On Burgundian (di)vine orators and other impostors: Stylometry of Late Medieval Rhetoricians

Camps, Jean-Baptiste

jean-baptiste.camps@chartes.psl.eu
École nationale des chartes | Université PSL, France

Salvati, Benedetta

benedetta.salvati@chartes.psl.eu
École nationale des chartes | Université PSL, France; Université de Lausanne, Suisse

Introduction

The attribution of medieval texts poses specific challenges. Many texts are anonymous or of dubious authorship. Very often, there is no known candidate for the authorship of a given text. In many other cases, there is only one single known candidate (or only one that can actually be tested), and the task becomes one of authorship verification. Widespread in the Middle Ages, anonymity is the hallmark of heraldic literature. Only rarely indeed do heralds claim ownership over their works, and, when they do, they generally use a pseudonym (Couhault 2020), which makes it difficult to identify the authors.

This paper intends to answer the question of whether or not the two pieces of topical poetry, *Le blazon des armes de Lalaing* (1509), and *L'epitaphe de Jehan de Luxembourg* (1508-1509 ca.), attested in the MS Paris, BnF, fr. 5229, at the folios 143r-144v and 150r-v, undoubtedly two works of an herald of arms, can be ascribed to Nicaise Ladam, *chroniqueur-poète* and herald of arms of the Burgundian-Habsburgs' House. The first poem is signed by *Luxembourg le herault*, while the running title of the second one attributes the epitaph to the *Songeur de Bapasmes*, both known to be the pen names of Nicaise (Ladam 1975; Lemaire and Tourneux 2001; Couhault 2020), who was born in Bapaume in 1465. In order to compare the results on those two pieces, we include a third anonymous one, *Or a permis la divine puissance* (1524-1546 ca.), for which there are no suspicions of Ladam's authorship.

Preparing the corpus

Writing among heralds is very poorly recorded. Not many French-speaking officers apart from Ladam seem indeed to have practised rhymed eloquence, which has been a major constraint in establishing our corpus. In the absence of documentation, we therefore gave priority to Burgundian contemporary authors, such as J. Molinet, *indiciaire* and official historiographer of the dukes of Burgundy (Devaux 1996; Frieden 2013), whose dialect and poetic institutes could correspond to the ones of the Anonymous of MS 5229, as well as French rhetoricians, namely G. Crétin, who has practised different forms of topical writing, including funeral lamentation and eulogy, genres to which the two poems mentioned above belong.

The texts included in the corpus have been acquired using handwritten text recognition (cer of 2.26% on test set) and ocr, and have then been carefully corrected and critically edited. As for the Ladam's ones, they have been edited by B. Salvati, as a part of her digital scholarly edition of *Chronique abrégée* (1488-1546) (Salvati 2020); the other ones have been taken from the critical reference editions, which we later emended where necessary (see Table 3 in appendix). We removed inserts in prose or in Latin from the texts.

Method

Due to manuscript based transmission, authorial idiolect in the documents is mixed with the linguistic particularities of the successive scribes. It is particularly true of spelling, probably the less stable element in the tradition. For this reason, we experiment not only with traditional stylometric features such as affixes (Sapkota et al. 2015), but we also use lemmatised function words and part-of-speech trigrams (Camps, Clérice, and Pinche 2021).

Dealing with short texts (as is often the case with occasional poetry) of variable lengths is also an issue to which there are no convenient solutions. To have a minimum sample size of 1000 words, we aggregate all poems of the same author, and then take samples of 1000 words each.

We use two different type of supervised analyses:

- Linear support-vector machines, trained to recognise the style of each author. To handle the differences in training corpus size, we use downsampling and class weights (Lemaître, Nogueira, and Aridas 2017);
- a generalised-impostors method (Koppel and Winter 2014; Kestemont et al. 2016), following the implementation provided by M. Eder (Eder, Rybicki, and Kestemont 2016).

