CATMuS-Medieval: Consistent Approaches to Transcribing ManuScripts

Ariane Pinche, Thibault Clérice, Alix Chagué, Jean-Baptiste Camps, Malamatenia Vlachou-Efstathiou, Matthias Gille Levenson, Olivier Brisville-Fertin, Federico Boschetti, Franz Fischer, Michael Gervers, et al.

CATMuS - Medieval: Consistent Approaches to Transcribing ManuScripts

A generalized set of guidelines and models for Latin scripts from Middle Ages (8^{th}-16^{th})

Ariane Pinche[1], Thibault Clérice[2], Alix Chagué[2,3,4], Jean-Baptiste Camps[5, 6], Malamatenia Vlachou-Efstathiou[4], Matthias Gille Levenson[5, 1], Olivier Brisville-Fertin[1, 6], Federico Boschetti[7, 8], Franz Fischer[8], Michael Gervers[9], Agnès Boutreux[9], Avery Manton[9], Simon Gabay[10], Patricia O'Connor[5], Wouter Haverals[11], Mike Kestemont[12], and Caroline Vandyck[12]

[1]CIHAM–UMR 5648, CNRS
[2]ALMAnaCH - Automatic Language Modelling and Analysis & Computational Humanities, Inria, Paris, France
[3]UdeM - Université de Montréal, Montréal, Canada
[4]EPHE - École Pratique des Hautes Études, Paris, France
[5]CJM - Centre Jean Mabillon, Paris, France
[6]ÉNC - École nationale des chartes, Paris, France
[7]ILC-CNR
[8]VeDPH - Venice Centre for Digital and Public Humanities, Ca'Foscari, Venice, Italy
[9]UToronto - Department of History, University of Toronto, Canada
[10]UNIGE - Université de Genève, Switzerland
[11]Princeton University
[12]Antwerp University, Belgium

December 2023

Introduction

In 2016, the OCRopy tool (Breuel 2014) –designed as an open-source, easily trainable OCR engine– piqued the interest of digital humanists in deploying its application to manuscripts Camps 2017. While the majority of the community explored its usefulness for *incunabula* and *fraktur* texts (Springmann et al. 2018), we delved into its potential beyond printed material. Between 2016 to 2019, the landscape of handwritten text recognition (HTR) underwent a profound transformation. The advent of general public GPUs for model training became feasible, and the availability of programming libraries like Tensorflow and PyTorch became standard. Notably, during this period, Kraken (Kiessling 2022) gained prominence, while Transkribus (Kahle et al. 2017) emerged as a comprehensive solution, enabling manual annotation, model training and post-correction.

Along with the growing popularity of applications such as Transkribus and eScriptorium (Stokes et al. 2021), the use of HTR has seen significant growth in the digital humanities. Each project varies widely and is often inclined to tailor its own transcription methods based on specific objectives: from automatically transcribing cartularies (Stutzmann et al. 2018) to extracting literary manuscript texts for further study (Camps, Clérice, and Pinche 2021; Haverals and Kestemont 2023). The need for a standardization of methods and datasets became evident in 2021 with the production of generic HTR models able to embrace diversified collections. The divergence in transcribing practices across projects prompted the creation of a set of unifying guidelines for transcribing Old French manuscripts from the 10^{th} to the 15^{th} century (Pinche 2022b). The collective dataset expanded through external additions, culminating in the development of the first generic model for Medieval manuscripts: the CREMMA medieval model (Clérice, Pinche, and Vlachou-Efstathiou 2022a). Subsequently, fuelled by multiple funding sources and collaborations, both the guidelines and models are extending their reach to encompass various languages, with a coming expansion for Old English and Middle Dutch.

The Consistent Approaches to Transcribing Manuscripts (CATMuS) guidelines, dataset, and model are the result of an international collaboration aimed at standardizing transcription practices for historic documents in Western Europe. CATMuS currently includes past and ongoing projects with collaborators from Europe and North America: France (CREMMA, HTRomance (Clérice, Chagué, et al. n.d.), Fabliaux (Pinche and Pierreville 2023), Liber (Aruta et al. 2023), Gallic(orpor)a, HTRogène[1]), Italy (HTRomance, HTRogene), Switzerland (Gallic(orpor)a, FoNDUE), Canada (DEEDS(Gervers, Boutreux, and Manton 2023)). While CATMuS aims at merging transcription practices for HTR across documents from the Middle Ages to

[1]This dataset is not yet public but collaborators were actively involved in the previous dataset.

