singles out shared 'intertexts' or formulae between pairs of charters, yielded useful assistance in interpreting the stylistic proximity of text clusters during a close reading of the relevant documents. For this part of the analysis, we have relied on the Python *text-matcher* package developed by Jonathan Reeve (Reeve, "Text-Matcher"). Although this package has been developed for modern English-language text, much of the approach is language-independent and we have applied it here with its default settings. Future research could aim to finetune the parametrization for medieval Latin, which might certainly pay off (Manjavacas). The results below nevertheless demonstrate that this approach, even 'out of the box', already yields very

useful results. The text-matcher package internally relies on Python's *difflib*, which implements a number of established routines for string alignment. This analysis, too, was restricted to the lowercased and preprocessed *dispositio* with masked named entities.

We have paired each charter with each other charter in the corpus and the resulting 'matches' have been saved in a tabular format, accommodating the easy perusal of the results (one file per document pair). An example is offered in figure 2: in the central column, the matching passage can be found, with the textual context in the columns on either side. The sources and targets of the matches alternate in the rows. The examples illustrate that the package allows for some 'fuzziness' when detecting matches (*e.g.* minor mismatches in the orthography of words).[18] The resulting matches have been taken into account during the analysis as additional arguments in the study of the dictamen.

18. Consult the package's documentation for more details (Reeve).

left context	match	right context
udo nos ammonet ut eos propensius confouere et promouere studeamus qui religiose conuersationis propositum sectari satagunt et amplecti inde est quod	dilectis in christo filiis xxxxxxx et sociis eius religionis habitum	in ecclesia sanctę xxxxx in pratis per dei inspirationem et institutionem consilii nostri suscipientibus eorum que sucessoribus altare sancti saluato
am eos precipue plenius confouere cohortari et adiuuare tenemur qui sanctę conuersationis iter artum et arduum amplecti et tenere ordiuntur propterea	dilectis in christo filiis xxxxxxx et sociis eius religionis habitum	secundum ordinem et institutiones ęcclesię sancti xxxxxxxx xxxxxxxxxxx assumere proponentibus terram illam quę est inter ecclesiam sancti saluatoris
us religionis habitum in ecclesia sanctę xxxxx in pratis per dei inspirationem et institutionem consilii nostri suscipientibus eorum que sucessoribus	altare sancti saluatoris et personatum eius atque omnes obuentus ad eundem personatum pertinentes preter annonam quam ecclesia xxxxxxxxxxx thesaurario ęcclesię sancte xxxxx xxxxxxxxxxxx annua pensione	persoluit deuote contulimus et nomine elemosinę in perpetuum assignauimus atque in sollemni synodo nostra recognouimus ea conditione quod prefata ecc
ibus libera electione instituatur preterea bono eorumdem proposito deo inspirante admodum delectati ipsa die susceptionis ordinis et habitus canonici	altare sancti saluatoris et personatum eius atque omnes obuentus ad eundem personatum pertinentes preter annonam quam ecclesia xxxxxxxxxxx thesaurario ęcclesię sanctę xxxxx xxxxxxxxxxxx annua pensione	persoluit contulimus et nomine elemosinę pro redemptione peccatorum nostrorum in perpetuum assignauimus eo tenore quod dies obitus nostri anniuersari
is presbyterum instituere unum de canonicis suis qui jus parrochiale prouideat et episcopo ac ministris eius de jure parrochiali respondeat cura uero	eiusdem parrochię in manu abbatis remaneat quoniam uero parrochia illa ad dignitatem thesaurarię ęcclesię sancte xxxxx xxxxxxxxxxxx pertinere dinoscebatur communi assensu capituli eiusdem ęcclesię simul que adhibita	coniuentia et beniuolentia xxxxxxx thesaurarii et decani predictę donationis sollemnitatem canonice terminauimus ac eidem xxxxxx eius que successorib
instituere presbyterum unum de canonicis suis qui ius parrochiale prouideat qui episcopo et ministris eius de iure parrochiale respondeat cura tamen	eiusdem parrochię in manu abbatis remaneat quoniam uero locus predictus in quo noua fundabatur ecclesia et prenominata parrochia sancti saluatoris ad dignitatem thesaurarię ęcclesię sancte xxxxx xxxxxxxxxxxx pertinere dinoscebatur communi assensu capituli eiusdem ecclesię simul que adhibita	coniuentia ac plenissima beniuolentia xxxxxxx thesaurarii et decani predicte donationis sollemnitatem canonice terminauimus ac eidem xxxxxx ac succes
atem thesaurarię ęcclesię sancte xxxxx xxxxxxxxxxx pertinere dinoscebatur communi assensu capituli eiusdem ęcclesię simul que adhibita coniuentia et	beniuolentia xxxxxxx thesaurarii et decani predictę donationis sollemnitatem canonice terminauimus ac eidem xxxxxx eius que successoribus ad eiusdem thesaurarię titulum pleniorem restitutionem reddidimus	tria uidelicet altaria quę expressis nominibus duximus annotare altare de xxxxx altare de xxxxxxxx altare de xxxxxx cum appendiciis suis statuentes ig
rarię ęcclesie sancte xxxxx xxxxxxxxxxx pertinere dinoscebatur communi assensu capituli eiusdem ecclesię simul que adhibita coniuentia ac plenissima	beniuolentia xxxxxxx thesaurarii et decani predicte donationis sollemnitatem canonice terminauimus ac eidem xxxxxx ac successoribus eius ad eiusdem thesaurarię titulum pleniorem restitutionem reddidimus	cum celebratione obitus sui qui in predicta ęcclesia annis singulis recolendus instituetur statuentes igitur et sub anathematis interdicto precipient

Figure 2. Tabular example of the detected intertexts using text reuse detection for DiBe ID 7273 (blue) and DiBe ID 7652 (green). The overlapping text is shown in red in this example; left and right context of the match in both document can be found on either side of the matches.

3 Charter map: the case of RogF/JeanE

RogF/JeanE

This case study is situated in the context of the Cambrai episcopal chancery, a rather well-organized example of the administrative writing activity in the circle of a bishop. From its beginnings in 1057, the dictamen has been fairly well established for each individual bishop (Van Mingroot, *De bisschoppelijke kanselarij*, 693–701; Van Mingroot, *Les chartes*, 1–15). In particular, the first half of the twelfth century constitutes an interesting period with the presence of chancellor Werinbold III, whose influence largely regulated the chancery dictamen (Van Mingroot, *De bisschoppelijke kanselarij* 728; Barré, "Chancellerie épiscopale" 139; Brunel 238–240; Bedos-Rezak 120). His influence partially remained during the second half of this century. By the end of this period, the episcopal chancery became a less stable environment, as the bishop encountered political turbulence (De Moreau 76–77).

It is during this more unstable episode that the presence of one particular dictator has been previously noted, instigated by the recurrent use of the phrase *committimus … memoriali*. In the databases *Diplomata Belgica* and *Chartae Galliae*, used during a preliminary study, this formula is present in nine episcopal charters from Cambrai, dated from 1185 to 1198 (Leclercq 411).

DiBe ID	Charter description	Date	Hand
1870	Roger of Wavrin for Grand-Bigard	1185/1187	Preserved as copy
7648	Roger of Wavrin for the Templars	1187	RogF/JeanE
1804	Roger of Wavrin for Affligem	1189	Preserved as copy
1989	Roger of Wavrin for Tongerlo	1189	RogF/JeanE
2119	Roger of Wavrin for Saint-Martin of Tournai	1189	Preserved as copy
2130	Jean II of Antoing for Saint-Martin of Tournai (confirmation)	1193	Preserved as copy
3681	Jean provost and Adam dean of Notre-Dame for Anderlecht	1195	Unknown hand
3682	Jean II of Antoing for Anderlecht (confirmation)	1195	RogF/JeanE
8694	Hugues elect for Saint-Saulve of Valenciennes (confirmation)	1198	Preserved as copy

Table 1. The Cambrai charters containing '*commitimus...memoriali*'.