Table 1. F1 Score for leave-one-out evaluation of the SVM models for each feature type

| Author	affixes	lemma (full)	lemma (fw)	POS	Support
Cretin	1.00	1.00	0.83	0.94	16
Ladam	0.99	0.99	0.94	0.96	37
Molinet	0.96	0.96	0.83	0.96	12
macro av.	0.98	0.98	0.87	0.95	
weighted av.	0.98	0.98	0.90	0.95	

All in all, affixes and lemma, closely followed by POS, seem to be the most efficient features, when used in isolation. If only the score is considered, they seem like the natural choice, but, when we look at the features with the highest coefficients in the classification models, the results are somewhat different (fig. 1). Some of them, such as the word ending **y** after **u**, seem indeed more likely to indicate copyists' scribal habits, rather than authorial style features.

Such bias also contaminates the POS: for instance, the rarer spelling for the definite article *ung*, is erroneously interpreted by the tagger as demonstrative, causing biased pseudo-features. In addition, lemmas themselves seem to retain a lot of thematic and generic features. For instance, the sting (*dart*), frequently associated to the representations of Death in funeral poetry, to which the major part of Ladam's chosen pieces belong, or the overuse of the second-person pronoun (*tu*).

For these reasons, selected function words, though less performing, seem like the less biased feature (the other, harder to imple-

ment, alternative would be to correct annotation and perform spelling normalisation). We decide to continue on using only function words.

Globally, the imbalance of the training corpus is a substantial obstacle, and the problem is to some extent an open set question, which makes the impostors approach an interesting alternative. The impostors method was first optimised through the procedure implemented by (Eder, Rybicki, and Kestemont 2016). In the leave-one-out training, the range of unreliable values were contained between 0.28 and 0.55.
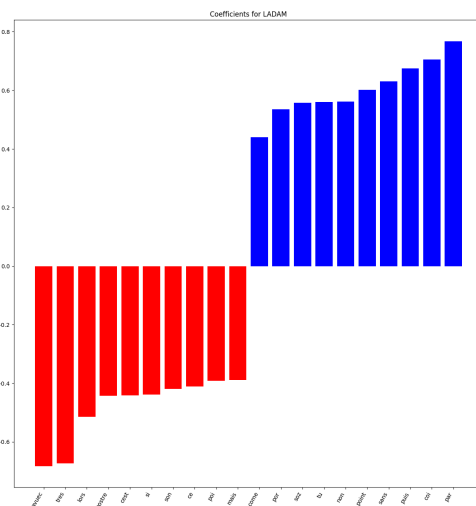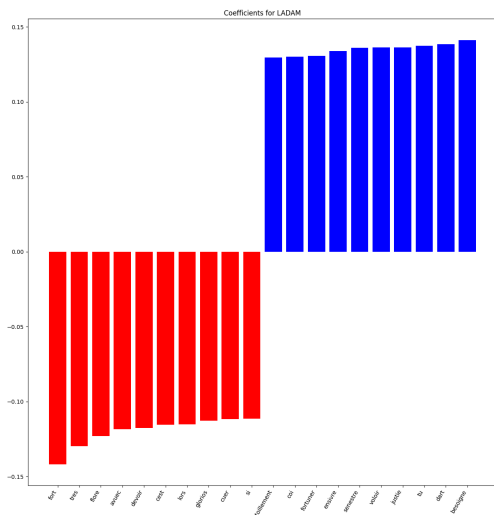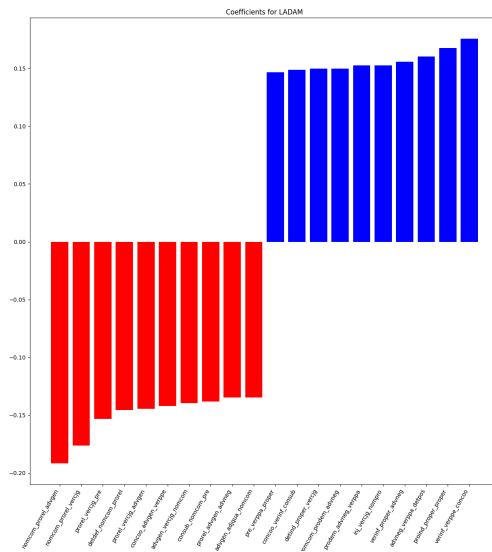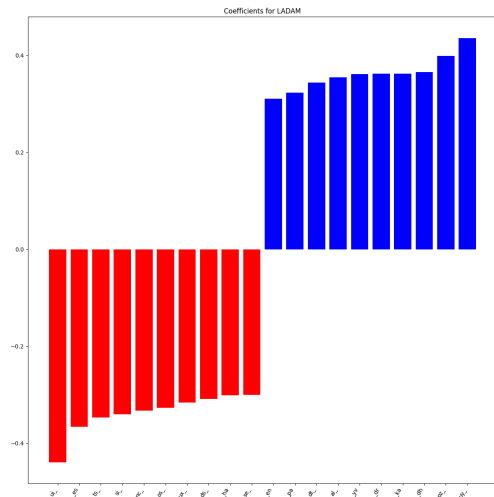
Figure 1: Coefficients for the Ladam models based on affixes, lemmas, POS 3-grams as well as lemmatised function words