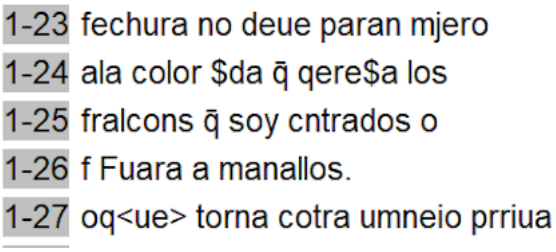


Figure 1: Prediction using Transkribus *Coloso Español* model. The model uses two different ways to deal with abbreviations, line 24 ‹q̄› and line 27 ‹<ue>›. Capture kindly provided by J. M. Fradejas Rueda.

contemporary times, this paper specifically focus on the medieval and 15-16$^{\text{th}}$ centuries early print (incunabula and print with gothic typefaces) from the datasets developed around the CATMuS Guidelines to help design general model(s) for HTR.

1 Guidelines

Transcribing involve a nuanced balance between fidelity to the source and interpretive choices. For texts in Anglo-Norman, for example, the historians' editions have neither accents nor apostrophes: we thus read *Dangleterre* and *labbe*, instead of modern *d'Angleterre* and *l'abbé* (Breuil 2019, Introduction, p. 5). On the contrary, French literary texts editions expect standardized punctuation, word segmentation, and accents. Moreover, while a human reader can handle the lack of homogeneity, it is a significant challenge for computers. The lack of standardization inflates character counts and introduces ambiguity. For example, the ‹p› with a stroke could be represented by the ‹ꝑ› (Unicode A751)or by a ‹ꝑ› (0554) from the Armenian alphabet. Such variations introduce noise during model training, leading to inconsistent predictions. Such issues have been found regarding a generic diachronic model for Spanish called *Coloso español* on Transkribus (Fradejas 2023) (cf. Figure 1).

In tandem with a seminar held at the École nationale des Chartesthe first guidelines for medieval French emerged, addressing documents spanning from the 10$^{\text{th}}$ to 15$^{\text{th}}$ centuries (Pinche 2022b). These guidelines aim to assist medievalists in translating manuscripts into a computer-readable format and in unifying their practices by:

1. restricting the characters used for specific purposes according to the MUFI[2];

[2] `https://mufi.info/q.php?p=mufi`.

2. applying a graphematic transcription[3];

3. keeping abbreviations (and thus, reducing the part of the language specific traits the model has to learn).

Decisions on abbreviations, diacritics, and phonetic representations aim to minimize characters in the HTR model. In medieval sources, distinctions like ‹u›/‹v› and ‹i›/‹j› that are typically variations in form rather than distinct phonetic realizations are reduce to ‹u› and ‹i›.

Ongoing efforts include addressing new transcription requirements arising from diverse language domains, illustrated by challenges introduced by Latin in the CREMMA project and Spanish in the HTRomance project. The addition of Latin as added new difficulties as the representation of specific abbreviations for morphological suffix representations such as ‹ꝛ̧› for ‹rum› or ‹;› for ‹bus›. The future addition of Middle English introduces the imperative need for new signs, like ‹ð› (eth) representing the voiced and voiceless dental fricative ([θ] and [ð]).

The establishment of transcription rules remains an ongoing and evolving effort, crucial for ensuring data quality and homogeneity. To ensure a better accessibility to the guidelines, the project aims at building a new set of guidelines, more interactive and more multilingual than the current PDF, with a delivery planned for the beginning of 2024.

2 Data

The CATMuS Medieval 1.0.0 dataset is built upon 17 different repositories across 10 different projects for which the guidelines have been applied, and for which Choco-Mufin character normalization and quality control have been applied Clérice and Pinche 2021. The corpus (c. 115.000 lines and 3.4M characters) is heavily influenced by its two first languages: Old French and Latin, and the existence of three personal projects[4] which produces the three overrepresentation of French-Narratives (29.2% of the dataset), Treatises-Latin (24.2%) and Treatises-Castillan (23%, *cf.* Figure 2a). Speeches, Epistolary and Drama are three under-represented genres but it is not expected that their abbreviation rate or vocabulary differs much from the Narrative and Poetry genres.