The dictamen of these charters is similar; in particular the phrases used in the protocol and eschatocol, combined with the *committimus memoriali* expression, hint at one possible dictator. Specifically, the charters of Roger of Wavrin for Affligem (1189), Tongerlo (1189) and Saint-Martin of Tournai (1189) show several similarities (underlined) but enough differences as well to conclude that they were not mere copies of each other (DiBe ID 1989):

In nomine domini. Rogerus diuina permissione Cameracensis episcopus tam posteris quam modernis in perpetuum. Quicquid sollemniter gestum est firmiter conualescit dum sacris litterarum annotatum apicibus thesauris memorie commendatur. Eapropter presenti committimus memoriali ut in communem noticie lucem prodeat quod pauperi ecclesie Tongerlensi altare de Ouelo sub elemosine titulo pro nostre predecessorum que nostrorum animarum salute libere contulimus ita tamen quod sorores eiusdem ecclesie apud Aiwen demorantes ter in anno cum sibi minute fuerint ex eiusdem prouentibus altaris refectione carnium contentur. Hanc igitur huius doni nostri et institute elemosine paginam ne quis mortalium attemptet infringere inhibentes in eius conseruatores diuinam diffundi benedictionem poscentes in eius uero detractores eterni penam promulgantes anathematis. Ipsam appensi caractere sigilli et subsignatione testium duximus communire. Signum Walteri et Sigeri archidiaconorum. Signum Walteri capellani. Signum Egidii de montibus. Signum Walteri Ciruiensis decani. Signum Iohannis Bruxellensis decani. Actum anno uerbi incarnati M° C° LXXX° IX° presulatus uero nostri decimo. Ego Walcherus cancellarius subscripsi et publice recognoui.

The above charter can be considered a typical example of this dictator's dictamen; the protocol formulas (*intitulatio*, address, *salutatio*, clause of corroboration, witness list, dating) adhere to chancery practice. The details of each text in the list are resumed in the analysis below. We start from the working hypothesis that a single dictator was responsible for all of these charters, with the exceptions of the charter of Jean II of Antoing for Saint-Martin of Tournai (DiBe ID 2130) and the charter of Hugues elect for Saint-Saulve of Valenciennes (DiBe ID 8694). As these are confirmations, parts of these texts could very well be copied by a dictator in service of the beneficiary.

Of the above list, four charters are preserved in their original form, of which three are written by the same hand (see figure 3). The charter issued by Jean and Adam, respectively provost and dean of Notre-Dame of Cambrai, for Anderlecht from 1195 (DiBe ID 3681), has been executed or engrossed by another hand, probably situated in the Anderlecht chapter, and is a textual copy from the episcopal charter of that same year regarding the same subject (DiBe ID 3682). From the preliminary palaeographic research, we also have the following list of charters engrossed by RogF/JeanE:

DiBe ID	Charter description	Date	Place of conservation
7648	Roger of Wavrin for the Templars	1187	Archives Nationales Paris, 5255/1
1989	Roger of Wavrin for Tongerlo	1189	Tongerlo, Abbey Archives n°37
3680	Provost Siger and dean Lambert of Anderlecht for Anderlecht	1195	Archives Départementales du Nord, 4G 704/6974
3682	Jean II of Antoing for Anderlecht	1195	Maison d'Érasme, n°6
1823	Siger d'Arras, archdeacon of Cambrai for Affligem	1196	Rijksarchief Leuven, A eccl. 4608/32

Table 2. The list of charters likely engrossed by RogF/JeanE

The charter of Siger d'Arras (DiBe ID 1823) does not contain "*commitimus… memoriali*" and is thus not included in table 1. This charter does contain a corroboration of the RogF/JeanE type: *"Nos igitur abbitrium nostrum ratum uolentes permanere et sigilli nostri appensione et testium digna subscriptione paginam presentem dignum duximus communire,"* but no additional elements we can link clearly to the established RogF/JeanE dictamen. As such, we prefer not to label this charter a RogF/JeanE product for lack of arguments. Plausibly, the other four charters of table 2 have been composed *and* engrossed by the same person, whom we term 'RogF/JeanE'; below, our aim is to assess, via stylometric means, the current working theory that dictator and scribe coincide in this specific case, and also, whether the oeuvre of this individual could potentially be enlarged.

The eight charters coined as RogF/JeanE products form our starting point for the analysis and are labeled as such in the visualizations below.

DiBe ID	Charter description	Date
1870	Roger of Wavrin for Grand-Bigard	1185/1187
7648	Roger of Wavrin for the Templars	1187
1804	Roger of Wavrin for Affligem	1189
1989	Roger of Wavrin for Tongerlo	1189
2119	Roger of Wavrin for Saint-Martin of Tournai	1189
3681	Jean provost and Adam dean of Notre-Dame for Anderlecht	1195
3682	Jean II of Antoing for Anderlecht (confirmation)	1195
3680	Provost Siger and dean Lambert of Anderlecht for	1195

Table 3. The list of charters considered as RogF/JeanE products at the start of the analysis

Figure 3. The hand RogF/JeanE in Tongerlo, Abbey Archives n°37.
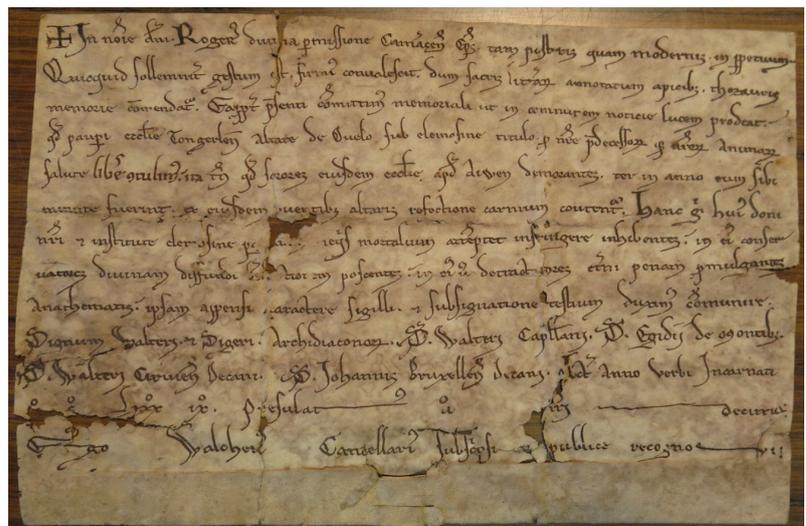
19. *Sigeri de Atrebato archidiaconorum* (DiBe ID 1870); *Sigeri archidiaconi* (DiBe ID 7648; DiBe ID 1804. DiBe ID 1989. DiBe ID 2119. DiBe ID 8694); *Sigeri [...] presbiterorum* (DiBe ID 3681. DiBe ID 3682); *Sigerus dei gratia Cameracensis archidiaconus* (DiBe ID 1823); *Sigeri [...] praesbiterorum* (DiBe ID 3682).

20. We arrive at the following list for the second half of the twelfth century: 1160: deacon Siger (*Chartae Galliae* Charter n° 212748); 1162: *magister* Siger (DiBe ID 4779); 1185-1186: archdeacon Siger of Arras (DiBe ID 1870. DiBe ID 1823); 1187-1205: archdeacon Siger, mentioned 1189 with the additional title of *magister* (DiBe ID 496. DiBe ID 2799. DiBe ID 496. DiBe ID 13757). An archdeacon Siger is mentioned in the Notre-Dame of Cambrai obituary on the twenty-fourth of October (Municipal Library of Cambrai, B 229, f°79v°); 1195 : the priest Siger (DiBe ID 3682). A *Segardus presbiter* is mentioned on the 24th of March in the Notre-Dame of Cambrai breviary (Municipal Library of Cambrai, B 46, f°2v°). The thirteenth-century cathedral chapter obituary shows *Sygerus presbiter* on the sixteenth of February, but this mention has been erased (Municipal Library of Cambrai, B 1161 f.12r.); 1195: *magister* Siger, provost of Saint-Peter's of Anderlecht (DiBe ID 3680). It is quite probable that a *magister* mentioned in 1162, the archdeacon and the *magister* mentioned from 1185 until 1205 refer to the same person.

21. DiBe ID 1981. DiBe ID 3518. DiBe ID 3685. DiBe ID 3686. DiBe ID 3517.

22. DiBe IDs 1872, 1803, 4482, 2886, 31032, 8692, 2902, 10897, 10898, 13045, 13563, 34033, 32274, 14484 and 14662.

The possible identity for this individual could be found in the person of a certain *Siger*, as this name is present in all the charters linked to RogF/JeanE.[19] We have compiled a list of all Sigers – a widespread name during this period – in the Cambrai charters and in the necrologic documentation.[20] A potential candidate for the identification of RogF/JeanE, is Siger of Arras, archdeacon of Cambrai, mentioned in the *Diplomata Belgica* charters from 1187 (DiBe ID 1872) to 1211 (DiBe ID 14662). He is specifically mentioned as *magister* in 1201 (DiBe ID 13101) and 1211 (DiBe ID 14662). Potentially, Siger moved to the Anderlecht chapter and took up the role of provost as well. A Siger is mentioned in the witness lists of charters pertaining to members of the Saint-Peter's chapter from 1173 onward,[21] and specifically as *magister* in 1175 (DiBe ID 3686). The provost Siger even issued a charter destined for Kortenberg in 1195 (DiBe ID 10476). All originals written by RogF/JeanE are issued by the bishop or by a Siger, the provost in 1195 and the archdeacon in 1196.