Results and discussion

Table 2: Coefficients of the pieces to be attributed based on the SVM models (value above 0 in bold), and impostors method (value above P2 are in bold; dubious values between P1 and P2 in italics))

	Cretin	Ladam	Molinet	SVM pred	Cretin	Ladam	Molinet	Impost. pred	
Or a…	-0.61	0.04	-0.48	Ladam?	0.55	0.65	0.24	Contra-dictory	
Blazon	-0.39	-0.38	-0.29	None	0.40	0.83	0.32	Ladam?	
Epitaphe	-0.73	0.43	-0.68	Ladam	0.39	0.79	0.06	Ladam?	

If Ladam is the most likely candidate inside the set for the *Blazon* and *Epitaphe*, the results are a bit more conclusive regarding the second. Could he be the author of only this piece? Nonetheless, the results are to be taken with caution, as the scores are partially contradictory between themselves. Current results make it likely that *Or a...* is indeed not of Ladam, and maybe even the *Blazon*, whose style seem somewhere between the three authors but inside the style of neither of them.

Future research

To avoid scribal biases, even when a normalisation such as lemmatisation is performed, it will be necessary to correct annotation errors. Our results tend indeed to show that these errors are not random and uniform, but are subject to biases when, for instance, a rare scribal spelling is encountered by the lemmatiser, creating issues in attribution. In itself, automated normalisation might not be sufficient to handle this.

Secondly, given the partially contradictory results of the two methods, it will be necessary to improve their results by enriching the dataset. For the impostors' method, the inclusion of more imposters is necessary to reach a really 'open set' approach. For the SVM, the addition of more samples, or, of additional features seem necessary. That's why, in the future, we plan to also integrate features of the poetic style, who sometimes have been shown to be the best performing feature for the analysis of poetry (Plecháč, Bobenhausen, and Hammerich 2018). In particular, the alternation of verses of varying length, the schema of verse-ending rhymes, the presence of internal rhymes and the place of caesura are features worthy of encoding and analysing, especially since some patterns are already suspected to be strong authorial markers (for instance, the presence of internal rhymes is suspected to be a feature of Ladam's style).

Data and materials availability

Supervised analyses were made with SuperStyl (https://github-b.com/SupervisedStylometry/SuperStyl/) and the impostors' implementation is based on *Stylo* R package.