Languages from the modern area of Italy and Spain (Castilian, Catalan, Navarrese and Venitian) are mostly underrepresented as only two projects are providing

[3]Transcription preserving letter sequences and reducing each form's meaning to an alphabetical system (Stutzmann 2011)

[4]Namely the PhDs and MA thesis of Pinche (2021), Gille Levenson (2023b), and Vlachou-Efstathiou (2023).

(a) by genre

(b) by script

Figure 2: Relative importance (in characters percentage) of categorical features in all CATMuS Medieval 1.0.0 training languages. Bars are accumulation of percentage per categorical feature.

ground-truth. The final corpus contains 180 different documents-hands, of which 20 are prints and 34 are microfilms. Its most represented centuries are the 15th century (74 document-hands), the 14th (45), and the 13th (34). The least represented centuries (8th, 10th, 11th with a total of 6 manuscripts) can only be represented by Latin manuscripts and Old English manuscripts, which reduces the opportunities to accumulate data.

3 Evaluation

For CATMuS Medieval 1.0.0 (Pinche, Clérice, et al. 2023), we provide a test based on out-of-domain manuscripts (see Table 1). Data were transcribed using the previous *CREMMA Generic 1.0.1* model and post-corrected in the eScriptorium interface by the same annotators than the training data. *Cambridge, CCC, ms. 041* represents one exception, as its language is completely out of domain (Old English and Middle Dutch is not in our training set), and its annotator has joined the initiative after the training of the model. Two competitor models are provided as baselines: the CREMMA Cortado mode (Pinche 2022a), a monolingual model trained on Old French (no character count available), and the CREMMA Generic model (Clérice, Pinche, and Vlachou-Efstathiou 2022b), trained on Latin and Old French (1.357 M characters).

Test results (*cf.* Figure 3) shows a large upgrade from the previous monolingual

model (CREMMA Cortado) and a relative upgrade compared to the bilingual model *CREMMA Generic*. The later outperforms our larger models on most of the Old French test set ($< 2\%$) except for *Paris, BnF, fr. 1553*. Looking at this specific manuscript, we hypothesize that its digitization (lower quality, a little blurry, grey-scale microfilm) provided noise for our original model which our new model, trained with more noisy digitization, was able to overcome. On all other languages (except Old English and Middle Dutch), the model severely outperforms the original model. For English specifically, we can see that moving from a monolingual to a multilingual model provides an important performance gain, while the language barrier remains stable for multilingual model, despite having more languages (+2) and more data (around 2 million more characters).

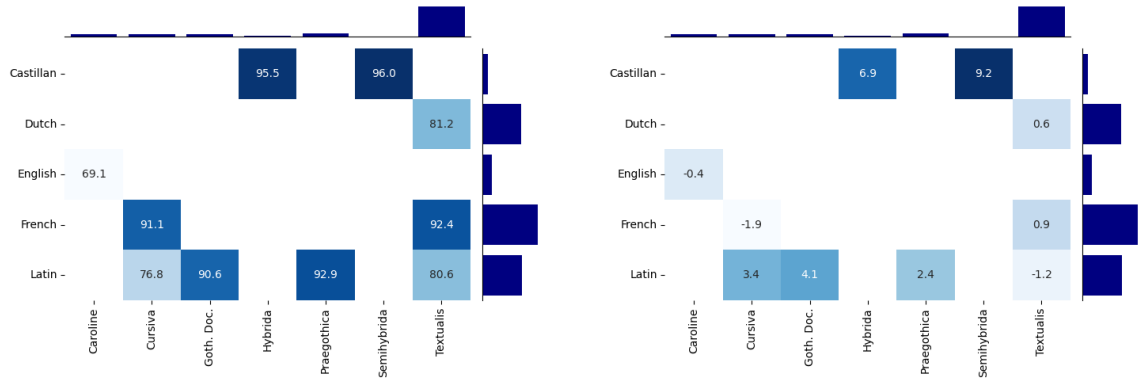
Overall, the new multilingual model provides a lower standard deviation across accuracies and a better average accuracy (Table 2), providing as such a better "generic" model according to this test set. The improvement from the Cortado model remains a huge leap, while the improvement from CREMMA Generic (and the detailed loss for some manuscripts) provides a picture of what can be enhanced in the near future.