As the 1196 charter (DiBe ID 1823) has been issued in this Siger's own name, and is of the hand of RogF/JeanE, it could reinforce the argument that all these references are to a single person. Unfortunately, save the corroboration we are uncertain of its status as a RogF/JeanE product. This charter does contain a corroboration of the RogF/JeanE type: "*Nos igitur abbitrium nostrum ratum uolentes permanere et sigilli nostri appensione et testium digna subscriptione paginam presentem dignum duximus communire.*" The charters in the *Diplomata Belgica* database which mention this archdeacon (1187-1211) additionally present a collection of writing styles that seemingly differ from RogF/JeanE; other dictators were probably involved in their conception[22].

Analysis of the charter map

The t-SNE visualization is shown in figures 4 and 5 with different colourings. In figure 4, the charter material is presented using colour codes encoding the different origins, *i.e.* charters issued by the beneficiaries and by the Cambrai bishops' charters. The scatter plot shows, for example, the papal charters (purple areas) on each side of the scatterplot, which have a very distinct dictamen. The charters for Ename abbey (blue areas), too, are distinctly regrouped.

Figure 4. Two-dimensional t-SNE scatter plot visualization of the 545 x 545 Euclidean distance matrix, obtained on the basis of a TF-IDF normalized bag of words model (for a vocabulary of 1,000 words, after masking). The colours refer to the episcopal charters (teal) versus the various other authorities. The episcopal charters are dispersed through the complete visualization as they have textual links to the various institutions the chancery wrote for.
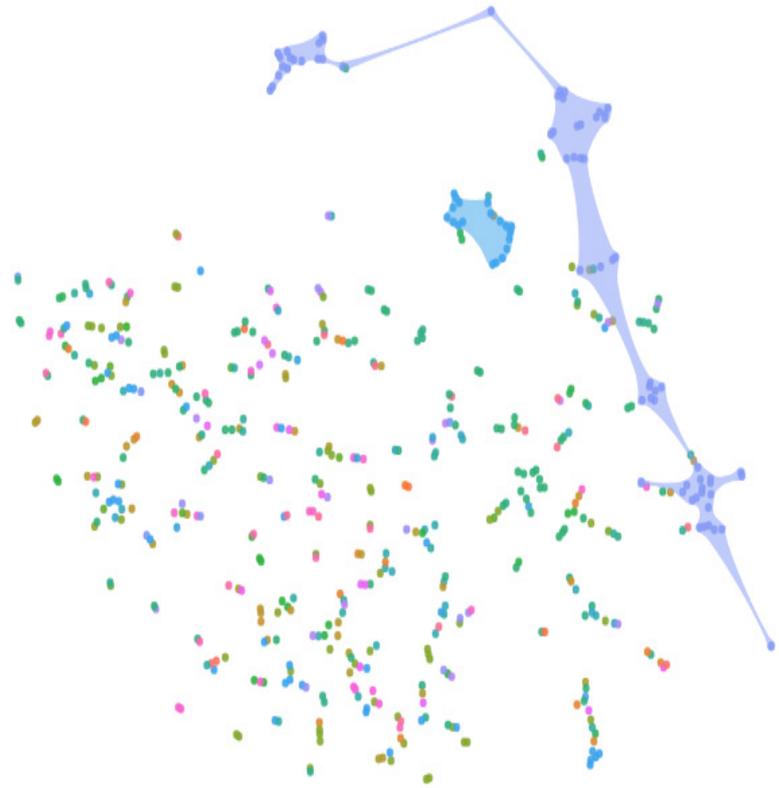
When looking closer at our dictator-scribe RogF/JeanE, the charters hypothetically attributed to him are located, surprisingly close to one another, within a group of Cambrai bishops' charters (see figure 5). To corroborate this impression of proximity, we collected all rows in the distance matrix, tentatively attributed to RogF/JeanE and paired them exhaustively to calculate the Euclidean distance between them. Next, we did the same for all other charters that were not attributed RogF/JeanE. The latter operation would of course yield much more distances, but nevertheless allows us to assess whether the distance between the hypothetical RogF/JeanE documents is indeed smaller than that for two random charters in the dataset. The boxplot visualization in figure 6 (first and second row) does suggest this to be the case.

The more detailed analysis of the charter texts below is organized in five, manually assigned clusters (see figure 7). Cluster A contains multiple charters linked to the abbey of Saint-Martin of Tournai. Cluster B includes multiple charters linked to the Saint-Peter's chapter of Anderlecht, as well as a few other charters. Cluster C consists almost exclusively of Cambrai bishop's charters. Cluster D presents a rather heterogeneous group which will not be discussed in detail in this paper as these texts show no connection to RogF/JeanE. Finally, cluster E comprises two bishop's charters that demand our attention in particular.

Figure 5. Same as figure 4, but with the RogF/JeanE charters explicitly labeled. Colouring based on the different beneficiaries, with the exception of the Cambrai episcopal charters (teal).

Figure 6. Boxplot visualization of the Euclidean distances between rows in the distance matrix, for all document pairs from two categories: RogF/JeanE and all documents not currently attributed to RogF/JeanE.
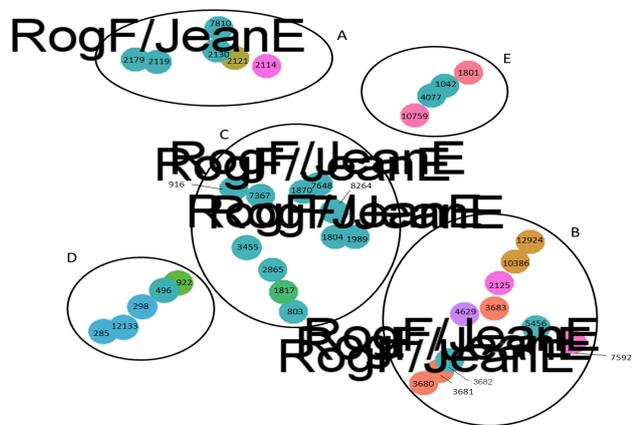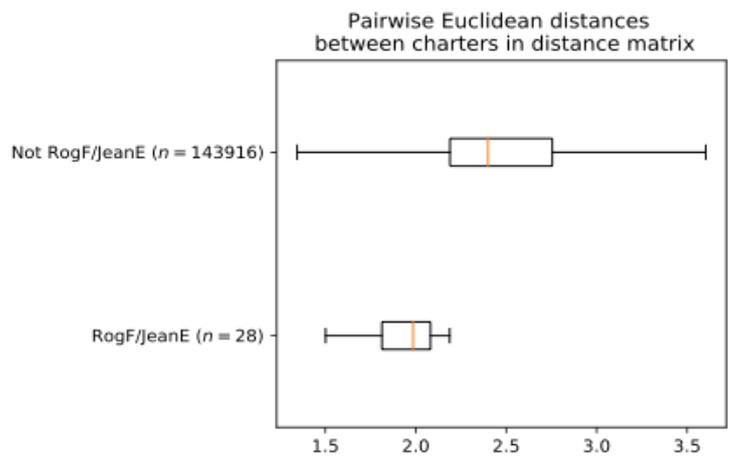

Figure 7. Detail of the visualisation featuring the *Diplomata Belgica* references.

Cluster A

In the first cluster, featured in the top left corner of this image, we can find the following charters:

Table 4. The description of the charters in cluster A.