Appendix: Composition of the corpus

Table 3: List of the texts of the corpus

Author	Genre	Date	Title	Nb. words	Source
Crétin	lament	ca. 1496	Deploration sur le trepas de feu Okegan	2847	(Crétin 1864)
Crétin	lament	3901	Plainte sur le trespas de messire Guillaume de Byssipat	3901	(Cretin 1932)
Crétin	lament	1335	Plainte sur le trespas de Jehan Braconnier	1335	(Cretin 1932)
Crétin	lament	7763	L'apparition du Mareschal sans reproche	7763	(Cretin 1932)
Crétin	occasional poem	1498-1501	La Rescription des femmes de Paris	153	
Crétin	epistle	1515	Ad illustrissimum seigneur monseigneur Franciscum valescorum	680	
Crétin	rhymed chronicle	1515-25	Chronique françoyse	443	
Ladam	lament	1492	Le Doeul d'Arras	509	(Salvati 2020)
Ladam	lament	1492	Rescription aux Allemantz	319	(Salvati 2020)
Ladam	lament	1494	L'Epytaphe de Philippes de Criefvecoeur	2073	(Salvati 2020)
Ladam	prayer poem	1495	De Job	757	(Salvati 2020)
Ladam	voyage	1498	Recreation de la paix faicte à Paris	858	(Salvati 2020)
Ladam	occasional poem	1498	Où est Jubal et Tubal, les chanteurs ?	278	(Salvati 2020)
Ladam	voyage	1499	Rescription à Hesdin, Aire et Bethune	498	(Salvati 2020)
Ladam	royal entry	1503	Le retour de nostre seigneur	551	(Salvati 2020)
Ladam	lament	1504	Epitaphe dudict feu bastard de Bourgongne	2251	(Salvati 2020)
Ladam	lament	1504	Epitaphe de Philibert duc de Savoie	331	(Salvati 2020)
Ladam	voyage	1504	Le voiaige de Philippes prinche des Hespaignes	422	(Salvati 2020)
Ladam	dit	1505	Le vergyer d'Austrice	948	(Salvati 2020)
Ladam	occasional poem	1506	Description de la couronne du roy Philippes	803	(Salvati 2020)
Ladam	lament	1506	Deploration sur la mort dudit roy Philippe de Castille	5122	(Salvati 2020)
Ladam	lament	1506	Epitaphe de maistre Alexandre Agricola	250	(Salvati 2020)
Ladam	lament	474	Epitaphe de Philippes de Bouzenton	2688	(Salvati 2020)
Ladam	occasional poem	1506	Le Confort du Pasteureau d'Austrice	2688	(Salvati 2020)
Ladam	occasional poem	1507	Alliance faicte à Calaix pour Charles d'Austrice	425	(Salvati 2020)
Ladam	occasional poem	1508	L'Attente de paix refformee	842	(Salvati 2020)
Ladam	occasional poem	1508	Sacre legat du siege apostolicque	230	(Salvati 2020)
Ladam	prayer poem	1508	Conditor alme siderum	943	(Salvati 2020)
Ladam	occasional poem	1508	L'an mil cincq cens et huict pour vray	242	(Salvati 2020)
Ladam	dit	1508	Constumes seignouralles de Bappalmes	1415	(Salvati 2020)
Ladam	occasional poem	1513	Recoeul de France et d'Angleterre	1105	(Salvati 2020)
Ladam	royal entry	1514	L'entree et couronnement du roy Franchois premier	655	(Salvati 2020)
Ladam	royal entry	1514	L'entree de Charles roy de Castille en Anvers	966	(Salvati 2020)
Ladam	lament	1516	L'Epitaphe de Ferdinand d'Aragon	5333	(Salvati 2020)
Ladam	dit	1517	Chasteaulx en Espaigne	2504	(Salvati 2020)
Ladam	lament	1518	Epytaphe de Maximilien d'Austrice	661	(Salvati 2020)
Ladam	voyage	1535	Le voyage de La Goulette	1994	(Salvati 2020)
Molinet	lament	1467	Le Trosne d'Honneur	1538	(Dupire 1939)
Molinet	occasional poem	1475	Le Temple de Mars	1691	
Molinet	lament	1482	La complainte pour le trespas de Marie de Bourgogne	3047	(Becker 1902)
Molinet	lament	1483	Epitaphe de Josse de Lalaing	859	(Thiry 1973)
Molinet	lament	1486	L'arbre de Bourgogne	1204	(Dupire 1939)
Molinet	lament	1489	Epitaphe de Simon Marmyon	340	(Dupire 1939)
Molinet	lament	1493	La mort Federicq empereur	1290	(Dupire 1939)
Molinet	lament	a. 1500	Le miroir de vie	2013	(Dupire 1939)
Molinet	lament	1500	Lamentables regrés pour le trespas d'Albert duc de Zassen	736	(Dupire 1939)
Molinet	lament	1502	Epitaphe de monseigneur Henry de Berghes	428	(Dupire 1939)
Molinet	lament	1504	Au comte Englebert de Nassau	896	(Dupire 1939)
Molinet	lament	1506	L'Epitaphe du duc Philippe de Bourgogne	322	(Dupire 1939)
O. St-Gelais	epistle	1496	Espitre de Dido à Enée	2929	
O. St-Gelais	lament	1498	Epitaphe de Charles VIII	465	
O. St-Gelais	lament	1498	Complainte de Charles VIII	3879	
O. St-Gelais	Allegoric poem	1492	Le Séjour d'honneur	888	
O. St-Gelais	translation/epics	1498	*Énéide* - beg.	10028	

