

# How to detect institutional and regional feature clusters in late medieval charters?

## Collaboration between more and less digital humanists in the project BeCoRe

### Barret, Sébastien

sebastien.barret@irht.cnrs.fr

Institut de recherche et d'histoire des textes, Centre national de la recherche scientifique, France

### Helias-Baron, Marlène

m.helias@irht.cnrs.fr

Institut de recherche et d'histoire des textes, Centre national de la recherche scientifique, France

### Stutzmann, Dominique

dominique.stutzmann@irht.cnrs.fr

Institut de recherche et d'histoire des textes, Centre national de la recherche scientifique, France

### Tscherne, Niklas

niklas.tscherne@uni-graz.at

Austrian Centre for Digital Humanities, University of Graz

### Vogeler, Georg

georg.vogeler@uni-graz.at

Austrian Centre for Digital Humanities, University of Graz

### Schindler, Jacqueline

jacqueline.schindler@gmx.at

Niederösterreichisches Institut für Landeskunde, Austria

Charters and deeds are texts recording juridical acts, and as such, they made use of a vast array of strictly defined signs of validation and authenticity; but their makers also sought to express their authority and validity through more “subtle” signs, devoid of “direct” juridical functions, for example layout, certain types of scripts, or ornamental elements. This was key to social integration, and to their efficacy. These methods have been studied with digital tools on small scales by manually curated databases (Models of Authority 2015-2017, Broun 2015, Webber 2015, Stokes et al. 2016, Stokes 2020) or primarily as a data collection (NOTAE project 2023, Ghignoli 2019). The project “Between Composition and Reception: The Authority of Medieval Charters” (BeCoRe, FWF-I-4502, ANR-19-CE27-0021) studies this phenomenon on a large scale and comparatively. One of our hypotheses is that such elements developed in regional frames, which specialists in mediaeval charters have named “Diplomatic landscapes” (Härtel 2015) and that this geographic influence was stronger than institutional

relationships on a wider geographic scale - which, for the time being, constitutes a theoretical framework, and not yet a result, of our endeavours. We study this in the combination of human experts in diplomatics and data science methods.

We identify recurring features in a large-scale documentation and relate them to individual observations; these are among others layout features (document size, margins, interlinear space) that are traditionally used by chanceries to highlight solemnity (Bischoff 1996): the formation of the final lines (e.g. Worm 2004), the use of specific signs such as monograms, ornate capitals and chancery or notarial signs (for a general view, Rück 1996); but also the script itself (for an example in royal charters, Gasparri 1973), or the use of spaces and punctuation marks. In order to achieve its goals, BeCoRe uses exemplary, digitised corpora of charters originating in French and Austrian monastic archives through the digital platform (Heinz 2010, Vogeler 2019, Monasterium [2023]), through which we use semantic features, which have to be detected via region detection and classification methods (e.g. YOLO5), in order to test the commonly accepted knowledge on the signs used for expressing power in charters, and their possible combination: do they only appear together, or does their use define a kind of “hierarchy of the visibility of power”?

The project is a good example for data re-use: It will add new data to Monasterium.net from existing data sets (CBMA 2023, Fontenay 2023, Gallica 2023 IIIF), enhance this data and the existing data in Monasterium to comply with the description standards needed for the project. Charters are transcribed with the help of Transkribus, and encoded with the established standards in Monasterium.net (Charters Encoding Initiative [2023], Vogeler 2006). This description, including text-image alignment, creates a solid basis for further database queries and in particular through machine learning and clustering of the relevant elements.

The project focuses first and foremost on the graphical level, but also takes into account the textual and typological aspects. In practice, this involves close and distant reading methods. Close reading will create solid data and sample cases on the relationship between regional and super-regional structures, like those we reported at the Digital Diplomatics 2022 conference (Bergk Pinto et al. 2022). Distant reading methods will apply to both image classification and object detection. We hope to compare the results of different architectures (e.g. ResNet) and pipelines (e.g. YOLO). We will then proceed to investigate whether the use of unsupervised clustering algorithms (such as k-means clustering) will result in diplomatically useful groupings of charters by graphical and textual features, along the lines of geographical regions, affiliation of monastic order, monastery, or others, and whether such a process is likely to generalise beyond the material we have selected. Here, also, we intend to compare the results of different algorithms.

Given that one of the major recent challenges within the field of diplomatic research has been the sheer amount of surviving research material, we hope that the cooperation between automated machine learning techniques of this kind and human interpretation can prove to be a fruitful approach toward facing this problem.

The poster will present the state of the project and discuss what role machine learning methods can have for the human interpretation of late mediaeval charters as means to convey authority. As one of its focus is on collaboration, it will include a visualisation of our workflow for data use and reuse.