DiBe ID	Charter description	Date
2114	Guillaume I of Champagne Archbishop of Reims for Tournai (Saint-Martin)	1186
2179	Roger of Wavrin for Mons (Sainte-Waudru)	1186
2121	Jean provost and Hugues dean of Cambrai (Notre-Dame) for Tournai (Saint-Martin)	1189
2119	Roger of Wavrin for Tournai (Saint-Martin)	1189
7810	Jean II of Antoing for Saint-Amand	1193
2130	Jean II of Antoing for Tournai (Saint-Martin)	1193

The presence of multiple charters linked to Saint-Martin of Tournai, issued by different authors, is immediately clear. Three of these charters can be neatly regrouped. The charter of Guillaume I of Champagne (DiBe ID 2114) shows only a tenuous link to the other texts. When comparing this charter to the other Cambrai texts, its dictamen does not feature the Cambrai characteristics. Rather, it contains rather general phrases.[23] These appear in multiple charters coming from different dioceses. The longer sentence "*presentis scripti patrocinio … audeat infringere*" is present in six other corpus texts, of which three are issued by this archbishop[24]. As such, the dictamen might be linked more to the Archbishop of Reims. This charter is linked to the other Saint-Martin of Tournai texts, as well as to a Vaucelles and an Affligem charter, through parts included in the clause of corroboration.[25] The other Saint-Martin of Tournai texts are connected through their subject.

The charter of Jean provost and Hugues dean of Notre-Dame of Cambrai for Saint-Martin of Tournai, dated 1189 (DiBe ID 2121), ratifies the donation of the Saint-Martens-Lierde, Beloeil and Ormegnies altars, donated the same year by Roger of Wavrin to Saint-Martin of Tournai (DiBe ID 2119). The charter of Jean II of Antoing from 1193 confirms the donation (DiBe ID 2130). These three texts share a particular phrase: "*altare de Lierde sancti Marini et altare de Bailluel cum altari de Ormegnies.*» It is valuable to note that the text reuse detection result seems very generic: "*xxxxxxx xxxxxxxxxxx altare de xxxxxx sancti xxxxxxx et altare de xxxxxxxx cum altari de*," yet it has yielded an interesting connection. The phrase "*altare de Lierde sancti Martini et Deiamicus clericus noster altare de Bailluel cum altari de Ormegnies*" can also be found in a charter of Roger of Wavrin for Saint-Martin of Tournai from 1189, which announces the donation to the deans of Chièvres and Grammont (DiBe ID 2120). This text is not presented near this cluster A: the remaining dictamen is indeed different.

The charter issued by Jean provost and Hugues dean of Notre-

23. "*ad quos littere iste peruenerint in domino salutem,*" "*nouerit uniuersitas uestra,*" "*presentis scripti patrocinio et sigilli nostri impressione confirmamus statuentes … ne quis hanc nostre confirmationis paginam audeat infringere.*"

24. DiBe ID 1677. *Chartae Galliae* charter n°211866. *Chartae Galliae* charter n°212636. *Chartae Galliae* charter n°212637. *Chartae Galliae* charter n°212640. *Chartae Galliae* charter n°212641.

25. "*Ne quis hanc nostre confirmationis paginam audeat infringere uel*" in the charter of Guillaume I of Champagne for Vaucelles, 1192. DiBe ID 36258. "*Presentis scripti patrocinio et sigilli nostri auctoritate confirmamus statuentes et*" in the charter of Guillaume I of Champagne for Affligem, 1188. DiBe ID 1677.

26. *"Eapropter nouerit uniuersitas paginam presentem inspicientium quoniam dilectissimus episcopus noster Rogerus pro salute anime sue et suorum concessit monasterio et fratribus."*

27. *"Hanc ergo donationem concedentes et per omnia ratam habentes presenti scripti subnixam esse et uolumus et sigilli nostri appensione corroborare dignum duximus.»*

28. *"Sacris litteram apicibus facta perpetuantur sollemnia et cauillationes littium ne aduersus ea suscitentur in posterum sopiuntur. Presenti eapropter memoriali committimus quod altare de Lierde sancti Marini et altare de Bailluel cum altari de Ormegnies que predecessor noster dominus Rogerus bone memorie Cameracensis episcopus assensu capituli sancte Marie Cameracensis ecclesie sollemniter et canonice sub elemosine titulo contulit."*

29. *"Uniuersis ad quos littere presentes peruenerint salutem in domino"*

30. *"Nulli ergo omnino hominum liceat hanc paginam nostre confirmationis et diffinitionis infringere uel ei temere obuiare salua auctoritate apostolice sedis et metropolitan"*

31. DiBe ID 4480, DiBe ID 1786, DiBe ID 2115, DiBe ID 809, DiBe ID 780, DiBe ID 779, DiBe ID 735, DiBe ID 732, DiBe ID 2257, DiBe ID 2823, DiBe ID 7248, *Chartae Galliae* charter n°201279, DiBe ID 1819.

Dame of Cambrai (DiBe ID 2121) shows a part of the *notificatio* close to the episcopal model;[26] the clause of corroboration in particular fits this mold.[27] The charter of Jean II of Antoing from 1193 (DiBe ID 2130) almost completely copies the preamble from the original donation by Roger of Wavrin.[28] These two texts were not drafted by RogF/JeanE. They show proximity due to their shared subject, as well as the few chancery characteristics mentioned above. They are featured closer to each other than to the Roger of Wavrin charter for Saint-Martin from 1189 (DiBe ID 2119) to which these two charters refer. This last charter has been established during the preliminary dictamen-research as a typical charter of RogF/JeanE, of which we have detailed the text above while introducing our dictator.

The charter issued by Roger of Wavrin for Sainte-Waudru of Mons, from 1186 (DiBe ID 2179), is shown closer to the typical dictamen of RogF/JeanE as well. This text contains an episcopal dictamen, especially in the clause of corroboration linked to the penal clause, which is close to the dictamen of RogF/JeanE: *"Hanc igitur iuste et honeste donationis nostrę et elemosinę paginam ne quis mortalium ausu temerario attemptet infringere sub anathematis interminatione districtius inhibentes et ut ratior perseueret eam sigilli nostri appensi karactere et sufficienti testium subsignatione duximus communire."* The protocol, the use of *eapropter* in the beginning of the *dispositio*, the use of *"pro nostra et predecessorum nostrorum Cameracensium Episcoporum"* as well as *"misericorditer … concessimus et firmauimus"* and finally the dating clause point to a RogF/JeanE dictamen. As such, the stylometric analysis has provided us with a first addition to the list of charters considered as RogF/JeanE products.

Finally, this cluster includes the charter of Jean II of Antoing for Saint-Amand, from 1193 (DiBe ID 7810). This charter's text shows chancery elements which differ from the RogF/JeanE dictamen, especially in the protocol[29] and the penal clause.[30] The formula *"Nulli ergo omnino hominum liceat hanc paginam nostre confirmationis"* has been found by the text reuse detection in twelve papal charters and one charter by Count Baudouin IX of Constantinople.[31] It is not exactly clear why the analysis has placed this charter closely to the charter issued by Jean II of Antoing for Saint-Martin of Tournai in 1193 (DiBe ID 2130): the general episcopal dictamen could be the defining factor.

Cluster B

In this second cluster, three Anderlecht charters are put close to each other by the applied algorithms and could well be considered prod-

DiBe ID	Charter description	Date
4629	Godefroid III Duke of Brabant for the Templars	1181
5456	Roger of Wavrin for Bellingen	1182
12924	Gautier abbot of Arrouaise for Arrouaise	1183
10386	Jean provost and Hugues dean of Cambrai (Notre-Dame) for Arrouaise	1188
2125	Jean abbot of Tournai (Saint-Martin) for Roger of Wavrin	1189
3680	Sigerus provost and Lambert dean of Anderlecht for church of Anderlecht	1195
3681	Jean provost and Adam dean of Cambrai (Notre-Dame) for Anderlecht	1195
3682	Jean II of Antoing for Anderlecht	1195
3683	Henri I Duke of Brabant for Anderlecht	1195
7592	Baudouin IX of Constantinople Count of Flanders-Hainaut for Valenciennes (Saint-Saulve)	1196

Table 5. The description of the charters in cluster B

ucts of RogF/JeanE. The charter issued by Jean provost and Adam dean of Notre-Dame of Cambrai, dated 1195 (DiBe ID 3681), concerns the transfer of the Anderlecht altar and all its possessions; this charter is almost completely copied from the one issued by Jean II of Antoing the same year (DiBe ID 3682). The same legal act is confirmed by the charter from Sigerus provost and Lambert dean of Anderlecht (Dibe ID 3680). Finally, a fourth charter intended for Anderlecht has been issued by the Duke of Brabant Henri I the same year (DiBe ID 3683): the text confirms this same transfer and copies a few parts of the dictamen.[32] The remainder of the dictamen differs from the formulas used by RogF/JeanE.