Model					CATMuS 1.0		CREMMA Cortado			CREMMA Generic 1.0.1		
	Scripta	**Lang.**	**Cent.**	**Char.**	**Errors**	**Accuracy**	**Errors**	**Accuracy**	**Diff.**	**Errors**	**Accuracy**	**Diff.**
Paris, BnF, fr. 12558	Textualis	French	14	17258	1021	94.08%	1669	90.33%	3.75%	913	94.71%	-0.63%
Paris, BnF, fr. 1443	Textualis	French	13	13233	1820	86.25%	2308	82.56%	3.69%	1710	87.08%	-0.83%
Paris, BnF, fr. 1450	Textualis	French	13	17930	2280	87.28%	3419	80.93%	6.35%	2416	86.53%	0.76%
Paris, BnF, fr. 1553	Textualis	French	13	13510	348	97.42%	1978	85.36%	12.07%	1500	88.90%	8.53%
Paris, BnF, fr. 17229	Textualis	French	13	15027	1002	93.33%	1434	90.46%	2.87%	835	94.44%	-1.11%
Paris, BnF, fr. 6447	Textualis	French	13	16309	624	96.17%	1563	90.42%	5.76%	581	96.44%	-0.26%
Paris, BnF, fr. 840	Cursiva	French	14	7564	672	91.12%	1491	80.29%	10.83%	528	93.02%	-1.90%
Paris, BnF, lat. 14354	Textualis	Latin	14	24498	4744	80.64%	7619	68.90%	11.74%	4442	81.87%	-1.23%
Dublin, TCD, ms. 524	Praegothica	Latin	13	9500	726	92.36%	2159	77.27%	15.08%	1013	89.34%	3.02%
Toronto, Fisher Library, ms. 01053	Goth. Doc.	Latin	14	10216	2368	76.82%	3955	61.29%	15.53%	2718	73.39%	3.43%
Cambridge, CCC, ms. 111	Praegothica	Latin	11	4943	414	91.62%	1264	74.43%	17.20%	590	88.06%	3.56%
London, BL, Egerton 3031	Goth. Doc.	Latin	12	14799	1387	90.63%	4488	69.67%	20.95%	2001	86.48%	4.15%
Mâcon, Archiv. Dep., G 443	Praegothica	Latin	13	8455	494	94.16%	1158	86.30%	7.85%	589	93.03%	1.12%
Madrid, BNE, ms. 3995	Semi-hybrida	Castil.	15	2842	113	96.02%	967	65.97%	30.05%	375	86.81%	9.22%
Madrid, BNE, ms. 6406	Hybrida	Castil.	14	1702	207	87.88%	584	65.81%	22.07%	385	77.46%	10.42%
Madrid, BNE, ms. 12732	Hybrida	Castil.	15	6178	151	97.56%	865	86.00%	11.56%	519	91.60%	5.96%
Brussels, KBR, 2485	Textualis	Dutch	14	10394	1047	89.93%	1475	85.81%	4.12%	1143	89.00%	0.92%
Brussels, KBR, 1805-1808	Textualis	Dutch	14	32421	6677	79.41%	8774	72.94%	6.47%	6530	79.86%	-0.45%
Vienna, ÖNB, 13.708	Textualis	Dutch	15	28421	5643	80.14%	7591	73.29%	6.85%	6087	78.58%	1.56%
Cambridge, CCC, ms. 041	Caroline	English	11	17071	5279	69.08%	7051	58.70%	10.38%	5212	69.47%	-0.39%
- *Without new chars**				16399	4607	71.91%	6379	61.10%	10.81%	4540	72.32%	-0.41%

Table 1: Out-of-domain tests for various manuscripts in languages represented in the CATMuS initiative. As CATMuS 1.0.0 has no Old English manuscript nor Middle Dutch. It is expected that Cambridge, CCC, ms. 041 might have a drop in performances because of its specific characters: we provide a second scoring method ignoring thorn, eth and other Middle English specific characters. For Middle Dutch, confusion of u/n was a high factor in performance drop.

Model		CATMuS	C. Cortado	C. Generic
Accuracy	Micro	86.40%	77.30%	85.28%
	Macro	88.59%	77.34%	86.30%
	STD Mac	7.73%	10.36%	7.81%
Differences	Average		11.26%	2.20%
	STD		7.17%	3.70%

Table 2: Scores across manuscripts depending on the model. Differences are computed against CATMuS. Micro average takes into account each manuscript character and error count, while Macro-average and STD take into account each manuscript accuracy.