## Bibliography

- Ares Legaspi, Adrián** (2021): “La presentación gráfica de las fórmulas diplomáticas en documentos y códices de Santiago de Compostela en los siglos XIV y XV”, in: Louviot, Élise / García, Charles / Morrison, Stephen (eds.): *La Formule au Moyen Age / Formulas in Medieval Culture IV*, Turnhout: Brepols: 85-109.
- Bergk Pinto, Nicole / Schindler, Jacqueline / Tscherne, Niklas** (2022), “BeCoRe–Between Composition and Reception”, in: *Digital Diplomats Conference*, Graz, 28-30 Sept. 2022: forthcoming, programme and abstract: <https://didip.hypotheses.org/conference-2022/digital-diplomats-2022-program> [28.04.2023].
- Bischoff, Frank M.** (1996): *Urkundenformate im Mittelalter. Größe, Format und Proportionen von Papsturkunden in Zeiten expandierender Schriftlichkeit (11.-13. Jahrhundert)*. Marburg: Institut für historische Hilfswissenschaften.
- Broun, Dauvit** (2015): “Introducing the ‘Models of Authority’ Project: Scottish charters c.1100–c.1250 | Models of Authority”, in: *Models of Authority* (blog), 23. Juli 2015: [28.04.2023].
- CBMA–Corpus Burgundiae Medii Aevi** (2023): <http://www.cbma-project.eu/> [28.04.2023].
- Charters Encoding Initiative** [2023]: <https://www.cei.lmu.de> [28.04.2023].
- Fontenay data set** (2023): <https://github.com/oriflamms/Fontenay> [28.04.2023].
- Gallica**–Digital publication platform of the French National Library (2023): <https://gallica.bnf.fr> [28.04.2023].
- Gasparri, Françoise** (1973): *L’écriture des actes de Louis VI, Louis VII et Philippe Auguste*. Geneva: Droz / Paris: Minard.
- Ghignoli, Antonella** (2019): “The NOTAE Project: A Research between East and West, Late Antiquity and Early Middle Ages”, in: *COMSt Bulletin* 5: 29–42.
- Härtel, Reinhard** (2015): “Urkundenlandschaften zwischen Donau, Rhein und Adria”, in: Schwarcz, Andreas / Kaska, Katharina (eds.): *Urkunden– Schriften– Lebensordnungen. Neue Beiträge zur Mediävistik*. Wien / Graz / Köln: Böhlau: 293–211.
- Heinz, Karl** (2010): “Monasterium.net: Auf dem Weg zu einem europäischen Urkundeportal”, in: Kölzer, Theo / Rosner, Willibald / Zehetmayer, Roman (eds.): *Regionale Urkundenbücher. Die Vorträge der 12. Tagung der Commission Internationale de Diplomatique*. St. Pölten: Niederösterreichisches Institut für Landeskunde: 139–45.
- Models of Authority**, Scottish Charters and the Emergence of Government 1100–1250 (2015–2017): <https://www.modelsofauthority.ac.uk/> [28.04.2023].
- Monasterium** ([2023]): <http://www.monasterium.net> [28.04.2023].
- NOTAE project**–Not a Written Word, (2023): <http://www.no-tae-project.eu/> [28.04.2023].
- Rück, Peter** (ed.) (1996): *Graphische Symbole in mittelalterlichen Urkunden. Beiträge zur diplomatischen Semiotik*. Sigmaringen: Thorbecke.
- Stokes, Peter** (2020): “La paléographie et la diplomatique numériques pour l’exploitation d’un corpus de chartes: l’exemple de Models of Authority”, in: Canteaut, Olivier / Guyotjeannin, Olivier / Poncet, Olivier (eds.): *Actes royaux et princiers à l’ère du numérique (Moyen Âge-Temps modernes)*. Pau: Presses universitaires de Pau et des Pays de l’Adour: 71–78.
- Stokes, Peter / Brookes, Stewart / Noël, Geoffroy / Davies, John / Webber, Teresa / Broun, Dauvit / Taylor, Alice / Tucker, Joanna** (2016): “The Models of Authority Project: Extending the DigiPal Framework for Script and Decoration”, in: *Digital Humanities 2016: Conference Abstracts*, Kraków: Jagiellonian University & Pedagogical University, 11–16 July 2016: 896–99. <https://dh2016.adho.org/abstracts/387> [28.04.2023].
- Vogeler, Georg** (2006): “Charters Encoding Initiative (CEI): Zu Möglichkeiten der Integration mit Hilfe eines Standards für Urkundendigitalisierung”, in: Aigner, Thomas / Winter, Karin (eds.), *Alte Archive– Neue Technologien*. St. Pölten: Diözesanarchiv: 181–198.
- Vogeler, Georg** (2019): “Monasterium.net–eine Infrastruktur für diplomatische Forschung”, in: *Das Mittelalter* 24, 1: 247–52 [28.04.2023].
- Webber, Teresa** (2015): “An Introduction to the Project ‘Models of Authority: Scottish Charters and the Emergence of Government, 1100–1250’”, in: *Scripta: An international Journal of Codicology and Palaeography* 8: 177–180.
- Worm, Peter** (2004): *Karolingische Rekognitionszeichen. Die Kanzlerzeile und ihre graphische Ausgestaltung auf den Herrscherurkunden des achten und neunten Jahrhunderts*. Marburg: Institut für historische Hilfswissenschaften.