The charter presented next to it, issued by the Duke of Brabant Godfrey III for the Templars in 1181 (DiBe ID 4629), is probably shown closer to the Henri I charter due to a possibly ducal dictamen. This could indicate the existence of some form of ducal chancery; this assumption seems to be incorrect for the twelfth century however: according to Michel Margue there was in the Mosan region no chancery linked to a prince in the strict sense; at the same time, true beneficiary products seem rare as well, as the prince could count on clerics, sometimes connected to one of the Hausklöster, who were charged with the writing of the charters (Margue 227, 235 and 242).

These two ducal charters do not show particular common formulas, with the exception of the *intitulatio* "*dei gratia dux Lotharingiae.*" A full study of the duke's possible dictamen does not fit within the confines of this paper; we can at least identify clear traces of a non-episcopal dictamen. As such, a redaction by RogF/JeanE is improbable. It is possible that the vectorial representation of the second ducal charter (DiBe ID 4629) is geometrically close to the RogF/JeanE texts due to the nature of our corpus: after all, the complete ducal charter corpus is not included, thus, ducal texts present in the visualization are scattered. More general formulas (in particular "*munimine roborauimus*" and "*sigilli nostri appensione*") can be noted in a few other ducal charters present in the databases.[33]

32. "*in publicam prodire uolumus notitiam, sub annua pensione 40 librarum Cameracensis monetae*" and "*inconuulsum manere uolentes sigilli nostri appensione.*"

33. DiBe ID 16388. DiBe ID 29782. DiBe ID 34161. DiBe ID 18107. DiBe ID 29772. *Chartae Galliae* charter n°211898.

The two charters studied next show an episcopal dictamen: the charter issued by Jean abbot of Saint-Martin of Tournai for Roger of Wavrin, dated 1189 (DiBe ID 2125), and the one issued by Jean provost and Hugues dean of Notre-Dame of Cambrai for Arrouaise, dated 1188 (DiBe ID 10386). These texts were not included in the preliminary dictamen study, as the bishop takes up the role of beneficiary and author respectively and not the one of issuer. It is certainly possible these documents were written in the episcopal chancery as the prelate played a prevalent role in their juridical conception, or was at least the superior authority implied in the legal act.

The charter by Jean abbot of Saint-Martin of Tournai (DiBe ID 2125) contains a distinctly episcopal dictamen: "*in nomine domini,*" "*tam modernis quam posteris in perpetuum,*" the use of "*eapropter*" at the beginning of the *dispositio,* "*libere et absolute contradidimus in usus,*" "*Actum anno uerbi incarnati.*" These elements, in particular combined with the corroboration "*Hanc igitur huius doni nostri paginam inconuulsam in euum manere uolentes eam nostre ymagine sigilli et sub signatione testium duximus roborare,*" relate to a dictamen by RogF/JeanE, and constitute our second addition.

The presence of an episcopal dictamen in the charter by Jean provost and Hugues dean of Notre-Dame for Arrouaise (DiBe ID 10386) is hardly surprising. After all, the chapter charters and the episcopal charters are considered to be produced by the same group of people (Barré 129-146). This case here shows multiple chancery characteristics in the eschatocol[34] as well as at the end of the text.[35] The dictamen is not particularly evocative of that of RogF/JeanE, while it is probably a chancery product.

This last charter shows a link to the charter issued in 1183 (DiBe ID 12924) by Gautier abbot of Arrouaise for that same institution, only through the use of the phrase "*rata habemus,*" rather frequently found in the *Diplomata Belgica* and the *Chartae Galliae* databases.[36] The only other connection remains their shared beneficiary. It is possible the computer code has classified these two texts close together due to the appearance of a literal citation in a non-exhaustive corpus. In addition, the Gautier abbot charter is very short and does not allow for much comparison.

The two last charters in the B cluster prove to be problematic. The charter issued by Roger of Wavrin for Bellingen in 1182 (DiBe ID 5456) and the charter issued in 1196 by the Count of Flanders Baudouin IX of Constantinople for Saint-Saulve of Valenciennes (DiBe ID 7592) show no resemblances in their dictamen, nor do they share their subject

34. "*In nomine sancte et indiuidue Trinitatis, amen. Iohannes, Dei gratia Beate Marie Cameracensis.*»

35. "*Presens scriptum appenso sacro beate Uirginis karactere muniuimus et testes legitimos ac sufficientes subscripsimus*" and "*Actum anno Uerbi Incarnati.*"

36. We can report 118 occurrences of "*rata habemus*" in the *Diplomata Belgica* from 1160 to 1250 (as well as a forgery from 1628 but dated 642) of which 22 cases are Cambrai episcopal charters (1178 tot 1250 and the aforementioned forged charter). The *Chartae Galliae* report 100 occurrences between 1145 and 1295, none in Cambrai charters. Accessed 16th Jan. 2019.

or beneficiary (DiBe ID 5456. DiBe ID 7592). The charter from Roger of Wavrin showcases a few episcopal characteristics, such as "*notum igitur sit tam posteris quam modernis*" shared with the charter from Roger of Wavrin for Aymeries, dated 1179 (DiBe ID 10910). This last text is represented in one of the large collections of Vaucelles charters, on the opposite side of the visualization. One would rather expect to find these two charters closer to one another on the graph. Whether this is due to a 'deeper' comparison than the simple reuse of words is currently unclear.

Cluster C

DiBe ID	Charter description	Date
2865	Roger of Wavrin for Ninove	1179
3455	Roger of Wavrin for Kortenberg	1180
7367	Roger of Wavrin for Hautmont	1185
916	Roger of Wavrin for Ename	1185
1870	Roger of Wavrin for Grand-Bigard	1185
8264	Roger of Wavrin for Arras (Saint-Jean)	1186
7648	Roger of Wavrin for the Templars	1187
1804	Roger of Wavrin for Affligem	1189
1989	Roger of Wavrin for Tongerlo	1189
803	Jean II of Antoing for Ghent (Saint-Pierre)	1192
1817	Lambert provost and Christine provost of Mons (Sainte-Waudru) for Forest	1195

Table 6 - The description of the charters in cluster C.

All charters included in this cluster show characteristics of the episcopal chancery. Four of these texts have indeed been identified as RogF/JeanE products during the preliminary dictamen research: the charters issued by Roger of Wavrin for Grand-Bigard in 1185 (DiBe ID 1870), for the Templars in 1187 (DiBe ID 7648), for Affligem in 1189 (DiBe ID 1804) and for Tongerlo the same year (DiBe ID 1989). These texts contain the phrase considered typical of this dictator: "*commitimus… memoriali*." In the Affligem charter (DiBe ID 1804) we find the use of the phrase "*diem obitus nostri patris que nostri (ac) matris,*" equally found in a charter from this dictator included in cluster A, the charter issued by Roger of Wavrin for Saint-Martin of Tournai in 1189 (DiBe ID 2119), which as discussed above showcases the dictator's typical dictamen.

These four texts are organized around Roger of Wavrin's charter for Saint-Jean of Arras, dated 1186 (DiBe ID 8264), which shows many similarities to the dictamen of RogF/JeanE and is considered our third addition. "*Presenti eapropter memoriali in omnium noticiam diffundi*" is close to the charter issued in the same year by Roger of Wavrin for Sainte-Waudru of Mons (DiBe ID 2179), equally presumed to be a RogF/JeanE product. "*Paginam rata manere volentes in ejus conserva-*

37. *Chartae Galliae* charter n°211854.

38. *Chartae Galliae* charter n°211872.

tores diffundentes benedictionem" is also found in a charter from Roger of Wavrin for Vaucelles, dated 1189 (DiBe ID 10760).[37] This last writ is presented on the other side of the visualization, in a group of charters connected to Vaucelles, next to the charter issued by Jean II of Antoing for Vaucelles in 1193 (DiBe ID 10772).[38] Finally, the penal clause *"eterni penam promulgantes anathematis"* is equally present in the charter issued by Roger of Wavrin for Tongerlo in 1189 (DiBe ID 1989) and the one issued by the same prelate for Hautmont in 1185 (DiBe ID 7376).

This last charter is shown close to the four RogF/JeanE products and clearly contains an episcopal dictamen. Even though no absolute argument such as *"commitimus…memoriali"* is present to rule this text a RogF/JeanE product, we consider this to be the case. This brings our current total of additions to four.