(a) Absolute accuracy

(b) Improvement over CREMMA Generic

Figure 3: Test micro-accuracy per language and scripta. Bars represents the total number of characters per categorical feature.

Future

Consistent transcription standards, collaboration and data sharing across languages have provided HTR for medieval manuscripts with a boost in terms of accuracy. Such boost results in the provision of general models for HTR that play a crucial role in setting common benchmarks and improving the quality of predictions (Camps, Baumard, et al. 2023). Thanks to the guidance provided by the guidelines and the support of our sponsors, we have seen a significant increase in the number of available, interoperable datasets for training HTR models in recent years. This increase has enabled us to develop new generic models adapted to medieval manuscripts.

Although this article focuses mainly on Romance languages, we plan to extend the project to Old English and Middle Dutch, demonstrating our willingness to further

broaden the linguistic eras treated. In addition, the project includes a section dedicated to the modern period, enabling the project not only to cover numerous linguistic eras, but also to extend chronologically from the Middle Ages to the contemporary era.

References

Aruta, Davide et al. (Apr. 2023). *Liber*. URL: https://github.com/CIHAM-HTR/Liber/data.

Breuel, Thomas M. (2014). *Ocropy: Python-based tools for document analysis and OCR*. URL: https://github.com/tmbdev/ocropy..

Breuil, Eddie (Nov. 2019). *Méthodes et pratiques de l'édition critique des textes et documents modernes*. fre. Bibliothèque de littérature du xxe siècle 27. ISBN: 978-2-406-08639-0. DOI: 10.15122/isbn.978-2-406-08639-0. URL: https://classiques-garnier.com/methodes-et-pratiques-de-l-edition-critique-des-textes-et-documents-modernes.html (visited on 09/10/2021).

Camps, Jean-Baptiste (Feb. 6, 2017). *Homemade manuscript OCR (1): OCRopy*. Sacré Gr@@l. URL: https://graal.hypotheses.org/786 (visited on 12/06/2023).

Camps, Jean-Baptiste, Nicolas Baumard, et al. (Dec. 2023). "Make Love or War? Monitoring the Thematic Evolution of Medieval French Narratives". In: *Computational Humanities Research (CHR 2023)*. CEUR Workshop Proceedings. URL: https://enc.hal.science/hal-04250657.

Camps, Jean-Baptiste, Thibault Clérice, and Ariane Pinche (Oct. 2021). "Noisy medieval data, from digitized manuscript to stylometric analysis: Evaluating Paul Meyer's hagiographic hypothesis". In: *Digital Scholarship in the Humanities* 36.Supplement_2, pp. ii49–ii71. ISSN: 2055-7671. URL: https://doi.org/10.1093/llc/fqab033 (visited on 12/03/2021).

Clérice, Thibault, Alix Chagué, et al. (n.d.). *HTRomance*. URL: https://htromance-project.github.io/.

Clérice, Thibault and Ariane Pinche (Sept. 2021). *Choco-Mufin, a tool for controlling characters used in OCR and HTR projects*. DOI: 10.5281/zenodo.5356154. URL: https://github.com/PonteIneptique/choco-mufin (visited on 10/29/2021).

Clérice, Thibault, Ariane Pinche, and Malamatenia Vlachou-Efstathiou (Oct. 2022a). *Generic CREMMA Model for Medieval Manuscripts (Latin and Old French), 8-15th century*. Version 1.0.0. DOI: 10.5281/zenodo.7234166. URL: https://doi.org/10.5281/zenodo.7234166.

— (Oct. 2022b). *Generic CREMMA Model for Medieval Manuscripts (Latin and Old French), 8-15th century*. Version 1.0.0. DOI: 10.5281/zenodo.7234166. URL: https://doi.org/10.5281/zenodo.7234166.

Fradejas, José Manuel (Nov. 2023). *Examination of the output of the previous ms shows one of the problems with this model...* URL: https://twitter.com/JMFradeRue/status/1730191566508060883 (visited on 11/23/2023).

Gervers, Michael, Agnes Boutreux, and Avery Manton (Dec. 2023). *DEEDS HTR Dataset*.

