



**HAL**  
open science

## Assessing collaborative writing

Sophie Dolto, Simon Gabay

► **To cite this version:**

Sophie Dolto, Simon Gabay. Assessing collaborative writing. DH Benelux 2024, Jun 2024, Leuven, Belgium. hal-04569687

**HAL Id: hal-04569687**

**<https://hal.science/hal-04569687v1>**

Submitted on 6 May 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Assessing collaborative writing. A stylometric study of Manchette's and Bastid's co-written novel

Sophie Dolto and Simon Gabay

Université de Genève

sophie.dolto@etu.unige.ch, simon.gabay@unige.ch

## Abstract

In the second half of the 20th, century, a small circle of French noir writers began publishing a large number of novels, some of which were co-written. This was notably the case of *Laissez bronzer les cadavres*, co-signed in 1971 by the foremost French crime fiction author of his period, Jean-Patrick Manchette, and another figure of Gallimard's famous "Série Noire" collection and of the "neo-polar" world in general, Jean-Pierre Bastid. Using stylometry, we attempt to analyze this joint work to unveil the writing process and the dynamic of creation between the two writers. Our results show that if the text was indeed co-written by the two authors, Manchette very probably played a leading role in finalizing the novel.

## 1 Introduction

At the beginning of the 1970s, in the wake of May '68, the French noir novel underwent a major revival with the emergence of what Jean-Patrick Manchette (1942-1995) would call the "néo-polar" (Dolto, 2023, p. 76-77), with authors such as Manchette of course (its most famous representative), but also Pierre Siniac (1928-2002), Jean Vautrin (1933-2015), Jean-Pierre Bastid (1937-)...

It is commonly acknowledged that, beyond their predilection for the *roman noir*, these authors also shared strong left-wing political convictions (Collovald and Neveu 2001 ; Gorrara 2003, p. 57-60), as well as a notable interest in cinema. These common traits lead "neo-polar" writers to collaborate regularly on novels and even screenplays. Within these circles of politicized authors, writing was primarily regarded as a means of earning a livelihood, and writing opportunities often circulated among friends.

Bastid and Manchette were part of this network of "intellectual proletarians" (Manchette, 2015, p. 311), and shared the practices of this circle. The former co-wrote numerous novels, in particular

with Michel Martens (1940-) (Bastid and Martens, 1977), while Manchette co-wrote television scripts with Michel Levine (1936-) (Levine and Manchette, 1968). Although both participated in this collective writing dynamic, the two authors differed in the extent to which they engaged. Bastid cosigned numerous books, in contrast to Manchette, who predominantly wrote his novels alone. Among the rare cases of books co-written by the latter, one was cosigned with the former and holds a special place : *Laissez bronzer les cadavres* (1971), the novel that marked the beginning of Manchette's career as a writer, and secured him a place in the prestigious "Série Noire" collection.

## 2 Authorship attribution

Although *Laissez bronzer les cadavres* is included in Manchette's complete noir novels published in the "Quarto" collection (Manchette, 2005), it was excluded from a significant translation project into English, revealing the ambiguous status of the text. Should we exclude this text because of the hybrid nature of the writing, or can we consider that the English-language "Manchettian" corpus will only be complete once this text has been translated?

Witnesses from the period, including those from family circles or insiders still alive, have reported in interviews that the novel was crafted during a two-week residency. Manchette and Bastid alternated in writing chapters, transposed a Western film outline into 1970s France. While the final proofreading's exact extent remains unknown, it likely rested with Manchette, potentially giving the text a more "Manchettian" than "Bastidian" flavor.