The charter issued by Roger of Wavrin for Ename in 1185 (DiBe ID 916) is shown closest to the Hautmont text (DiBe ID 7376) and contains chancery characteristics as well. It is certainly not drafted by RogF/JeanE, in particular when we look at the use of *"dei uoca-tione"* in the intitulatio and *"pie fauentibus cauterio anathematis in-festantes,"* both phrases this dictator never used as far as we know. The preamble contains *abiecta mundi sarcina*, a phrase linked by B.-M. Tock, D. Van de Perre and J. Barrow to a reform movement started in the abbey of Saint-Martin of Laon (Van de Perre 56; Tock, "Les chartes de fondation" 159-163 ; Barrow 26). This element is also found in the charter issued by Roger of Wavrin for Kortenberg in 1180 (DiBe ID 3455); this last one does show a few of the RogF/JeanE elements but not enough to be considered one of his products.

The charter for Ninove from 1179 (DiBe ID 2865) was already considered as a chancery product by M. Van den Noortgate (Van den Noortgate 167-168), in which we observe multiple episcopal chancery characteristics, especially in the penal clause combined with the corroboration. No specific dictamen elements currently point at RogF/JeanE as its dictator. The remaining two charters (DiBe ID 1817 and DiBe ID 803) probably are not drafted/composed by RogF/JeanE, but they do display a few chancery characteristics.[39]

JeanB: a second dictator-scribe?

Finally, even though cluster E does not add to our knowledge of charters written by RogF/JeanE, and thus will not be discussed at the same depth as cluster D, the group merits our attention due to two Cambrai charters which feature a chancery dictamen which clearly

39. The charter from Lambert provost and Christine provost of Sainte-Waudru of Mons for Forest (DiBe ID 1817) contains *"tam presentibus quam futuris in perpetu-um," "libera traditione contulimus"* and a corroboratio modeled on the chancery type. The charter issued by Jean II of Antoing to Saint-Pierre of Ghent from 1192 (DiBe ID 803) contains *"Johannes dei gratia Cameracensis episcopus,"* the corroboratio *"nullus ergo hanc nostrę donationis paginam infringere uel ipsi temere contraire presumat"* and finally the subscription and datation *"Data per manum magistri Danihelis Cameracensis ecclesię canonici anno uerbi incarnate."*

differs from the RogF/JeanE model. They (DiBe ID 1042 et 4077) are clearly aligned due to their shared expressions, in particular in the clause of corroboration: "… *nostrum presentis pagine patrocinio et nostri sigilli … confirmantes et ne quis … sub periculo … inhibentes.*" The use of "*suscepte administrationis*" is limited to these two charters in the *Diplomata Belgica* database as well as in this corpus according to the text reuse detection results. This element could explain their proximity. In the *Chartae Galliae* database, we observe this same expression in a papal charter, from Celestinus III for the abbey of Saint-Jean-des-Vignes, dated 1197 (*Chartae Galliae* charter n° 210698).

As these two charters have been transcribed by the same hand (fig. 8), "JeanB" (1194–96), we consider the possibility of determining another case where a scribe and dictator coincided in the Cambrai chancery.

In the visualization, two other charters by this hand are located in each other's close vicinity: one issued by the cathedral chapter of Notre-Dame of Cambrai in 1195 and one issued by the Holy Roman Emperor Henry VI in 1196, in favour of Jean II of Antoing (DiBe ID 7373. DiBe ID 8686). A second charter issued by Henry VI in the same year, intended for Jean II of Antoing and the abbey of Saint-André-du-Cateau, is from the same hand (DiBe ID 8687). This charter was not included in our test corpus as the Cambrai bishop is not its author nor its beneficiary; as a confirmation of the 1196 Cambrai charter, however, the text is almost a complete copy.
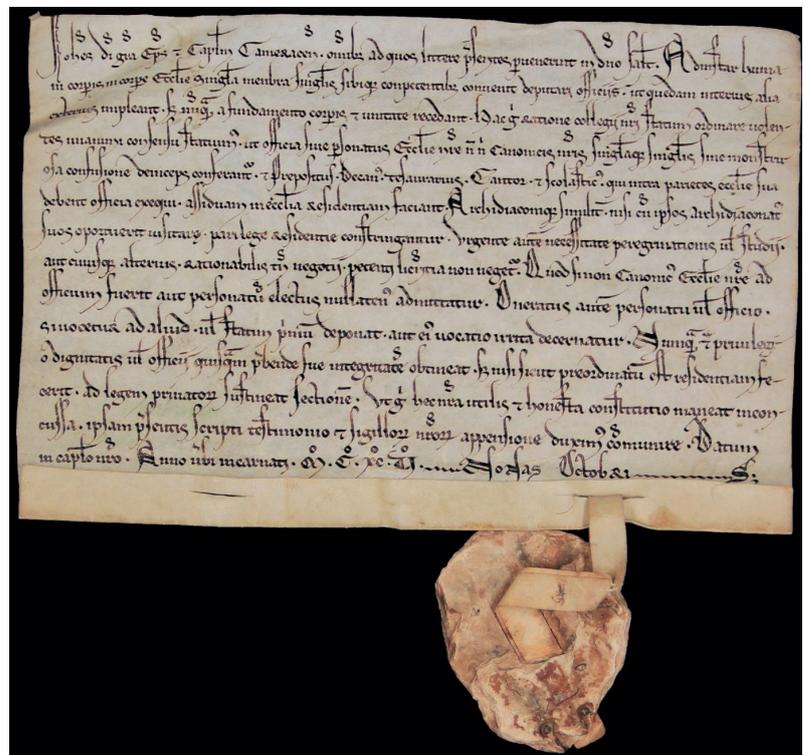
Figure 8. The hand JeanB in Archives Départementales du Nord, 4 G 58/796.

The four JeanB originals are represented in the same region of the visualization:
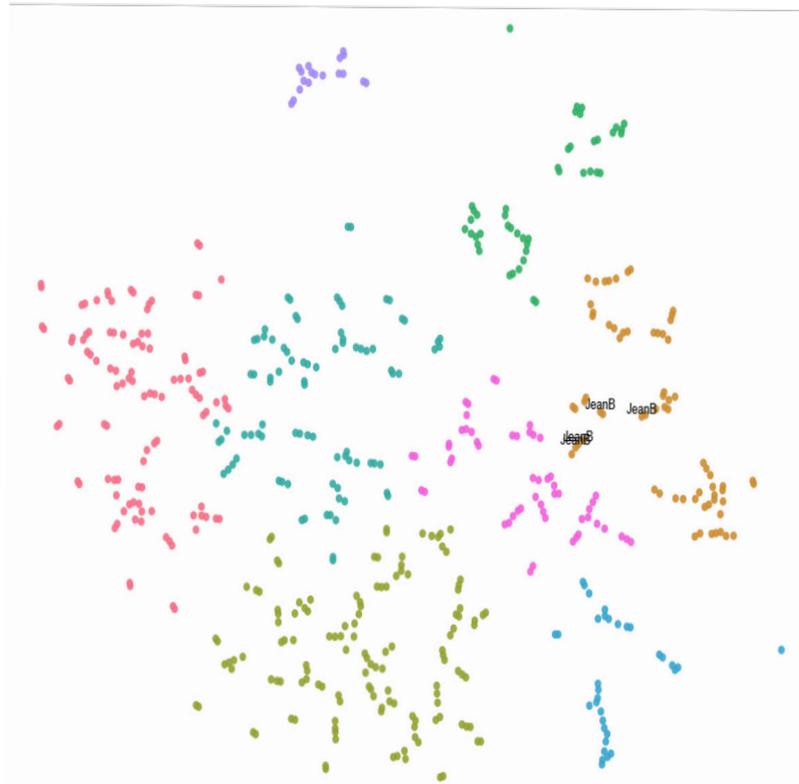
Figure 9. The vicinity of the charters of JeanB in the complete visualization as shown in figure 3. Colouring based on a clustering protocol (8 clusters) that assigns the same colour to nearby points, as a reading aid.