Bibliography

Becker, Philipp August. 1902. "Autobiographisches von Jehan Molinet." *Zeitschrift Für Romanische Philologie* 26 (6): 641–51. https://doi.org/doi:10.1515/zrph.1902.26.6.641.

Camps, Jean-Baptiste, Thibault Clérice, and Ariane Pinche. 2021. "Noisy Medieval Data, from Digitized Manuscript to Stylometric Analysis: Evaluating Paul Meyer's Hagiographic Hypothesis." *Digital Scholarship in the Humanities* 36 (Supplement_2): ii49–ii71. https://doi.org/10.1093/llc/fqab033.

Couhault, Pierre. 2020. *L'Étoffe Des Hérauts. L'office d'armes Dans L'Europe Des Habsbourg à La Renaissance*. Classiques Garnier.

Cretin, Guillaume. 1932. *Œuvres Poétiques, éd*. Edited by Kathleen Chesney. Paris.

Crétin, Guillaume. 1864. *Déploration de Guillaume Crétin Sur Le Trépas de Jean Okeghem: Musicien, Premier Chapelain Du Roi de France et Trésorier de Saint-Martin de Tours, Remise Au Jour*. A. Claudin.

Devaux, Jean. 1996. *Jean Molinet, Indiciaire Bourguignon*. Champion.

Dupire, Noël. 1939. *Les Faictz et Dictz de Jean Molinet*. Société des anciens textes français.

Eder, Maciej, Jan Rybicki, and Mike Kestemont. 2016. "Stylometry with R: A Package for Computational Text Analysis." *R Journal* 8 (1): 107–21. https://journal.r-project.org/archive/2016/RJ-2016-007/index.html.

Frieden, Philippe. 2013. *La Lettre et Le Miroir: écrire L'histoire d'actualité Selon Jean Molinet*. Honoré Champion.

Kestemont, Mike, Justin Stover, Moshe Koppel, Folgert Karsdorp, and Walter Daelemans. 2016. "Authenticating the Writings of Julius Caesar." *Expert Systems with Applications* 63: 86–96.

Koppel, Moshe, and Yaron Winter. 2014. "Determining If Two Documents Are Written by the Same Author." *Journal of the Association for Information Science and Technology* 65 (1): 178–87.

Ladam, Nicaise. 1975. *Mémoire et épitaphe de Ferdinand d'Aragon: édition Critique Par Claude Thiry*. Les Belles Lettres.

Lemaire, Jacques Charles, and André Tourneux. 2001. *Le Mémoire de L'aigle et de La Salamandre de Nicaise Ladam, Roi d'armes de Charles Quint*. Peeters.

Lemaître, Guillaume, Fernando Nogueira, and Christos K Aridas. 2017. "Imbalanced-Learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning." *The Journal of Machine Learning Research* 18 (1): 559–63.

Plecháč, Petr, Klemens Bobenhausen, and Benjamin Hammerich. 2018. "Versification and Authorship Attribution. A Pilot Study on Czech, German, Spanish, and English Poetry." *Studia Metrica et Poetica* 5 (2): 29–54.

Salvati, Benedetta. 2020. "Édition Critique Numérique de La Chronique Abrégée: Œuvres et Recueil (1488-1546) de Nicaise Ladam." Thèse en préparation, Université Paris Sciences & Lettres.

Sapkota, Upendra, Steven Bethard, Manuel Montes, and Thamar Solorio. 2015. "Not All Character N-Grams Are Created Equal: A Study in Authorship Attribution." In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 93–102.

Thiry, Claude. 1973. "Un inédit de Jean Molinet: l'épitaphe de Josse de Lalaing, sire de Montigny (1483)." *Bulletin de La Commission Royale d'Histoire* 139 (1): 29–66.