## 3 Co-writing

In addition to the question of author attribution, another emerges : that of the study of the collaboration between the two authors. Four-handed

Author	Book Title	Publisher	Date	Collection	Length
Jean-Pierre Bastid et Jean-Patrick Manchette	<i>Laissez Bronzer les cadavres</i>	Gallimard	1971	Série noire	145 pages
Jean-Pierre Bastid	<i>Méchoui massacre</i>	Gallimard	1974	Super noire	163 pages
Jean-Pierre Bastid	<i>Parcours fléché</i>	Gallimard	1994	Série noire	267 pages
Jean-Patrick Manchette	<i>L’Affaire N’Gustro</i>	Gallimard	1971	Série noire	127 pages
Jean-Patrick Manchette	<i>Ô dingos, ô châteaux!</i>	Gallimard	1972	Série noire	129 pages
Jean-Patrick Manchette	<i>Le Petit bleu de la côte ouest</i>	Gallimard	1976	Série noire	129 pages
Jean Vautrin	<i>À Bulletins rouges</i>	Gallimard	1973	Série noire	129 pages
Jean Vautrin	<i>Billy-ze-Kick</i>	Gallimard	1974	Série noire	109 pages

TABLE 1 – Studied corpus with (i) the problematic text, (ii) the works of J.-P. Bastid, (iii) the works of J.-P. Manchette, (iv) the works of Jean Vautrin.

writing is a well-known phenomenon in literature (Lafon and Peeters, 2006) : Dumas and Maquet, the Goncourt brothers, Boileau and Narcejac... The examples are innumerable, yet never identical. Although the duo formed by Edmond and Jules de Goncourt may have been almost fused (Cabannes and Dufief, 2020), the same cannot be said for Alexandre Dumas’ various associates, most of whom have faded into oblivion (Fillaire, 2002).

Co-writing is thus a complex phenomenon, spanning from perfect symbiosis to the complete delegation of writing to a third party. In our study, the inquiry into who wrote which part – the traditional focus of stylometry – is intricately connected to the dynamics and methods adopted by the two authors. These questions not only help us establish the significance of the text within Manchette’s body of work but also shed light on the modalities of this collaborative effort executed within a confined time frame.

## 4 State of the art

Stylometry has established itself as the benchmark technique for author attribution (Cafiero and Camps, 2022), and numerous studies on French literature have demonstrated its effectiveness, particularly in the realm of classical theater (Schöch, 2013; Cafiero and Camps, 2019; Gabay, 2021). In recent years, research has shifted towards more contemporary texts, such as those by Willy and Collette (Cafiero and Puren, 2022), and has explored topics beyond authorship attribution, such as the examination of the French canon during the 19th and 20th centuries (Barré et al., 2023).

Among the many studies conducted, several

have focused on written texts deemed “co-written.” Examples include the *Roman de la Rose* (Eder, 2015b), initiated by Guillaume de Lorris and continued by Jean de Meun during the 13th century, and *Psychée*, a tragedy-ballet collaboration involving Molière, Pierre Corneille, and Philippe Quinault (Cafiero and Camps, 2021).

## 5 Corpus

The first task is to gather documents that are as comparable as possible : noir novels, written solely by one author, if possible in the 1970s (when *Laissez bronzer les cadavres* was written). The significant presence of co-authored works in Bastid’s bibliography creates the difficulty of locating sufficient works only authored by him, especially during the 1970s. We also need to find a third author, to diversify the data and not only have texts from Bastid and Manchette. Our choice fell on Jean Vautrin, who is part of the same literary circle as the other two authors. Since these authors write novels, we do not have a major problem regarding the amount of data for a stylometric study.

The works we have selected are presented in the tab. 1. The digital facsimiles have been segmented with the FoNDUE-Boscaiola model (Solfrini et al., 2024) and then OCRized with the CATMuS Print (large) model (Gabay et al., 2024) to extract the text from the PDFs.