Figure 10. Detail of the charters transcribed by JeanB.
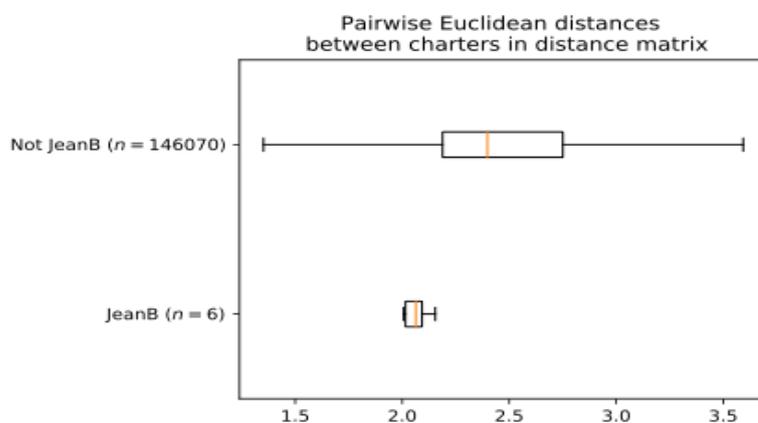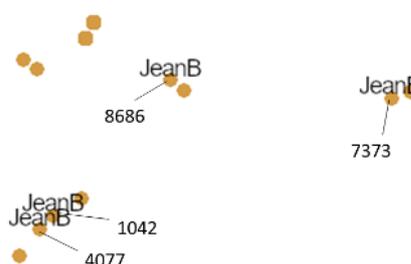
Figure 11. Boxplot visualization of the Euclidean distances between rows in the distance matrix, for all document pairs from two categories: JeanB and all documents not currently attributed to JeanB.

40. These texts are the two 1195 charters as well as one issued by Jean II of Antoing for Ename in 1194 (DiBe ID 950), probably transcribed by the hand JeanA.

In figure 13, we have also included the pairwise distance for the JeanB group (as did for RogF/JeanE in figure 6): these show a distribution that is similar to RogF/JeanE, with smaller than average distances. A short comparison of their dictamen reveals a few textual links. "*Omnibus ad quos littere presentes peruenerint*" is present in two of these charters, both dating from 1195: the one issued by the cathedral chapter (DiBe ID 7373) and the one issued by Jean II of Antoing (DiBe ID 4077). In the *Diplomata Belgica*, this expression is found in 23 charters ranging from 1194 to 1241, of which three are Cambrai charters.[40] In the *Chartae Galliae* this exact wording can also be found in the charter issued in 1175 by papal legate Pierre of Saint-Chrysogonus for Saint-Bertin and in the one issued in 1214 by Aubry archbishop of Reims for Arrouaise (*Chartae Galliae* Charter n° 210234. *Chartae Galliae* charter n° 201367). The *Diplomata Belgica* includes this phrase in archiepiscopal charters from Reims starting in 1186 (DiBe ID 2340). The exact penal clause "*Hanc nostre confirmationis paginam audeat infringere aut ei in aliquo temeraria presumptione contraire*" is present in the two charters issued in 1196 by Emperor Henry VI, as well as in a charter issued in 1192 by Guillaume I of Champagne, of which the hand is extremely close to JeanB (DiBe ID 36258). This close hand can also be found in two other charters issued by the same archbishop (DiBe ID 7455; DiBe ID 7606).

The two Cambrai charters from cluster E seem to have been written by the same person, who differs from RogF/JeanE. While the hypothesis of a single dictator-scribe for the charters transcribed by JeanB is certainly credible, a closer study of the Cambrai charters of which the scribes are known seems necessary. Such a study should, in particular, pay attention to the texts of which the Cambrai bishop is the beneficiary, or implied in the legal action (disposant) rather than the authority issuing it.

Comparison of the digital text analysis with the established diplomatic method has allowed for the confirmation and the expansion (bold) of the determined RogF/JeanE dictamen in the clusters A, B and C, which has led to the following list of charters produced by this same dictator:

DiBe ID	Charter description	Date
7367	**Roger of Wavrin for Hautmont**	**1185**
1870	Roger of Wavrin for Grand-Bigard	1185
8264	**Roger of Wavrin for Arras (Saint-Jean)**	**1186**
2179	**Roger of Wavrin for Mons (Sainte-Waudru)**	**1186**
7648	Roger of Wavrin for the Templars	1187
2125	**Jean abbot of Tournai (Saint-Martin) for Roger of Wavrin**	**1189**

1804	Roger of Wavrin for Affligem	1189
1989	Roger of Wavrin for Tongerlo	1189
2119	Roger of Wavrin for Tournai (Saint-Martin)	1189
3681	Jean provost and Adam dean of Cambrai (Notre-Dame) for Anderlecht	1195
3682	Jean II of Antoing for Anderlecht	1195
3680	Sigerus provost and Lambert dean of Anderlecht for Anderlecht	1195

Table 7. The expanded list of plausible RogF/JeanE products

As for further confirmation as to whether we should consider this individual to have been the dictator as well as the scribe of these charters, we have no access to supplementary original charters from these additions; therefore, this identification must remain a working theory for the time being. When we return to the problem of identifying RogF/JeanE with a name, a Siger is mentioned again in two of these additions:

- *magistri Sigeri Anderlectensis ecclesie prepositi* (DiBe ID 3680)
- *Signum Walteri et Sigeri Cameracensium archidiaconorum* et *Sygeri […] presbiterorum* (DiBe ID 2125)

In spite of this complementary information, the identification must remain uncertain.

4 A new approach?

In this paper, our contributions have been the following. We have used stylometry to assess the oeuvre of single dictator-scribe (RogF/JeanE): our quantitative analyses not only corroborated pre-existing theories about the products of this scribe, but have also enabled us to semi-automatically enlarge his oeuvre via new, highly plausible attributions. Secondly, our analyses have pointed towards the possible existence of another, similar, yet clearly distinct, oeuvre in the vicinity of RogF/JeanE. Tentatively, we have attributed this group of charters to 'JeanB' and we have called for further research into the matter by experts in this domain. Moving beyond this specific case study, we would now like to discuss, in conclusion, some of the wider implications of applying stylometric methods to diplomatic sources. Currently, even though researchers are increasingly developing digital tools to assist in such an endeavour, the use of statistical and stylometric methods is far from mainstream in the field of diplomatics. In what way, then, could this approach add to the current scholarly practice, now and in the future?

Much like Nicolas Perreaux, we believe that the development of new methods need not exclude existing practices. The largest ad-

vancement concerns increased efficiency: panoramic visualizations allow the diplomatist to extract starting points for the study of the dictamen and certainly implies saving time in the detailed analysis of the charter texts. The matches yielded by text reuse detection have additionally provided welcome support in interpreting the results. Analysing the frequencies of phrases and word combinations would allow to better determine their 'weight' in the corpus, as Perreaux has done in part through lemmatization. De Paermentier's method does this to a large extent as well, albeit still in a manual fashion, with word groups being tested against the full *Diplomata Belgica* Database. Automating this process through text reuse detection is a significant step forward into saving time during the analysis as well as a more statistical approach allowing for clear numbers. How often certain formulas appear in the texts, and in which period and/or region, could tell us more about the dictamen of a certain group, how far formulas spread, about standard formulas shared by the dictators of a particular writing office, and thus allow for a study of 'influences'. This would also allow to spot 'outliers' such as the *commitimus … memoriali* for RogF/JeanE more easily. Perreaux is working in this direction in observing "the lexical dynamics of the scriptorium of the abbey. We can thus identify the periods in which the writing of acts evolved most strongly, the breaks, the concentration or, conversely, the variability of the forms" (Perreaux, "Possibilities").

The expansion of the corpus, the explorative assessment of other case studies and the paleographic comparison of scribal hands should allow fine-tuning the method and determining more precisely to what degree and which stylometric techniques could be of use to the diplomatist. Especially the question of corpus bias needs to be reviewed, as the influence of charters by other great authorities distorts the results to a certain degree (*i.e.* the clear regrouping of the papal charters). It could be more interesting to compare episcopal charters of multiple dioceses with each other, to allow less bias from 'higher' authorities as well as a better basis for comparison. Enlarging the corpus to test this could be possible with Perreaux's *Cartae Europae Medii Aevi*, which is set to be made available to the public shortly.

Furthermore, the use of methods based on lexical features could take into account the grammatical grasp of dictators and help in the determination of which charter was written by whom. Elements such as the use of literary devices could maybe even come in to play (the use of rhetorical devices, references to other texts such as the Bible and the Church fathers …) and maybe even the presence of rhymed

prose. Another element to address would be the influence of the charter type. In literature, writing style could be influenced by the genre: charters also have a genre typology. In the current test, the vast majority of texts concerns the types of donations and confirmations, but adding the dimension of charter types in future tests could certainly be interesting.

In conclusion, the present paper demonstrates the feasibility of using a stylometric method on a reasonably large charter corpus. We currently cannot speak of an end-to-end 'Distant Diplomatics', as the detailed analysis clearly still asks for manual corroboration via close reading, in order to suggest plausible explanations as to why certain texts are placed in each other's vicinity. This case study has thus served as a first foray into the possibilities of using stylometry on charters for detailed *dictamen* research; especially the automation provided by text reuse detection could prove to be a valuable addition to the diplomatist's toolbox.