Gille Levenson, Matthias (2023a). "Le *Regimiento de Los Prínçipes* et Sa Glose : Étude et Édition Numérique de La Partie Sur Le Gouvernement de La Cité En Temps de Guerre (III, 3)." PhD thesis. École Normale Supérieure de Lyon.

— (2023b). "Towards a general open dataset and models for late medieval Castilian text recognition (HTR/OCR)". In: *Journal of Data Mining and Digital Humanities*: *Special Issue: Historical documents and automatic text recognition*. Ed. by Ariane Pinche and Peter Stokes. DOI: 10.46298/jdmdh.10416.

Haverals, Wouter and Mike Kestemont (2023). "The Middle Dutch Manuscripts Surviving from the Carthusian Monastery of Herne (14th century): Constructing an Open Dataset of Digital Transcriptions". In: *Proceedings http://ceur-ws. org ISSN* 1613, p. 0073.

Kahle, Philip et al. (Nov. 2017). "Transkribus - A Service Platform for Transcription, Recognition and Retrieval of Historical Documents". In: *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*. Vol. 04. ISSN: 2379-2140, pp. 19–24. DOI: 10.1109/ICDAR.2017.307.

Kiessling, Benjamin (Apr. 2022). *The Kraken OCR system*. Version 4.1.2. URL: https://kraken.re.

Pinche, Ariane (2021). "Edition nativement numérique des oeuvres hagiographiques "Li Seint Confessor" de Wauchier de Denain d'après le manuscrit 412 de la Bibliothèque Nationale de France." PhD thesis. Lyon 3. URL: http://www.theses.fr/s150996 (visited on 05/09/2017).

— (July 2022a). *CREMMALab Project: Handwritten Text Recognition for medieval manuscripts*. Digital Humanities. Poster. URL: https://hal.science/hal-03724041.

— (June 2022b). "Guide de transcription pour les manuscrits du Xe au XVe siècle". URL: https://hal.archives-ouvertes.fr/hal-03697382.

Pinche, Ariane, Thibault Clérice, et al. (Nov. 2023). *CATMuS Medieval*. Version 1.0.0. DOI: 10.5281/zenodo.10066219. URL: https://doi.org/10.5281/zenodo.10066219.

Pinche, Ariane and Corinne Pierreville (Apr. 2023). *Fabliaux*. URL: https://github.com/CIHAM-HTR/Fabliaux/data.

Springmann, Uwe et al. (Sept. 2018). *Ground Truth for training OCR engines on historical documents in German Fraktur and Early Modern Latin*. arXiv:1809.05501 [cs]. DOI: 10.48550/arXiv.1809.05501. URL: http://arxiv.org/abs/1809.05501 (visited on 11/29/2023).

Stokes, Peter A. et al. (2021). *The eScriptorium VRE for Manuscript Cultures – Classics@ Journal*. en-US. URL: https://classics-at.chs.harvard.edu/classics18-stokes-kiessling-stokl-ben-ezra-tissot-gargem/ (visited on 12/06/2023).

Stutzmann, Dominique (2011). “Paléographie statistique pour décrire, identifier, dater... Normaliser pour coopérer et aller plus loin ?” fr. In: Publisher: BoD, p. 247. URL: https://halshs.archives-ouvertes.fr/halshs-00596970 (visited on 11/08/2021).

Stutzmann, Dominique et al. (2018). “Handwritten Text Recognition, Keyword Indexing, and Plain Text Search in Medieval Manuscripts”. en. In: URL: https://dh-abstracts.library.cmu.edu/works/6324 (visited on 03/08/2021).

Vlachou-Efstathiou, Malamatenia (2023). “Éditer les manuscrits grammaticaux glosés : solutions numériques face aux défis paléographiques : Le cas du Voss.Lat. O.41 d’Eutychès”. MA thesis. PSL - École nationale des chartes.