## 6 Experiments

### 6.1 Set up

For this study, we use SUPERSTYL (Camps, 2021), which relies on Support Vector Machines

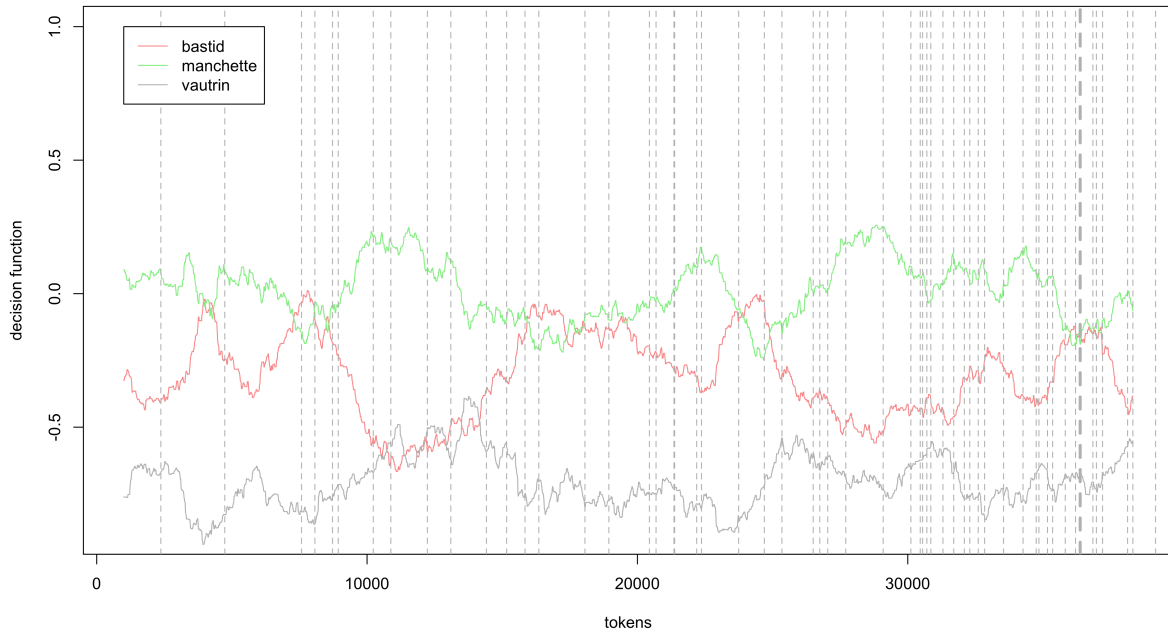


FIGURE 1 – Rolling analysis of *Laissez bronzer les cadavres*. Each dashed-grey vertical line represents a section of the book. The thicker vertical line separates Friday from Saturday.

(SVM). Several configurations are tested to find the optimal one :

- we try both stop words (Kestemont, 2014) and characters 3-grams (Sapkota et al., 2015) to compare the efficiency of both methods ;
- several sizes of text samples are tested, between 1 000 and 6 000 – previous studies showing that samples from 2 000 to 5 000 words are the most satisfactory (Eder, 2015a) ;
- Various method of sampling are tested, especially downsampling and upsampling.

All data are normalized using an L2 normalization. We use class weights to penalize mistakes during the classification process, a linear support vector classifier and a group 10-fold validation.

## 6.2 Classification task

	precision	recall	f1-score	support
Bastid	0.98	1.00	0.99	54
Manchette	1.00	0.97	0.98	59
Vautrin	0.98	1.00	0.99	43
accuracy			0.99	156
macro avg	0.99	0.99	0.99	156
weighted avg	0.99	0.99	0.99	156

TABLE 2 – Classification results on the corpus.

If function words offer satisfactory results (93%),

the best results (99% of accuracy, cf. tab. 2) are obtained with a classic configuration : 2 000 words samples, 3-grams and downsampling.

## 6.3 Rolling stylometry

Using this configuration, we train an SVM classifier and we then apply to each subsequent length  $n$  segment at a step of 10 (resulting in an overlap of  $n - 10$  between two following segments, such as words 1-2 000, 10-2 010, 30-2 030, etc.). Next, we plot the decision function for each classifier after extracting the classification (cf. fig. 1). The different sections of the book are signaled by dotted gray vertical lines.