Bibliography

Barré, Nathalie. "Chancellerie épiscopale, chancellerie canoniale: unicité ou pluralité des institutions à Cambrai au XIIᵉ siècle ?" *Bulletin de la Commission royale d'Histoire*, 176.2 (2010): 129–46.

Barrow, Julia. "Les évêques et leur conception de la réforme ecclésiastique dans l'Europe du Nord-Ouest (XIᵉ-XIIᵉ siècles)." *Autour de Lanfranc (1010-2010): Réforme et réformateurs dans l'Europe du Nord-Ouest (XIᵉ - XIIᵉ siècles): Colloque international de Cerisy 29 septembre-2 octobre 2010*. Ed. Julia Barrow, Fabrice Delivré and Véronique Gazeau. Caen: Presses universitaires de Caen, 2015. 21–32.

Bedos-Rezak, Brigitte. *When Ego was Imago: Signs of Identity in the Middle Ages*. Leiden and Boston: Brill, 2010.

Brunel, Ghislain. "Chartes et chancelleries épiscopales du Nord de la France au XIᵉ siècle." *A propos des actes d'évêques: hommage à Lucie Fossier*. Ed Michel Parisse Nancy: Presses universitaires de Nancy, 1991. 227–44.

Cárcel Ortí, Maria Milagros. *Vocabulaire international de diplomatique.* Valéncia: Universitat de Valéncia, 1997.

Chartae Galliae. Web. Accessed 6th July 2020.

Culler, Jonathan. "Presupposition and Intertextuality." *MLN. Comparative literature* 91.6. (1976): 1380–96.

De Hemptinne, Thérèse, Jeroen Deploige, Jean-Louis Kupper, Walter Prevenier, eds. *Diplomata Belgica*. *Les sources diplomatiques des Pays-Bas méridionaux au Moyen Âge*. Bruxelles: Commission royale d'Histoire, since 2015. Accessed 6th July 2020.

De Moreau, Edouard. *Histoire de l'Église en Belgique, tome III: l'Église féodale 1122–1378*. Bruxelles: Ed. Universelle, 1946.

De Paermentier, Els. "In cuius rei testimonium et firmitatem. Oorkonden en kanselarijwerking in de entourage van de graven en gravinnen van Vlaanderen en Henegouwen (1191–1244). Een diplomatische en paleografische studie." Ph.D. dissertation, Universiteit Gent, 2011.

Diederich, Joachim, Jörg Kindermann, Edda Leopold and Gerhard Paass. "Authorship Attribution with Support Vector Machines." *Applied Intelligence* 19.1–2(2003): 109–23.

Dierkens, Alain and David Guilardian. "Actes princiers et naissance des principautés territoriales: du duché de Basse-Lotharingie au duché de Brabant (XIe - XIIIe siècles)." *Bulletin de la Commission royale d'Histoire* 176.2 (2010): 243–58.

Dijkhof, Everardus Cornelis. *Het Oorkondenwezen van enige kloosters en steden in Holland en Zeeland, 1200–1325.* Leuven: Peeters, 2003.

Gervers, Michael and Michael Margolin. "Managing Meta Data in a Research Collection of Medieval Latin Charters." München, 2007. Web. Accessed 8th November 2021.

Hirst, Graeme and Olga Feiguina. "Bigrams of Syntactic Labels for Authorship Discrimination of Short Texts." *Literary and Linguistic Computing* 22.4 (2007): 405–17.

Kestemont, Mike. "What Can Stylometry Learn From Its Application to Middle Dutch Literature?" *Journal of Dutch Literature* 2.2 (2011): 46–65.

---. "Stylometry for Medieval Authorship Studies. An Application to Rhyme Words." *Digital Philology: A Journal of Medieval Cultures* 1/1 (2012): 42–72.

Kestemont, Mike, Sara Moens and Jeroen Deploige. "Collaborative Authorship in the Twelfth Century: a Stylometric Study of Hildegard of Bingen and Guibert of Gembloux." *Digital Scholarship in the Humanities* 30.2 (2015): 199–224.

Koppel, Moshe, Jonathan Schler and Shlomo Argamon. "Authorship Attribution in the Wild." *Language Resources and Evaluation* 45.1 (2011): 83–94.

Leclercq, Eveline. "L'élaboration des chartes médiévales. L'exemple des évêchés d'Arras, Cambrai et Liège (XIe–XIIe siècles)." Ph.D. dissertation, Université de Strasbourg, 2019.

Luyckx, Kim and Walter Daelemans. "The Effect of Author Set Size and Data Size in Authorship Attribution." *Literary & Linguistic Computing* 25 (2010): 35–55.

Margue, Michel. "Actes princiers et naissance des principautés territoriales: chartes et pouvoirs laïques dans les espaces mosan et mosellan (fin Xe – début XIIe siècle)." *Bulletin de la Commission royale d'Histoire* 176.2 (2010): 219–44.

Municipal Library of Cambrai. B 229.

---. B 46.

---. B 1161.

Perreaux, Nicolas. "L'écriture du monde (I). Les chartes et les édifices comme vecteurs de la dynamique sociale dans l'Europe médiévale (VIIe – milieu du XIVe siècle)." *Bulletin du centre d'études médiévales Auxerre* 19/2 (2015). Web. Accessed 6th May 2021.

---. "L'écriture du monde (II). L'écriture comme facteur de régionalisation et de spiritualisation du mundus: études lexicales et sémantiques." *Bulletin du centre d'études médiévales Auxerre* 20.1 (2016). Web. Accessed 6th May 2021.

---. "Possibilities, Challenges and Limits of a European Charters Corpus (Cartae Europae Medii Aevi – CEMA)." 2021. Archive ouverte HAL. Accessed 10th May 2021.

Prevenier, Walter. *De oorkonden der graven van Vlaanderen (1191 – aanvang 1206).* Vol. 1. Bruxelles: Palais des Académies, (1964).

Reeve, Jonathan. "Text-matcher." Accessed 9th Feb. 2019.

Rybicki, Jan and Maciej Eder. "Deeper Delta Across Genres and Languages: Do We Really Need the Most Frequent Words?" *Literary and Linguistic Computing* 26.3 (2011): 315–21.

Sanderson, Conrad and Simon Guenter. "Short Text Authorship via Sequence Kernels, Markov Chains and Author Unmasking: An Investigation." *2006 Conference on Empirical Methods in Natural Language Processing (EMNLP 2006), University of Technology, Sydney, Australia, 22-23 July 2006.* Eds. Dan Jurafsky and Eric Caussier. Sydney: Association for Computational Linguistics, 2006. 482–91.

Sebastiani, Fabrizio. "Machine Learning in Automated Text Categorization." *ACM Computing Surveys* 34.1 (2002): 1–47.

Spiegel, Gabrielle. *Romancing the Past. The Rise of Vernacular Prose Historiography in Thirteenth-Century France.* Berkeley: University of California Press, 1993.

Stamatamos, Efstathios. "Masking Topic-related Information to Enhance Authorship Attribution." *Journal of the Association for Information Science and Technology* 69.3 (2018): 461–73.

Tock, Benoît-Michel. "Les chartes de fondation des abbayes de Vicoigne (1129) et de Château-Dieu (1155), dans le diocèse d'Arras." *Analecta Praemonstratensia* 63 (1987): 159–63.

---. "Auteur ou impétrant ? Réflexions sur les chartes des évêques d'Arras au XIIe siècle." *Bibliothèque de l'École des Chartes* 149 (1991): 215–48.

Van den Noortgate, Machteld. "Diplomatisch-paleografisch onderzoek van de oudste oorkonden van de Sint-Cornelius- en Cyprianusabdij van Ninove (1137–1300)." Master thesis, Universiteit Gent, 1990.

Van de Perre, Dirk. "De oudste bisschoppelijke en grafelijke oorkonden van de abdij van Ninove (1138–1167)." *Analecta Praemonstratensia* 77 (2001): 54–93.

Van der Maaten, Laurens and Hinton, Geoffrey. "Visualizing Data using t-SNE." *Journal of Machine Learning Research* 9 (2008): 2579–2605.

Van Mingroot, Erik. "De bisschoppelijke kanselarij te Kamerijk (1057–1130)." Ph.D. dissertation, Katholieke Universiteit Leuven, 1969.

---. *Les chartes de Gérard Ier, Liébert et Gérard II, évêques de Cambrai et d'Arras, comtes du Cambrésis (1012–1092/93).* Leuven: Leuven University Press, 2005.