Appendix

Scripta	Century	Language	Type	Genre	Color	Print	Chars (K)
Caroline	10	Latin	prose	Poetry	False	False	9.5
				Speeches	True	False	0.9
				Treatises	True	False	1.6
			verse	Poetry	True	False	0.8
	11	Latin	prose	Narratives	True	False	3.7
	12	Latin	prose	Narratives	True	False	2.3
				Treatises	True	False	1.1
	8	Latin	prose	Treatises	True	False	1.6
	9	Latin	prose	Narratives	True	False	2.4
				Treatises	True	False	642.1
Cursiva	14	Castilian	prose	Treatises	True	False	4.2
		Catalan	prose	Treatises	False	False	11.8
		French	prose	Narratives	True	False	12.1
		Italian	prose	Narratives	True	False	8.6
			verse	Narratives	True	False	11.8
		Latin	prose	Treatises	True	False	35.6
	15	Castilian	prose	Treatises	True	False	10.1
		Catalan	prose	Treatises	False	False	12.0
		French	prose	Narratives	False	False	71.0
					True	False	5.4
				Treatises	True	False	76.0
			verse	Narratives	True	False	85.6
				Poetry	True	False	7.2

		Latin	prose	Documents of practice	True	False	57.0
	16	Latin	prose	Documents of practice	True	False	8.6
Goth. Doc.	12	Latin	prose	Documents of practice	True	False	12.8
	13	Latin	prose	Documents of practice	False	False	12.0
Humanistica	14	Italian	prose	Narratives	True	False	15.6
		Venitian	prose	Documents of practice	True	False	1.6
	15	Italian	verse	Narratives	True	False	17.3
		Latin	prose	Treatises	True	False	7.0
			verse	Poetry	False	False	7.0
					True	False	3.6
	16	Castilian	prose	Treatises	False	False	7.6
		Italian	prose	Narratives	True	False	4.7
			verse	Poetry	True	False	4.1
		Latin	prose	Narratives	True	False	7.8
Hybrida	14	Castilian	prose	Narratives	True	False	16.8
				Treatises	False	False	8.2
	15	Castilian	prose	Documents of practice	True	False	1.3
				Espistolary	True	False	8.5
				Narratives	True	False	12.8
				Treatises	True	False	151.4
			verse	Poetry	True	False	1.1
		French	prose	Narratives	False	False	11.5
		Italian	verse	Drama	True	False	5.1
		Latin	prose	Narratives	True	False	40.7
				Treatises	False	False	10.5
			verse	Poetry	True	False	0.8
	16	Catalan	verse	Poetry	True	False	5.8
Personal	16	Latin	prose	Treatises	True	False	7.2
Praegothica	12	Latin	prose	Documents of practice	True	False	10.8
				Treatises	True	False	27.7
Print	15	Castilian	prose	Treatises	True	True	174.8
		French	mixte	Treatises	True	True	24.8
			prose	Narratives	False	True	53.2
					True	True	59.3
				Treatises	True	True	19.3
			verse	Narratives	True	True	49.2
				Poetry	True	True	20.5
		Latin	prose	Narratives	True	True	56.2
	16	French	prose	Narratives	True	True	68.7
				Treatises	True	True	39.1

			verse	Narratives	True	True	54.8
Semihybrida	13	Castilian	prose	Documents of practice	False	False	11.1
	15	Castilian	prose	Narratives	False	False	7.1
					True	False	12.9
		Latin	prose	Narratives	True	False	23.3
	16	Catalan	prose	Documents of practice	False	False	6.6
Semitextualis	14	Castilian	prose	Treatises	True	False	279.1
	15	Castilian	mixed	Poetry	True	False	12.1
Textualis	13	Castilian	prose	Treatises	True	False	12.9
		French	prose	Narratives	False	False	7.8
					True	False	17.8
				Treatises	True	False	71.6
			verse	Narratives	True	False	326.2
				Poetry	True	False	28.3
		Latin	prose	Treatises	False	False	17.0
					True	False	69.9
			verse	Poetry	True	False	5.5
	14	Castilian	prose	Treatises	True	False	7.4
		Catalan	prose	Narratives	True	False	12.9
		French	prose	Narratives	False	False	4.4
					True	False	90.6
				Treatises	True	False	24.6
			verse	Narratives	False	False	47.0
					True	False	18.3
		Italian	prose	Narratives	True	False	9.3
		Latin	prose	Espistolary	True	False	12.4
				Narratives	True	False	14.7
				Treatises	True	False	52.9
			verse	Drama	False	False	16.2
				Poetry	True	False	5.6
		Navarrese	prose	Treatises	False	False	16.8
		Venitian	prose	Narratives	True	False	6.4
	15	Castilian	prose	Treatises	True	False	191.2
		French	prose	Narratives	True	False	76.7
			verse	Narratives	True	False	15.2
		Latin	prose	Speeches	True	False	4.8
				Treatises	True	False	17.4