## 7 Discussion

### 7.1 Stylometric results

Based on these results, we are able to analyze the writing process. The graph (cf. fig. 1) unveils a sporadic presence of Bastid and the omnipresence of Manchette. The fact that the different sections of the book do not coincide with changes in authorship does not argue for a distribution of work *a priori*, made according to the plan of the book, but more for interventions *a posteriori*, not taking into account the structure. This confirms the idea that the novel was written by the two authors, and then

was (almost) completely rewritten by Manchette, whose style is therefore predominant.

## 7.2 Stylistic confirmation

The convergence of the stylometric markers produced by our experiment with the observations of literature specialists can help consolidate our results. Among the most distinctive 3-grams of Manchette found by the SVM, we find `_je`, which conversely has the least correlation with Bastid's writing. We find a symmetrically opposite situation with the trigram `_ui` positively linked to the writing of Bastid, and negatively to that of Manchette. These two features seem to contrast Manchette's oral style with Bastid's more written style.

The trigram `_ui` refers to the relative pronoun *qui* ("who"), which is typical of complex constructions that orality tends to avoid in favor of asyndetic structures (Berrendonner, 2011). In Bastid's texts, the use of the relative pronoun is also used to introduce a relative subordinate proposition which specifies the principal one. It often functions as a stage direction, which sees the narrator intervene in the middle of the direct speech, mixing the effects of orality with typical elements of description.

- Oh, pardon, dit Loulou *qui* éclate de rire grossièrement.
- Oh, sorry, says Loulou, who bursts out laughing rudely.

Such interventions of the narrator are almost nonexistent in Manchette's writing, which is, on the contrary, much more oral. The trigram `_je` refers to the first person pronoun *je* ("I"), typical of dialogue :

- *Je* descends vers le sud.
- I'm heading south.
- C'est parfait, ça. Moi, *je* vais sur Sully.
- That's perfect. Me, I'm going to Sully.

We note here the repetition of the personal pronoun (*moi, je*), typical of orality (Krötsch, 2007).

If Manchette and Bastid both use dialogue (respectively 1439 and 1614 dashes introducing direct speech in the two corpora), their treatment of orality differs, that of Manchette appearing as more vivid and less "polluted" by the interventions of the narrator. Such results converge with the assertion of Serge Quadrupani, according to whom "no author of contemporary noir novel can measure up to him in the field of dialogue" (Quadrupani, 2020).

## 8 Conclusion

Our study would greatly benefit from a greater quantity of data, particularly with regard to the stylistic part. Nevertheless, the stylometric analysis of *Laissez bronzer les cadavres* shows that if it was clearly written by two authors, it reveals a specific type of co-writing, heavily marked by the rewriting by one of the two authors : Manchette. It therefore seems clear that the text belongs to the Manchettian corpus, and that his works translated into English, from which this text is absent, should be supplemented with a new translation.

## References

- Jean Barré, Jean-Baptiste Camps, and Thierry Poibeau. 2023. [Operationalizing Canonicity](#). *Journal of Cultural Analytics*, 8(1) :88113.
- Jean-Pierre Bastid and Michel Martens. 1977. *Adieu la vie...* Gallimard.
- Alain Berrendonner. 2011. [Unités syntaxiques & unités prosodiques](#). *Langue française*, 170(2) :81–93.
- Jean-Louis Cabanès and Pierre Dufief. 2020. *Les Frères Goncourt*. Fayard, Paris.
- Florian Cafiero and Jean-Baptiste Camps. 2019. [Why Molière most likely did write his plays](#). *Science Advances*, 5(11) :1–14.
- Florian Cafiero and Jean-Baptiste Camps. 2021. [‘Psyché’ as a Rosetta Stone? Assessing Collaborative Authorship in the French 17th Century Theatre](#). In *Proceedings of the Conference on Computational Humanities Research 2021*, volume 2989, pages 377–391. CEUR Workshop Proceedings.
- Florian Cafiero and Jean-Baptiste Camps. 2022. [Affaires de style : du cas Molière à l'affaire Grégory, la stylométrie mène l'enquête](#). Le Robert, Paris.
- Florian Cafiero and Marie Puren. 2022. [Claudine à l'atelier: l'impact de Willy et ses secrétaires sur l'écriture de colette](#). In *Humanistica 2022*, Montréal, Canada.
- Jean-Baptiste Camps. 2021. [SUPERvised STYLometry \(SuperStyl\)](#).
- Annie Collovald and Érik Neveu. 2001. [Le « néopolar »: du gauchisme politique au gauchisme littéraire](#). *Sociétés & Représentations*, 11(1) :77–93.
- Sophie Dolto. 2023. [« Il faudrait cesser d'écrire des romans »: The Paradoxical Influence of the Situationist International on Jean-Patrick Manchette](#). *New Readings*, 19 :67–90.
- Maciej Eder. 2015a. [Does Size matter? Authorship Attribution, Small Samples, Big Problem](#). *Literary and Linguistic Computing*, 30(2) :167–182.
- Maciej Eder. 2015b. [Rolling stylometry](#). *Digital Scholarship in the Humanities*, 31(3) :457–469.
- Bernard Fillaire. 2002. *Alexandre Dumas et associés*. Bartillat, Paris.

- Simon Gabay. 2021. [Beyond Idiolectometry? On Racine's Stylometric Signature](#). In *Proceedings of the Conference on Computational Humanities Research 2021*, volume 2989, pages 359–376. CEUR Workshop Proceedings.
- Simon Gabay, Thibault Clérice, Pauline Jacsont, Elina Leblanc, Marie Jeannot-Tirole, Sonia Solfrini, Sophie Dolto, Floriane Goy, Carmen Carrasco Luján, Maddalena Zaglio, Myriam Perregaux, Juliette Janès, Benoît Sagot, Rachel Bawden, Rasul Dent, Oriane Nédey, and Alix Chagué. 2024. [Reconnaissance des écritures dans les imprimés. CATMuS print: un modèle générique, multilingue et diachronique](#). In *Humanistica 2024*, Meknes, Morocco.
- Claire Gorrara. 2003. *The Roman Noir in Post-War French Culture : Dark Fictions*. Oxford University Press.
- Mike Kestemont. 2014. [Function words in authorship attribution. from black magic to theory?](#) In *Proceedings of the 3rd Workshop on Computational Linguistics for Literature (CLFL)*, pages 59–66, Gothenburg, Sweden.
- Monique Krötsch. 2007. [Répétition et progression en français parlé](#). *Linx - Revue des linguistes de l'université Paris X Nanterre*, 57.
- Michel Lafon and Benoît Peeters. 2006. *Nous est un autre. Enquête sur les duos d'écrivains*. Flammarion, Paris.
- Michel Levine and Jean-Patrick Manchette. 1968. *Les Globe Trotters : Objectif Mexico*. G.P. Presses de la Cité.
- Jean-Patrick Manchette. 2005. *Romans noirs*. Gallimard.
- Jean-Patrick Manchette. 2015. *Journal, 1966-1974*. Gallimard, Paris.
- Serge Quadruppani. 2020. [Jean-patrick manchette, l'écriture de la radicalité](#). *Lundi matin*, 246.
- Upendra Sapkota, 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*, pages 93–102, Denver, USA.
- Christof Schöch. 2013. [Fine-Tuning our Stylometric Tools: Investigating Authorship, Genre, and Form in French Classical Theater](#). In *Digital Humanities 2013 : Conference Abstracts*, pages 383–86, Lincoln, USA.
- Sonia Solfrini, Simon Gabay, Maxime Humeau, Ariane Pinche, Pierre-Olivier Beaulnes, Aurélia Marques Oliveira, Geneviève Gross, and Daniela Solfaroli Camillocci. 2024. [Océriser les imprimés du xvème siècle en langue française : le cas d'un corpus romand en caractères gothiques](#). In *Humanistica 2024*, Meknes, Morocco.